Commodity Return Predictability: Economic Value and Links to the Real Economy

John Cotter, Emmanuel Eyiah-Donkor, and Valerio Potì*
Michael Smurfit Graduate Business School, University College Dublin, Ireland

Abstract

This paper finds out-of-sample predictability of commodity excess returns using forecast combinations of 28 potential predictors. Such gains in forecast accuracy translate into economically significant improvements in certainty equivalent returns and Sharpe ratios for a mean-variance investor. Commodity return forecasts are closely linked to the real economy. Return predictability is countercyclical: stronger in business cycle recessions vis-à-vis expansions, and the combinations of individual predictors have strong predictive power for future economic activity. By using forecast combination methods, which provide insurance against model instability and model uncertainty, we reconcile our findings with the literature that documents the poor out-of-sample performance of individual predictive models.

Keywords: Commodity futures; Return predictability; Out-of-sample forecasts; Asset allocation; Real economy

JEL classification: C52, C53, G11, G13, Q02

*john.cotter@ucd.ie (John Cotter); emmanuel.eyiah-donkor@ucd.ie (Emmanuel Eyiah-Donkor); valerio.poti@ucd.ie (Valerio Potì). The authors would like to acknowledge the support of Science Foundation Ireland under Grant Numbers 16/SPP/3347 and 17/SP/5447. We also acknowledge comments from Thomas Conlon, Pasquale Della Corte, Anna-Maria Fuertes, Richard Levich, Marta Szymanowska, and conference participants at the 2017 International Symposium of Forecasting 2017 and the 2018 Commodity and Energy Markets Annual Meeting.

1. Introduction

Compared to the vast literature on the predictability of aggregate stock, bond, and currency returns (see, for example, Cochrane, 2011 and the references therein), the predictability of aggregate commodity returns has received little attention. This is despite the fact that commodity futures play an important role in explaining fluctuations in and forecasting macroeconomic activity (Hamilton, 2009), and interest in commodities as an alternative investment asset class has grown tremendously in recent years (Fuertes, Miffre, and Rallis, 2010; Erb and Harvey, 2016).

In this paper, we provide a comprehensive study of the time-series predictability of aggregate commodity excess returns–the return on the S&P Goldman Sachs Commodity Index (S&P GSCI) less the return on the one-month T-bill rate. To examine this issue, we consider return-forecasting models that differ in the way they specify variability of the mean return based on 28 potential predictors. In addition to forecasts from individual predictive models, we also consider forecast combinations to account for model instability and uncertainty that typically plague individual return-forecasting models leading to their poor out-of-sample performance. We implement 16 combination forecasts ranging from simple averaging schemes of individual forecasts to more sophisticated ones, for a total of 44 time-varying expected return models. As a benchmark model, we consider a simple no-predictability (historical average) return model against which we compare the performance to our time-varying mean models.

To measure the statistical performance of the return forecasts, we use the out-of-sample (OOS) R^2 which measures the proportional reduction in mean square forecast error (MSFE) between the time-varying forecasts relative to the historical average forecast. We measure the statistical significance of the OOS R^2 using the p-value of the MSFE-adjusted statistic of Clark and West (2007). Consistent with the evidence for equities in Welch and Goyal (2008), we find that the majority of the individual predictive model forecasts generate negative and statistical insignificant OOS R^2, implying that the historical average forecast generates a lower OOS MSFE. On the other hand, we find positive out-of-sample R^2 values of 0.30–4.96% which are statistically significant for all the combination forecasts. We reconcile the results for the individual and combination forecasts by testing the stability of individual predictive models using the t-statistic from the Giacomini and Rossi (2009) forecast breakdown test. The results of the forecast breakdown tests suggests that the reason for the poor performance of the individual models is breakdown in the correlation between returns and individual predictors. It also explains the superior performance of the combination forecasts as the later provides insurance against model instability and model uncertainty. Our findings show strong evidence of instability between commodity returns and the predictors.

We evaluate the economic significance of return predictability by examining the portfolio benefits for an investor. We consider a mean-variance investor with a relative risk aversion of three who exploits this predictability when forming optimal portfolios composed of commodities and risk-free T-bills. We find that the gains in predictive accuracy from combination forecasts of commodity excess returns translate into higher Sharpe ratios and certainty equivalent return gains for the investor. For example, the investor who uses the combination forecasts would have realized an annualized Sharpe ratio of 0.38 compared to 0.02 for the historical average benchmark forecast. Moreover, the investor would be willing to pay a fee of 3.46 per annum to have access to the portfolios generated by the combination forecasts relative to the portfolio generated by the benchmark forecast. These results imply that, on a risk-adjusted basis, the dynamic strategies performed better than the static strategy. The investor, however, would earn lower Sharpe ratios and face certainty equivalent return losses for optimal portfolios generated using the majority of the individual forecasts compared to portfolios generated using the historical average forecast.

We further examine the drivers of commodity return predictability. First, we address the sources of predictability by analysing the extent to which commodity return predictability is related to the business cycle. Using the NBER-dated business cycle indicator, we find that commodity return predictability is largely concentrated in economic recessions, with R^2 values as high as 18.18 percent. Second, and in the spirit of Cochrane (2008), we test whether our predictor variables forecast economic activity. If time-varying commodity excess return forecasts are more plausibly related to macroeconomic risk, then the predictors used to forecast commodity returns should also have forecasting power for macroeconomic activity. We find that forecasts of economic activity variables such as growth in industrial production based on combinations of the individual predictors display statistically significant predictability. Forecast of economic activity based on the individual predictors are, however, poor. These findings show that combination forecasts for commodity returns are closely linked to the real economy, and that time variation in commodity excess returns are only explained by the combination forecasts.

As a final contribution, we endeavour to improve on the forecasting performance of our models by implementing the conditional forecasting procedure of Giacomini and White (2006) that augments forecast selection by conditioning on a set of monitoring instruments. It is a decision-rule that tracks forecast performance over time by selecting the best forecast between a benchmark (the historical average return in our case) and the individual (combination) forecasts. Using money stock (M2) as a monitoring instrument,[1] we find only marginal improvement in forecast performance. That is the individual forecasts continue

[1] We experiment with other monitoring instruments including growth in consumer price index, the VIX, and the macroeconomic uncertainty measure of Jurado, Ludvigson, and Ng (2015) and find that M2 performs best in forecasting the squared error loss differential.

to display statistically and economically insignificant degree of predictability. However, there is marginal improvement in the performance of the combination forecasts.

Our study is related to a strand of literature that investigate the time series predictability of commodity returns. Consistent with the classical theories of storage (Kaldor, 1939; Brennan, 1958) and normal backwardation (Keynes, 1930; Hicks, 1939), many articles have provided evidence of the predictive power of commodity market variables such as basis, hedging pressure, and inventory. For example, Gorton, Hayashi, and Rouwenhorst (2013) find that individual commodity futures risk premia are driven by the basis and inventory levels. Other studies also examine the relationship between commodity returns and macroeconomic variables. Because of short-term mismatches between the demand and supply of commodities due to the business cycle, the general state of the economy is expected to influence commodity prices (see, for example, Bessembinder and Chan, 1992). Gargano and Timmermann (2014) show that macroeconomic variables such as the 3-month treasury bill rate, the term spread, the growth rate of consumer price index, money supply, among others, have forecasting power for raw industrials and metals commodity index returns.

Our study departs from the previous ones along the following dimensions. First, we examine commodity return predictability using a much broader set of predictors selected from the commodity return, stock return, bond return, and macroeconomic predictability literature. By considering this large set of candidate predictors, not only do we address the issue of potentially ignoring important predictors, but we also abstract from the debate on the validity of specific theories that have been proposed for understanding the determinants of commodity futures excess returns. Instead, we aim to identify the broad set of variables that have predictive power for commodity excess returns.

Second, whereas previous studies examine the predictability of either individual commodities, commodity spot indices, or an equally weighted portfolio of individual commodity futures (see, for example, Hong and Yogo, 2012; Gorton et al., 2013; Ahmed and Tsvetanov, 2016, and the references therein), we examine the predictability of aggregate commodity excess returns by focussing on an investable and widely used commodity futures index, namely the S&P GSCI.[2] Because of the high storage, transportation and insurance costs associated with holding the physical commodities, investors have traditionally relied on commodity futures to gain exposure to commodities (Edwards and Park, 1996; Jensen, Johnson, and Mercer, 2000).

Another difference between our study and prior studies on the predictability of commodity returns, except the study of Gargano and Timmermann (2014), is that we address the impact of structural instability and uncertainty about return prediction models by implementing several forecast combination methods. We are motivated by studies such as

[2]The S&P GSCI is the benchmark commodity index used by investor to gain broad exposure to commodities through investment vehicles such as commodity-linked exchange traded products.

Pesaran and Timmermann (1995) and Welch and Goyal (2008), among others, who show that the poor out-of-sample forecasting performance of predictive models of U.S. equity excess returns might be a direct consequence of structural breaks and model uncertainty in the underlying data generating process. Pesaran and Timmermann (1995) show that this is true even when very informative predictors are used. Such uncertainty about the correct model and structural breaks in the data, which is difficult to anticipate, may lead to the best prediction model changing over time. Combination forecasts provide insurance against these features by diversifying across the individual model forecasts. This leads to a reduction in forecast variance and improved forecasts as shown in Stock and Watson (2004) for forecasting the output growth, and Timmermann (2006) and Rapach, Strauss, and Zhou (2010) for forecasting the U.S. equity risk premium.

Our study also address concerns raised by previous studies on return predictability, that find that statistical evidence of predictability does not always translate into economic significance (Della Corte, Sarno, and Tsiakas, 2008; Thornton and Valente, 2012; Sarno, Schneider, and Wagner, 2016). We do this by examining the benefit of predictability to a risk-averse investor in a mean-variance optimal asset allocation framework. Finally, we extend prior studies by examining whether predictability has links to the real economy. First, we test whether predictability is concentrated in certain phases of the business cycle, and second, we examine whether, as a potential explanation for which forecasts (individual or combinations) capture time-variation in commodity risk premia, commodity return forecasts have predictive power for macroeconomic activity .

The rest of the paper is as follows. Section 2 describes the return prediction models we consider, and the framework for evaluating out-of-sample return predictability. Section 3 describes the commodity returns data and predictor variables, and presents the empirical results. In Section 4, we analyse the link between commodity return predictability, portfolio performance and the business cycle. Section 5 examines the economic drivers of commodity return predictability. In Section 6, we discuss a conditional predictability test that track forecast performance over time and implement a forecast decision-rule designed to improve upon our forecasts. Section 7 concludes.

2. Methodology for Predicting Commodity Returns

We next introduce the return prediction models: individual predictive models that condition on each of the 28 predictors at a time to generate a forecast, and the forecast combination methods that combine the individual predictive model forecasts using different combining weights resulting in 16 combination forecasts. We also detail the statistical and economic measures that we use in evaluating the performance of out-of-sample excess commodity return forecasts.

2.1. Return Predictability Models

2.1.1. Individual Predictive Model Forecasts

Consider the following bivariate predictive regression model for excess commodity returns,

$$r_{t+1} = \alpha + \beta x_{i,\,t} + \varepsilon_{i,t+1}, \tag{1}$$

where r_{t+1} is the realized log excess return on commodity futures index from time t to $t+1$, $x_{i,\,t}$ is a predictor available at time t, and $\varepsilon_{i,t+1}$ is a zero-mean error term. The step-ahead forecast of log excess returns is given by

$$\hat{r}_{t+1} = \hat{\alpha}_t + \hat{\beta}_t x_{i,\,t}, \tag{2}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the OLS estimates of α and β in Equation (1), respectively.

2.1.2. Combination Forecasts

A big issue that arises when studying return predictability is to decide what economic variables have predictive power for asset returns, especially when the set of possible predictors is large. One possibility is to use financial theory to guide the selection of the relevant variables. The difficulty is that economic theory alone does not provide enough guidance and so one is likely to ignore potentially important predictors. It is also very difficult to identify a single best model because of the difficulty in identifying structural breaks in the data and the uncertainty that surrounds return forecasting models. Combination forecasts incorporate information from many predictors and therefore should provide insurance against model uncertainty and parameter instability by diversifying across the individual predictive model forecasts. We consider three types of combination forecasts in this paper: 4 simple combination forecasts; 9 performance-based combination forecasts; and 3 factor model forecasts. Our combination forecasts differ in the way we compute weights assigned to the individual predictive model forecasts. As noted by Timmermann (2008), which combination method is ex ante optimal is an empirical question and justifies why we consider different forecast combination methods.

Let $\hat{r}_{i,\,t+1}$ denote the pseudo out-of-sample forecast of the realization r_{t+1} computed at time t based on the ith predictor variable as given by Equation (2). Most of the forecast combination methods we consider take the following form:

$$\hat{r}^{\text{CF}}_{t+1} = \sum_{i=1}^{N} w_{i,\,t}\hat{r}_{i,\,t+1}, \tag{3}$$

where $\hat{r}^{\text{CF}}_{t+1|t}$ is the combination forecast and $w_{i,\,t}$ is the weight assigned to the ith forecast with $\sum_{i=1}^{N} w_{i,t} = 1$.

The first set of combination methods we consider use simple averaging schemes: mean, trimmed mean, median, and weighted-mean forecasts. Stock and Watson (2004) find that simple combining methods work well in forecasting inflation and output growth using a large number of potential predictors. Similarly, Rapach et al. (2010) find that simple methods work well for forecasting the U.S. equity risk premium. The mean combination forecast, $\hat{r}_{t+1}^{\text{Mean}}$, is the average of the N individual forecasts that assign equal weights, $w_{i,t} = 1/N, i = 1, ..., N$, to each forecast defined in Equation (2):

$$\hat{r}_{t+1}^{\text{Mean}} = \frac{1}{N}\hat{r}_{1,t+1} + \frac{1}{N}\hat{r}_{2,t+1} + ... + \frac{1}{N}\hat{r}_{N,t+1}. \tag{4}$$

The trimmed mean forecast, $\hat{r}_{t+1}^{\text{Trimmed mean}}$, sets in Equation (3) $w_{i,t} = 0$ for the smallest and largest forecasts, and $w_{i,t} = 1/(N-2)$ for the remaining individual forecasts. The median combination forecast, $\hat{r}_{t+1}^{\text{Median}}$, is the sample median of the N individual predictive model forecasts. The weighted-mean forecast ($\hat{r}_{t+1}^{\text{Weighted-mean}}$) proposed by Bates and Granger (1969) specifies the combination weights to be proportional to the inverse of the estimated residual variance, $\sigma_{i,t}^2$, for the individual predictive regression models given by Equation 1,

$$\hat{r}_{t+1}^{\text{Weighted mean}} = \frac{1/(\hat{\sigma}_{1,t}^2)}{\sum_{i=1}^{N} 1/(\hat{\sigma}_{i,t}^2)}\hat{r}_{1,t+1} + \frac{1/(\hat{\sigma}_{1,t}^2)}{\sum_{i=1}^{N} 1/(\hat{\sigma}_{i,t}^2)}\hat{r}_{2,t+1} + ... + \frac{1/(\hat{\sigma}_{1,t}^2)}{\sum_{i=1}^{N} 1/(\hat{\sigma}_{i,t}^2)}\hat{r}_{N,t+1}, \tag{5}$$

The second set of combination methods consist of several performance-based combination forecasts. First, we compute the discounted mean squared forecast error (DMSFE) combination forecast following Stock and Watson (2004). Here, the combining weights are specified as functions of the historical performance of the individual predictive model forecasts over a holdout out-of-sample period,

$$w_{i,t}^{\text{DMSFE}} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}}, \quad \phi_{i,t} = \sum_{s=1}^{t-1} \theta^{t-1-s} \left(r_{s+1} - \hat{r}_{i,s+1}\right) \tag{6}$$

where θ is the discount factor.[3] When $\theta < 1$, greater importance is attached to the individual predictive model forecast with the lower mean square forecast error (MSFE). That is, the individual predictive model that generates the least MSFE is assigned a greater weight because it signals better forecasting performance. In the special case where there is no discounting ($\theta = 1$) and forecasts are uncorrelated, this leads to the optimal combin-

[3]The DMSFE combination forecast require a holdout evaluation period to estimate the combining weights. However, note that the first out-of-sample forecast of this method is simply calculated as the mean combination forecast because there is no past individual forecast used to form the DMSFE weight at this time point.

ation weights proposed by Bates and Granger (1969) given by Equation (5). We consider θ values of 0.7 and 0.9. Rapach et al. (2010) also show that the DMSFE combination forecasts of U.S. equity excess returns consistently outperforms a constant expected excess return benchmark forecast. Second, we consider an Approximate Bayesian Model Averaging (ABMA) combination forecast following Garratt, Lee, Pesaran, and Shin (2003) and choose the combining weights as follows:

$$w_{i,\,t}^{\text{ABMA}} = \frac{\exp(\triangle_{i,t})}{\sum_{j=1}^{N}\exp(\triangle_{i,t})}, \tag{7}$$

where $\triangle_{i,t} = \text{AIC}_{i,t} - \max_j(\text{AIC}_{j,t})$ and $\text{AIC}_{i,t}$ is the Akaike Information Criterion of model i. The ABMA thus gives higher weight to models with better historical fit as measured by the AIC. The ABMA combination has the advantage that, it has a firmer information-theoretic rationale. For example, Detzel and Strauss (2017) find that DMSFE and ABMA combination forecasts generate more accurate forecasts of the value weighted return on Fama-French thirty eight and forty eight industry portfolios.

Third, Elliott, Gargano, and Timmermann (2013, 2015) propose a new class of combination forecasts which they call Subset Regression forecasts. Their approach use an equally weighted combination of forecasts based on all possible predictive regression models that include a subset of the predictor variables. As noted by the authors, by keeping the number of predictors to be included in the predictive model fixed, they are able to control estimation error by trading off the bias and variance of the forecast errors similarly to generating the mean-variance efficient frontier of individual assets in portfolio theory. In an application to the predictability of U.S. equity excess returns, Elliott et al. (2013) find that combination of subset regression forecasts produce more accurate forecast than approaches based on equally weighting individual forecast or forecasts generated by Bayesian model averaging. Suppose the number of potential predictors that enter a regression is K. A subset regression is then defined by the set of regression models that include a specified number of regressors, $k \leq K$. The $k \leq K$ dimensional subset forecasts are then averaged to generate the forecasts. In our analysis, we use a maximum K value of 7. Given K regressors in full and k regressors chosen for short models, one has to average over $C_k^K = K!/(k!(K-k)!)$ subset regression forecasts. As a special case, when $k = 1$, this results in the mean combination forecast given by Equation (4). Generally, the Subset Regression forecast is given by

$$\hat{r}_{t+1}^{\text{Subset}} = \frac{1}{C_k^K}\sum_{i=1}^{C_k^K}\hat{\beta}_{i,\,t}x'_{i,\,t}, \tag{8}$$

where $\dim(x_{i,\,t}) = k$.

Finally, we generate out-of-sample forecasts by estimating a predictive regression based on diffusion index that assumes a latent factor structure following Stock and Watson (2002a,b):

$$\hat{r}^{\mathrm{PC}}_{t+1} = \hat{\alpha} + \sum_{k=1}^{K} \hat{\beta}_{k,\,t} F_{k,\,t}, \tag{9}$$

where $F_{k,\,t}$ is the kth principal component extracted from our 28 predictor variables. Diffusion indexes provide a convenient way of extracting common factor from a large number of potential predictor variables. Neely, Rapach, Tu, and Zhou (2014), for example, show that this approach improves the forecasting performance of U.S. equity excess returns. We consider models where the principal components are selected via the Akaike information criterion (AIC),[4] the Bayesian information criterion (BIC), and the adjusted R^2 statistical model selection criterion. We set the maximum number of principal components to 4.

2.2. Historical Average Return Benchmark Forecast

As a simple no-predictability benchmark, we use a constant expected excess return model:

$$r_{t+1} = \alpha + \varepsilon_{t+1}, \tag{10}$$

This is a popular benchmark model that has been used widely in studies of return predictability (see, for example, Welch and Goyal, 2008; Rapach and Zhou, 2013; Ahmed and Tsvetanov, 2016; and the references therein). The use of this model as the benchmark is also consistent with the hypothesis that commodity futures prices follow a random walk so their returns are unpredictable (Alquist and Kilian, 2010; Chinn and Coibion, 2014). We refer to it as the historical average (HA) model. We use the forecast from this model as the benchmark forecast against which all other forecasts are compared in assessing commodity return predictability.

[4] The Akaike information criterion (similarly to the adjusted R^2 selection criterion), unlike the Bayesian information criterion (BIC), is not statistically consistent, in the sense of selecting the "true" model, as the sample size increases without bounds. However, Pesaran and Timmermann (1995) note that in the context of forecasting asset returns where the correct list of regressors is unknown and may be changing over time, the consistency property of a model selection criterion is not as important as it may first appear. They suggest that of greater importance is to select a forecasting model that could be viewed at the time as being a reasonable approximation to the data generating process. Although AIC is statistically inconsistent, it has the property of yielding an approximate model. Shibata (1976), for example, shows that AIC strikes a good balance between giving biased estimates when the order of the model is too low, and the risk of increasing the variance when too many regressors are included.

2.3. Statistical Evaluation of Commodity Return Predictability

2.3.1. Out-of-sample Return Forecasts

We generate the out-of-sample forecasts using a recursive (expanding window) estimation scheme as follows. Suppose T observations are available for r_t and $x_{i,t}$. To initialize our parameter estimates for the individual predictive model forecasts, we use the first $n = 167$ observations (February 1976 to December 1989) as the in-sample estimation period and the remaining $T - n = 312$ observations (beginning in January 1990) as the out-of-sample forecast evaluation period. The choice of length of the in-sample estimation period enables us to have a sufficiently long out-of-sample forecasts evaluation period. Hansen and Timmermann (2012), for example, show that using a relatively large proportion of the available sample for forecast evaluation leads to better size properties of the test statistics of predictive ability. The model parameters are updated recursively as new data becomes available. Meaning that the estimation sample always starts in 1976:02 and we expand the estimation window by one month as additional observations become available. Only data up to the previous month is therefore used to estimate the model parameters and generate the pseudo out-of-sample forecast of excess commodity returns corresponding to each predictor variable for the month $t + 1$.

2.3.2. Statistical Measures of Performance

Following the convention in the return predictability literature, we measure the accuracy of the out-of-sample return forecasts generated by time-varying return prediction models relative to the benchmark HA return forecast using the Campbell and Thompson (2008) out-of-sample R^2 statistic, R^2_{OOS}, given by:

$$R^2_{\text{OOS}} = 1 - \frac{\text{MSFE}(\hat{r}_t)}{\text{MSFE}(\bar{r}_t)} = 1 - \frac{\sum_{t=n+1}^{T} (r_t - \hat{r}_t)^2}{\sum_{t=n+1}^{T} (r_t - \bar{r}_t)^2}, \tag{11}$$

where r_{n+1} is the realized log return at time $n + 1$ and $\hat{r}_{n+1}(\bar{r}_{n+1})$ is an alternative (HA forecast), individual predictive regression or the combination forecast, forecast. The R^2_{OOS} statistic measures the proportional reduction in mean square forecast error (MSFE) for an alternative forecast relative to the HA forecast. A positive R^2_{OOS} implies the alternative forecast, because it has a lower MSFE, outperforms the HA forecast.

We evaluate the statistical significance of the R^2_{OOS} statistic using the p-value of the MSFE-adjusted statistic of Clark and West (2007). The statistic tests the null hypothesis of equal out-of-sample predictive ability of the alternative model forecasts against the benchmark HA model forecast. That is, $R^2_{\text{OOS}} \leq 0$ against the alternative hypothesis that $R^2_{\text{OOS}} > 0$. Under the null of no-predictability, the HA return forecast is expected to have a lower MSFE. The Clark and West (2007) procedure accounts for the fact that under the null of equal predictive accuracy, the MSFE of the HA model is expected to be lower compared to the alternative models. This is because the alternative model introduces noise into its forecasts by attempting to estimate parameters whose population values are zero. As such, finding that the HA model forecast has a lower MSFE is not clear evidence against the alternative model. Clark and West (2007) recommend to reject the null of equal MSFE if the test statistic has critical values greater than 1.282 for a one-sided 10% test, 1.645 for a one-sided 5% test, and 2.326 for a one-sided 1% test.

In addition to our formal test of significance of the R^2_{OOS}, we also use the forecast breakdown test of Giacomini and Rossi (2009) to test the stability of the individual predictive models. This test is designed to detect forecast breakdowns by assessing whether a model that display good forecasting performance in one sample period will continue to do so in other sample periods. In our framework, the null hypothesis of the forecast breakdown test is that the out-of-sample MSFE of a model is equal to its in-sample MSFE. We test this hypothesis using a one-sided t-statistic for our recursive forecasts. The one-sided t-test focusses on the alternative hypothesis that the out-of-sample MSFE of a model is higher than its in-sample MSFE.

2.4. Economic Evaluation of Commodity Return Predictability

We next detail the asset allocation framework that we use to evaluate the economic significance of commodity return predictability. We test whether any statistical evidence of commodity return predictability translates into economic gains for a risk-averse investor.[5] We are motivated by studies such as Della Corte et al. (2008) and Potì (2018) for exchange rate returns, and Thornton and Valente (2012) and Sarno et al. (2016) for bond return predictability, who find that statistical evidence of return predictability does not always translate into economic significance. By evaluating return predictability from the economic perspective, we also address the limitations of studies on commodity return predictability that only focus on statistical tests of return predictability.

2.4.1. Dynamic Asset Allocation

Following Campbell and Thompson (2008), among others, we consider a mean-variance investor who monthly allocates her wealth between commodities and risk-free T-bills using either the individual predictive regression forecasts (combination forecasts) or HA forecast

[5]The R^2_{OOS} statistical evaluation metric does not take into account the risk that an investor would have to bear over the out-of-sample forecast evaluation period. Leitch and Tanner (1991) in the context of studying why firms purchase professional forecasts of economic and financial variables argue that the profitability of a forecast is a more relevant metric for assessing forecasts. They show that using such a metric explains the value of forecasts to firms even when forecasts fail to beat simple models in terms of MSFE.

of excess commodity returns. The investor optimally allocates the following share of her portfolio to commodities during the subsequent month $t + 1$

$$w_t = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right), \tag{12}$$

where γ is the investor's relative risk aversion coefficient, $\hat{r}_{t+1}$ is the simple excess return forecast and $\hat{\sigma}^2_{t+1}$ is the excess return variance. Following Campbell and Thompson (2008), among others, we assume that the investor uses a five-year rolling-windows of past returns to estimate the variance of commodity futures excess return. As in Campbell and Thompson (2008) and Rapach et al. (2010), we set risk aversion coefficient equal to 3. Since we use futures, we avoid short sale restrictions. However, we impose a realistic constraint on portfolio leverage by limiting leverage to 50% of wealth, similar to Campbell and Thompson (2008) and Rapach, Ringgenberg, and Zhou (2016). These constraints should mitigate excessive risk taking and produce plausible portfolio weights considering the well-known sensitivity of mean-variance optimal weights to return forecasts.

2.4.2. Economic Measures of Performance

The investor who uses the individual and combination forecasts to compute portfolio weights realizes average utility or certainty equivalent return (CER) is given by

$$\text{CER}(r_p) = \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}^2_p, \tag{13}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p$ are the mean and standard deviation, respectively, of portfolio excess returns over the forecast evaluation period. The CER is the return on a risk-free T-bill that an investor would be willing to accept rather holding a risky portfolio. The CER for the investor who uses the historical average forecast to compute portfolio weights is calculated similarly. Our direct measure of the economic significance of return predictability is the CER gain ($\triangle$): the difference between the CER of the portfolio generated by the individual or combination forecasts and the portfolio generated by the HA return forecast. We annualize the CER gain so that it can be be interpreted as the annual portfolio management fee that the investor would be willing to pay to have access to the portfolio generated by the individual or combination forecast (dynamic portfolio strategy) relative to the portfolio generated by the HA return forecast (static portfolio strategy). Positive values indicate that the time-varying predictability models perform better the HA model. A CER gain of 2% or more is usually considered to be economically significant (see, for example, Rapach et al., 2010, and the references therein.). We also report the annualized Sharpe ratio (SR) computed as the ratio of the mean of portfolio excess returns to its standard deviation.

A realistic assessment of the profitability of any dynamic asset allocation strategy should take into account the effect of transaction costs. With sufficiently high costs of trading, we should expect the dynamic portfolio strategies to be costly to implement relative to the static strategy because of fluctuations in their portfolio weights. We account for the effect of transaction costs in two ways. First, we compute our performance measures for the investor's realized portfolio returns net of transaction where we set We set the proportional transaction cost to 20 basis points per dollar of trading. Second, we calculate the break-even proportional transaction costs, τ^{BE}, that will render the investor indifferent between two competing portfolio strategies following Della Corte et al. (2008) as

$$\tau^{\mathrm{BE}} = \frac{\bar{r}_p^{\mathrm{FC}} - \bar{r}_p^{\mathrm{HA}}}{\mathrm{TO}^{\mathrm{FC}} - \mathrm{TO}^{\mathrm{HA}}}, \tag{14}$$

where $\bar{r}_p^{\mathrm{FC}}$ and $\bar{r}_p^{\mathrm{HA}}$ are the portfolio mean returns of the dynamic and static portfolio strategies, respectively, and $\mathrm{TO}^{\mathrm{FC}}$ and $\mathrm{TO}^{\mathrm{HA}}$ are their respective average turnover. In comparing a dynamic portfolio strategy with a static strategy, an investor who pays actual transaction costs lower than the break-even cost will prefer the dynamic strategy. We report the τ^{BE} in basis points, and to facilitate the interpretability of our results, do so only when the CER gain is positive.

3. Empirical Results

This section describes our empirical results. We first describe and provide descriptive statistics for the data and predictor variables. Next, we report results based on full-sample estimates and the out-of-sample analysis of the statistical and economic evidence on commodity return predictability.

3.1. Data on Commodity Futures and Predictor Variables

The dataset used for our empirical analysis is based on end-of-month total return index data on the S&P Goldman Sachs Commodity Index (S&P GSCI) downloaded from Bloomberg. We compute excess return as the log return on S&P GSCI less the log return on a one-month T-bill.[6] The sample period is January 1976 to December 2016.

Panel A of Table 2 presents descriptive statistics of monthly simple returns on the S&P GSCI for the full sample period. The table shows that the mean excess return was 0.143% with volatility close to 6%. On a risk-adjusted basis, the index recorded annualized Sharpe ratio of 0.089. The low mean return and high standard deviation suggests that commodities would be unattractive as a stand-alone investment strategy on a risk-adjusted

[6] The T-bill rate is downloaded from the webpage of Amit Goyal's website, `http://www.hec.unil.ch/agoyal/`.

basis.

3.2. Predictor Variables

We consider a set of 28 predictor variables. They include commodity market, stock market, Treasury market, corporate bond market, currency market, and macroeconomic variables. The commodity and currency market variables include those predictors that have been shown to have predictive power for individual and indexes of commodity spot and futures returns. They include, for example, the basis, crude oil production, and crude oil inventory, and the exchange rate of major commodity exporting countries such as Canada, South Africa, India, and New Zealand against the U.S. dollar. The stock, Treasury, and corporate bond market variables include, for example, the dividend-price ratio, the return on the S&P 500 index, the yield on 3-month Treasury bill rate, term spread default premium, among others. Most of these variables are analysed in the equity risk premium predictability study of Welch and Goyal (2008), and studies on Treasury and corporate bond return predictability including Gargano, Pettenuzzo, and Timmermann (2017) and Lin, Wu, and Zhou (2017), and the references therein. Finally, the macroeconomic variables include, among others, inflation, money stock, unemployment rate, industrial production growth, degree of capacity utilization in U.S. manufacturing, and global real economic activity index. Considering the important role commodity prices play in explaining fluctuations in economic activity make these variables candidate predictors for forecasting commodity returns.

All the predictor we consider have been used in prior studies on commodity return predictability of commodity returns (see, for example, Gargano and Timmermann, 2014; Hong and Yogo, 2012; Groen and Pesenti, 2011; Chen, Rogoff, and Rossi, 2010). Table 1 outlines and defines the predictors that we use, the motivation for their use, and relevant prior commodity return studies.

[Insert Table 1 about here]

Panel B of Table 2 presents the summary statistics for the predictor variables. Except for the commodity currencies, LTR, DFR, M1, and UNRATE, the other predictors are strongly positively autocorrelated with first-order autocorrelation coefficients of around 0.28 to 0.99 which suggests the use of test statistics that are robust to autocorrelation when testing the evidence of predictability.

[Insert Table 2 about here]

3.3. In-Sample Predictability

To test whether the individual predictors have forecasting power for commodity excess returns, we run bivariate predictive regression of log commodity excess returns on each of the lagged values of predictors at a time for the full sample period (1976:02-2016:12). Table 3 report estimates of the slope coefficient, β, the associated Newey and West (1987) autocorrelation and heteroskedasticity-consistent t-statistics, and the R^2 statistic. From the table, we can see that of the 28 predictor variables, one third of these namely LTR, CDFP, DFR, INDPRO, CUTIL, CFNAI, CLI, BCI, and CCI display statistically significant predictive power for excess commodity futures returns at conventional levels.[7] The R^2 statistics for the nine variables range from 1.20% to 6.35%. The CLI and BCI predictors display substantial predictive ability at the 1% level with R^2 statistics of 3.65% and 6.35%, respectively.

The last two columns of Table 3 report R^2 statistics separately for the National Bureau of Economic Research (NBER)-dated business-cycle expansions and recessions. To gauge the strength of predictability during the business cycle, we compute the following version of the conventional R^2 statistic for business-cycle expansions (EXP) and recessions (REC):

$$R^2_{\text{C}} = 1 - \frac{\sum_{t=1}^{T} I_t^c \hat{\varepsilon}_{i,t}^2}{\sum_{t=1}^{T} I_t^c (r_t - \bar{r})^2} \quad \text{for} \quad c = \text{EXP, REC}, \tag{15}$$

where I_t^{EXP} (I_t^{REC}) is an indicator function that takes a value of one when month t is an expansion (recession) and zero otherwise, $\hat{\varepsilon}_{i,t}$ is the fitted error based on the full estimate of the predictive regression model in Equation (1), $\bar{r}$ is full sample mean of r_t, and T is the full sample observations. The table shows that commodity return predictability is stronger in recessions relative to expansions for twenty four out of the 28 predictors. For example, the R^2_{REC} for the DFR, INDPRO, CLI, and BCI more than quadruple during recessions compared to the expansions. In-sample statistical evidence of predictability, while useful, forms only a small part of the story. The true extent of predictability can only be assessed in formal out-of-sample tests and the economic value of such predictability for investor's asset allocation decisions.

[Insert Table 3 about here]

3.4. Statistical Evaluation of Commodity Return Forecasts

The in-sample tests of predictability reported in Table 3 are not based on truly ex-ante measures of future expected commodity futures returns as the predictions would not have

[7] We also examined the forecasting power of additional predictor variables, namely hedging pressure, open interest and the Baltic dry. The results for these variables are not reported as they do not have sufficiently long data record.

been available to an investor in real time because we use the full sample data for estimation. Further, there is the concern of in-sample overfitting which could overstate the true extent of predictability resulting in unusually high Sharpe ratios of the returns on trading strategies. To circumvent this problem and guard against overfitting, Table 4 reports R^2_{OOS} for each of the individual predictive regression model and the combination forecasts relative to the benchmark HA forecast. The statistical significance of $R^2_{\text{OOS}} > 0$ is assessed using the p-value of the MSFE-adjusted statistic of Clark and West (2007).

Panel A of Table 4 report results for the individual predictive regression forecasts. Most of the individual predictors fail to beat the HA forecast in terms of MSFE similarly to the evidence reported in many of the return predictability studies. Only twelve out of the 28 predictors have positive R^2_{OOS}. Four of the positive R^2_{OOS} statistics for CDFP, DFR, CUTIL, and CFNAI range from 1.13% to 3.15% with MSFEs significantly less than the MSFE of the HA forecast at the 5% level. Out of step with these findings are impressive R^2_{OOS} statistics of 3.80% and 6.92% recorded for CLI and BCI, respectively, and are significantly greater than zero at the 1% level. Panel B of Table 4 report results for the combination forecasts where the findings are much more supportive of predictability. The R^2_{OOS} generated by each of the combination forecasts are impressive ranging from 0.30% for the Mean combination forecast to 4.96% for the PC (IC = BIC) combination forecasts, and outperform the HA benchmark forecast. All the combination forecasts have R^2_{OOS} that are significantly greater than zero at the 1% level except the Median and PC (IC=BIC) forecasts that have R^2_{OOS} significantly greater than zero at the 5% level. Later in the paper, we will investigate what might be explaining the good performance from CLI, BCI, and the combination forecasts.

[Insert Table 4 about here]

As earlier hypothesized, structural instability and uncertainty of the individual return prediction models may explain their poor out-of-sample performance. Table 5 reports the t-statistics and the associated p-values of the forecast breakdown test, a test of the stability of the individual predictive models, of Giacomini and Rossi (2009). The out-of-sample forecasts are generated using the same recursive estimation approach, and the test statistics are computed similarly using a quadratic loss function. The null hypothesis that the out-of-sample MSFE of a model is equal to its in-sample MSFE is rejected at the 1% level for all the individual predictors. That is, the models' forecasts fail to display consistently good forecasting performance throughout our sample. These results represents a strong evidence of the structural instability of the individual predictive model forecasts, and explain their poor out-of-sample performance.

[Insert Table 5 about here]

3.5. Economic Evaluation of Commodity Return Forecasts

Table 6 reports the economic significance of predictability as measured by the CER gains, Sharpe ratios, portfolio turnover ratios, and break-even transaction costs. The mean-variance investor's risk-aversion coefficient is set equal to 3, the optimal portfolio weights are between $-\infty$ and 1.5, and transaction costs are set to 20 basis points. CER gains are annualized percent values, Sharpe ratios are annualized values, and break-even transaction costs are reported in basis points. The turnover ratio is the ratio of the average turnover of the portfolio generated by the HA return forecast (static portfolio strategy) to the average turnover of the portfolio generated by the individual or combination forecast (dynamic portfolio strategy). The break-even transaction cost is the transaction cost that will render the investor indifferent between the dynamic portfolio strategy and the static portfolio strategy. A positive CER gain indicates that the CER of the dynamic portfolio strategy is greater than that of the static strategy. A CER gain of 2% or more is usually considered to be economically significant.

From Column 6 of Panel A in Table 6, we see that almost all the individual predictor realize negative CER gains in accord with the R^2_{OOS} statistical significance results reported in Table 4. Positive values well above the 2% significance level are documented for four predictors, namely CDFP, CUTIL, CLI and BCI, out of the 28 predictors, and provide support for the out-of-sample statistical evidence of predictability reported in Table 4. The CER gains associated with the combination forecasts are reported in Panel B of Table 6. Again strong support is given for their predictive power for commodity return and we realize positive CER gains for all combination forecasts. The Subset ($k = 2, ..., 7$) and the PC, $\text{IC} = R^2$ forecasts all record CER gains well above the 2% significance level. The results for the combination forecasts are also in accord with the statistical evidence of predictability reported in Table 4.

Similar to the results for the CER gains portfolio performance measure, Sharpe ratios of the commodity portfolio generated by the individual predictive model forecasts are lower, for almost all predictors, than the portfolio that relies on the historical average return forecast. These results are also consistent with the poor statistical performance of the individual forecasts which has translated into bad economic performance. Turning to the results based on the combination forecasts, a commodity portfolio that exploits predictability using any of the combining forecasts would have generated Sharpe ratios that are substantially higher than those generated by the historical average forecast.

[Insert Table 6 about here]

Column 7 of Panel A in Table 6 reports the CER gains net of transaction costs results for the individual predictive model forecasts. From the table, we observe that just as in the results without transaction costs, almost all the portfolio generated by individual predictive mode forecasts realize negative CER gains. On the other hand, accounting for the effect of transaction costs does not erode the performance of the combination forecasts as they continue to deliver positive CER gains net of transaction costs well above the 2% significance level. This performance, however, comes at the cost of a higher average turnover. For example, the Subset ($k = 2, ..., 7$) combination forecasts deliver CER gains net of costs of 2.2% compared to 2.6% without transaction costs. The break-even transaction costs values are also much higher than the actual proportional transaction cost for the combination forecasts meaning that investors would prefer the portfolios based on the combination forecasts. Sharpe ratios net of transaction costs results are consistent with the CER gains net of transaction costs results.

4. Commodity Return Forecasts and the Macroeconomy

We conduct further analysis to shed more light on the economic drivers of commodity return predictability and portfolio performance by investigate the link between return forecasts and the real economy. Such links should provide additional support for the performance of the forecasts based on the CLI and BCI predictors, and the forecast combination methods, and the economic rationale for the portfolio performance.

4.1. Commodity Return Forecasts and the Business-cycle

Studies such as Rapach et al. (2010), Henkel, Martin, and Nardari (2011) and Gargano and Timmermann (2014) show that the predictability of stock and commodity returns is stronger during business-cycle recessions compared to expansions. These findings suggest a link between return predictability and cyclical variation of expected returns. To test this hypothesis, we use the same version of the conventional R^2 statistic for business-cycle expansions (EXP) and recessions (REC) defined earlier for our analysis of in-sample tests of predictability.

Table 7 reports the out-of-sample R^2, the Clark and West (2007) MSFE-adjusted statistic and associated p-values separately for NBER-dated business-cycle expansions (R^2_{EXP}) and recessions R^2_{REC}. Panel A of the table report results for the individual predictive model forecasts. Almost all the individual predictive model forecasts fail to outperform the historical average forecast in terms of MSFE during both recessions and expansions. DFR, CLI and BCI predictors, however, continue to show significant levels of predictability with significantly greater than zero R^2 statistic at the 5% level during recessions and expansions. These results are is in sharp contrast to our earlier in-sample findings where we documented that 24 out of the 28 predictors display statistically significant predictability in recessions relative to expansions. One take away from these results is that evidence of in-sample predictability does not always guarantee significant evidence of out-of-sample predictability.

Panel B of Table 7 report results for the combination forecasts. None of the forecasts has statistically greater than zero R^2_{OOS} statistics during expansions. However, during business-cycle recessions, the combination forecasts deliver R^2 values ranging 0.73% to 18.18% which are statistically greater than zero at the 5% level. The results show that predictability is stronger in recessions relative to expansions and are supportive of the findings in Gargano and Timmermann (2014).

[Insert Tables 7 about here]

4.2. Economic Performance of Commodity Portfolios and Links to the Macroeconomy

We now examine whether the portfolios generated by the commodity return forecasts are also related to the business-cycle. We use the same asset allocation framework detailed earlier, and report results separately for the NBER-dated business cycle expansions and recessions.

4.2.1. Variation in Risk Premia

Does the documented statistically significant evidence of commodity return predictability is countercyclical and related to time-variation in risk premia? Asset pricing models featuring habit persistence such as Campbell and Cochrane (1999) suggest that risk premia move countercyclically and that the Sharpe ratio of the aggregate stock market should be higher during recessions relative to expansions due to a reduced surplus consumption ratio. Wachter (2006) derives implications for bond risk premia and the term structure of interest rates in a setting with habit persistence. If risk premia varies with the business-cycle, then the portfolios generated by the return forecasts should perform better in recessions relative to expansions.

Table 8 reports Sharpe ratios (Sharpe ratios net of transaction costs of 20 basis points) results computed separately for NBER-dated business cycle expansions and recessions based on the same asset allocation framework detailed earlier. We use the full out-of-sample forecast evaluation period so as to ensure that there are enough observations for the separate analysis of recessions. The Sharpe ratios for the individual forecasts reported in Panel A are mixed. For example, the Sharpe ratios of the portfolio based on CLI and BCI predictors have substantially high Sharpe ratios in recessions relative to expansions. In contrast, the Sharpe ratio results for all the combination forecasts provide strong support for the suggestion of Campbell and Cochrane (1999). We can see that Sharpe ratios of portfolios based on all the forecasts from the combination methods are substantially higher in recessions relative to expansions.

4.2.2. Business-cycle phases and Economic Performance of Commodity Portfolios

Table 8 reports results of economic significance as measured by CER gains (CER gains net of proportional transaction costs of 20 basis points) separately for business-cycle expansions and recessions. The out-of-sample portfolio performance analysis demonstrates the economic value of commodity return predictability with results concentrated in the recessionary phases of the business-cycle relative to expansions, especially for all the combination forecasts. The results for the individual predictive model forecasts are mixed. However, this is not surprising considering their poor performance as earlier shown.

[Insert Tables 8 about here]

Our results taken together show that predictability tracks business conditions so that expected returns, and for that matter portfolio performance, are high when business conditions are weak and vice-versa.

5. What Drives Commodity Return Predictability?

Cochrane (2008, 2017) suggests that time-varying equity risk premium forecasts are more likely related to macroeconomic risk if the predictors used to forecast returns also have predictive power for the business-cycle. Stock and Watson (2003, 2004), for example, show that forecasts of the output growth and inflation based on individual predictor variables are highly unstable over time compared to combination of forecasts. This provides a potential explanation for the poor forecasting performance of the individual predictors and the impressive performance of the combination forecasts. Since the individual predictors produce return forecasts that are highly unstable overtime as indicated by the forecast breakdown test results of Giacomini and Rossi (2009), we should observe a lack of forecasting power when the same predictors are used to predict macroeconomic activity. In contrast, the significant performance of the combination forecasts means that we should expect significant combination forecasts of macroeconomic activity. We provide support for this explanation by examining whether combinations of individual predictors have forecasting power for economic activity.

Consider the following autoregressive distributed lag model:

$$y_{t+1} = \alpha + \beta y_t + \gamma x_t + \varepsilon_{t+1}, \tag{16}$$

where y_t is either industrial production growth, Chicago Fed National activity index, yield on three-month Treasury bill rate, and Default yield spread (the difference between the yield on Moody's Baa-rated bond and Aaa-rated bond), and x_t is a predictor. We generate out-sample forecasts of y_t using the same recursive estimation procedure employed earlier and use the historical average as the benchmark model. Statistical significance of R^2_{OOS} is tested using the MSFE-adjusted statistic of Clark and West (2007).

Panel A of Table 9 reports results for the individual predictors. The results show that many of the individual predictors fail to outperform the HA benchmark across the four macroeconomic activity variables with mostly negative R^2_{OOS}.

The results for the combination forecasts are reported in Panel B of Table 9. We can see that almost all the R^2_{OOS} values are positive and statistically significant at the 1% level. These results mirror those reported for the combination models in Tables 4.

Taken together, these findings provide another explanation, in addition to structural instability, for the poor performance of the individual predictive model forecasts and the gains associated with the combination forecasts which deal with model uncertainty and structural instability.

[Insert Tables 9 about here]

6. Can we Improve Forecasts by Monitoring Performance?

We now implement the conditional forecasting approach of Giacomini and White (2006) who develop a framework for out-of-sample predictability testing and forecast selection when the forecasting model is subject to misspecification. Their framework aids forecast selection by linking them to instruments that tell us something about current economic conditions, which should lead to improved forecasting performance.

Timmermann and Zhu (2017) extend the work of Giacomini and White (2006) and develop conditions under which the expected predictive accuracy of a set of competing forecasts can be ranked conditionally based on a set of monitoring instruments. They characterize properties that monitoring instruments should possess and show that these reflect both the accuracy of the predictors used and the strength of the monitoring instruments. Timmermann and Zhu (2017) further show that in an environment with weak predictors, selecting between a benchmark forecast and an alternative forecast based on instruments that track their forecasting performance overtime should leading to better forecasts than relying solely a single forecast.

Let $\hat{r}_{1,t+1}$ and $\hat{r}_{2,t+1}$ be two individual one-step ahead forecast of r_{t+1} generated using information up to time t. Define the square error loss

$$L(\hat{r}_{t+1}, r_{t+1}) = (r_{t+1} - \hat{r}_{t+1})^2. \tag{17}$$

Under square error loss, the loss differential between the two forecasts is

$$\triangle L_{t+1} = e_{1,t+1}^2 - e_{2,t+1}^2, \tag{18}$$

where $e_{j,t+1} = r_{j,t+1} - \hat{r}_{j,t+1}$ for $j = 1, 2$ are individual forecast errors.

Following Giacomini and White (2006), the null hypothesis of conditional predictive ability is given by

$$H_0 : \mathbb{E}\left[\triangle L_{t+1} | Z_t\right] = 0, \tag{19}$$

where Z_t are monitoring instruments. We test this hypothesis using the linear regression

$$\triangle L_{t+1} = \theta_0 + \theta_1 z_t + \varepsilon_t, \tag{20}$$

where $\mathbb{E}[\varepsilon_t] = 0$, and $z_t \in Z_t$. Under the null of equal conditional predictive ability, $\theta = 0$ and $\theta_1 = 0$ in Equation (20). Non-zero values of $\theta_1 = 0$ suggests that the monitoring instrument, z_t, can help forecast differences in predictive accuracy across the two forecasts.

Using Equation (20), the expected future loss is given by $\mathbb{E}\left[\triangle L_{t+1}|Z_t\right] = \theta_0 + \theta_1 z_t$. Following Giacomini and White (2006), we consider a forecasting switching rule that chooses forecast 1 if $\mathbb{E}\left[\triangle L_{t+1}|Z_t\right] \leq 0$ or otherwise choose forecast 2:

$$r_{t+1}^{\text{SW}} = \hat{r}_{1,t+1}\mathbf{1}\{\mathbb{E}\left[\triangle L_{t+1}|Z_t\right] \leq 0\} + \hat{r}_{2,t+1}\mathbf{1}\{\mathbb{E}\left[\triangle L_{t+1}|Z_t\right] > 0\} \tag{21}$$

where $\mathbf{1}\{\mathbb{E}\left[\triangle L_{t+1}|Z_t\right] > 0\}$ is an indicator variable that takes the value one if the $r_{1,t+1}$ has the highest expected loss on $Z_t = z_t$ and zero otherwise. In our analysis, $r_{1,t+1}$ is always the HA return forecast and $r_{2,t+1}$ is either an predictive or combination forecast. As monitoring instruments, we consider growth in US consumer price index (CPI), US money stock (M2), a measure of macroeconomic uncertainty of Jurado et al. (2015), and the Chicago Board Options Exchange Volatility Index (VIX). The choice of the variables are motivated by the fact that they are drivers of commodity prices in general.

Table 10 reports results for the test of conditional predictability based on the four instruments. As shown in Panel A for the individual predictive model forecasts, the GW test fails to reject the null of equal predictive ability at the 5% level for all the monitoring instruments. However, there is some evidence that the monitoring instruments have predictive power for the future loss differential $\triangle L_{t+1}$ based on the t-statistic of the significance of θ_1 term in Equation (20).

The results for the combination forecasts are reported in Panel B of Table 9. We can see that the GW test rejects the null of conditional predictive ability for the M2 instrument whereas we fail to reject the null for the other monitoring instruments. Interestingly, the rejection of the null is driven solely by the significance of the information content of M2 as indicated by the significant slope coefficient, θ_1, in Equation (20). Based on this result, we use the M2 instrument to implement the forecast switching decision-rule in Equation (21).

[Insert Tables 10 about here]

Table 11 reports portfolio performance results for the forecasting switching rule. As can be seen from the table, monitoring forecasting performance results in only marginal improvement in portfolio performance for the individual predictive models. However, the individual forecasts still underperforms the HA return forecast as we continue to realize low and negative SR and utility gains. The forecasting strategy benefits the combination forecast leading to improved portfolio performance.

[Insert Tables 11 about here]

7. Conclusion

In this paper, we provide a comprehensive study on aggregate commodity return predictability using a large set of predictors predictors selected from the commodity return, stock return, bond return, and macroeconomic predictability literature. We consider both individual predictive regression models and forecast combination methods that pool information from a large set of predictors. We find that almost all of the individual predictive regression model forecasts fail to outperform the benchmark historical average return forecast except the OECD composite leading indicator and the business confidence index variables. Combination forecasts on the other hand perform substantially better than the historical average forecasts, and delivers statistically significant evidence of predictability. The superior forecasting performance of the combination forecasts can be attributed to their ability to provide insurance against model uncertainty and structural instability of the individual predictive models. Findings from a forecast breakdown test shows strong evidence of instability between excess commodity returns and the individual predictors, and provides an explanation for their inconsistent out-of-sample performance.

Commodity return predictability is also found to be countercyclical with predictability stronger during business-cycle recessions relative to expansions similar to the findings in studies such as Gargano and Timmermann (2014), Henkel et al. (2011), Rapach et al. (2010), and Lin et al. (2017) for commodity spot indexes, stocks and bond returns, respectively. We also show that the sources of predictability of combination forecasts for commodity returns has links to the real economy. Combination forecasts display predictive power for macroeconomic activity as measured by growth in industrial production, growth in consumer price index, changes in 3-month Treasury bill rate, changes in default spread, and changes in the Chicago Fed National activity index, and explain their significant out-of-sample performance as a result of picking up genuine variation in discount rates.

Economic significance, measured by Sharpe ratios and certainty equivalent return gains, of commodity return forecasts indicates that the statistically significant evidence of commodity return predictability translates into economic significance in an asset allocation exercise for a risk-averse investor. We find that recession is the main economic driver of these results. The investor realizes substantially high Sharpe ratios and certainty equivalent return gains in recessions compared to losses in expansions.

In an attempt to improve the forecasting performance of the individual models, we implement a forecast monitoring strategy that selects either the individual (combination) forecast or the HA benchmark based on whether a monitoring instrument such as US money stock predicts positively their square error loss differential. We show that whereas this strategy does not improve the individual models, it leads to marginal improvements in the performance of the combination forecasts both statistically and economically.

References

Ahmed, S., Tsvetanov, D., 2016. The predictive performance of commodity futures risk factors. Journal of Banking & Finance 71, 20–36.

Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? Journal of Applied Econometrics 25, 539–573.

Alquist, R., Kilian, L., Vigfusson, R. J., 2013. Forecasting the price of oil. In: *Handbook of economic forecasting*, Elsevier, vol. 2, pp. 427–507.

Bates, J. M., Granger, C. W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Baumeister, C., Kilian, L., 2012. Real-time forecasts of the real price of oil. Journal of Business & Economic Statistics 30, 326–336.

Baumeister, C., Kilian, L., 2014. What central bankers need to know about forecasting oil prices. International Economic Review 55, 869–889.

Baumeister, C., Kilian, L., 2015. Forecasting the real price of oil in a changing world: a forecast combination approach. Journal of Business & Economic Statistics 33, 338–351.

Baumeister, C., Kilian, L., 2016. A general approach to recovering market expectations from futures prices with an application to crude oil. CESifo Working Paper Series No. 5782. Available at SSRN: https://ssrn.com/abstract=2754033 .

Bessembinder, H., 1992. Systematic risk, hedging pressure, and risk premiums in futures markets. The Review of Financial Studies 5, 637–667.

Bessembinder, H., 1993. An empirical analysis of risk premia in futures markets. Journal of Futures Markets 13, 611–630.

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

Bjornson, B., Carter, C. A., 1997. New evidence on agricultural commodity return performance under time-varying risk. American Journal of Agricultural Economics 79, 918–930.

Brennan, M. J., 1958. The supply of storage. The American Economic Review 48, 50–72.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

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

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

Chinn, M. D., Coibion, O., 2014. The predictive content of commodity futures. Journal of Futures Markets 34, 607–636.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cochrane, J. H., 2008. Financial markets and the real economy. In: *Handbook of the Equity Risk Premium*, Elsevier, pp. 237–325.

Cochrane, J. H., 2011. Presidential address: Discount rates. The Journal of Finance 66, 1047–1108.

Cochrane, J. H., 2017. Macro-finance. Review of Finance 21, 945–985.

Della Corte, P., Sarno, L., Tsiakas, I., 2008. An economic evaluation of empirical exchange rate models. The Review of Financial Studies 22, 3491–3530.

DeRoon, F. A., Nijman, T. E., 2001. Testing for mean-variance spanning: a survey. Journal of empirical finance 8, 111–155.

Detzel, A., Strauss, J., 2017. Combination return forecasts and portfolio allocation with the cross-section of book-to-market ratios. Review of Finance 22, 1949–1973.

Edwards, F. R., Park, J. M., 1996. Do managed futures make good investments? Journal of Futures Markets: Futures, Options, and Other Derivative Products 16, 475–517.

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. Journal of Econometrics 177, 357–373.

Elliott, G., Gargano, A., Timmermann, A., 2015. Complete subset regressions with large-dimensional sets of predictors. Journal of Economic Dynamics and Control 54, 86–110.

Erb, C. B., Harvey, C. R., 2016. The strategic and tactical value of commodity futures. In: *The World Scientific Handbook Of Futures Markets*, World Scientific, pp. 125–178.

Fama, E. F., French, K. R., 1987. Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. Journal of Business pp. 55–73.

Fama, E. F., French, K. R., 1989. Business conditions and expected returns on stocks and bonds. Journal of financial economics 25, 23–49.

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, 2530–2548.

Gargano, A., Pettenuzzo, D., Timmermann, A., 2017. Bond return predictability: Economic value and links to the macroeconomy. Management Science .

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

Garratt, A., Lee, K., Pesaran, M. H., Shin, Y., 2003. Forecast uncertainties in macroeconomic modeling: An application to the uk economy. Journal of the American Statistical Association 98, 829–838.

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. The Review of Economic Studies 76, 669–705.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Gorton, G. B., Hayashi, F., Rouwenhorst, K. G., 2013. The fundamentals of commodity futures returns. Review of Finance pp. 35–105.

Groen, J. J., Pesenti, P. A., 2011. Commodity prices, commodity currencies, and global economic developments. In: *Commodity Prices and Markets, East Asia Seminar on Economics, Volume 20*, University of Chicago Press, pp. 15–42.

Hamilton, J. D., 2009. Causes and consequences of the oil shock of 2007–08. Brookings Papers on Economic Activity .

Hamilton, J. D., Wu, J. C., 2015. Effects of index-fund investing on commodity futures prices. International economic review 56, 187–205.

Hansen, P. R., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation. Working paper, European University Institute. Available at http://cadmus.eui.eu/handle/1814/21454 .

Henkel, S. J., Martin, J. S., Nardari, F., 2011. Time-varying short-horizon predictability. Journal of Financial Economics 99, 560–580.

Hicks, J. R., 1939. Value and capital. Oxford University Press.

Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? Journal of Financial Economics 105, 473–490.

Jensen, G. R., Johnson, R. R., Mercer, J. M., 2000. Efficient use of commodity futures in diversified portfolios. Journal of Futures Markets 20, 489–506.

Jones, C. M., Kaul, G., 1996. Oil and the stock markets. The Journal of Finance 51, 463–491.

Jurado, K., Ludvigson, S. C., Ng, S., 2015. Measuring uncertainty. The American Economic Review 105, 1177–1216.

Kaldor, N., 1939. Speculation and economic stability. The Review of Economic Studies 7, 1–27.

Keynes, J. M., 1930. A Treatise on Money: In 2 Vol. The Applied Theory of Money. Macmillan, London.

Kilian, L., 2009. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. The American Economic Review 99, 1053–1069.

Kilian, L., Murphy, D. P., 2014. The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Econometrics 29, 454–478.

Leitch, G., Tanner, J. E., 1991. Economic forecast evaluation: profits versus the conventional error measures. The American Economic Review pp. 580–590.

Lin, H., Wu, C., Zhou, G., 2017. Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. Management Science .

Neely, C. J., Rapach, D. E., Tu, J., Zhou, G., 2014. Forecasting the equity risk premium: the role of technical indicators. Management Science 60, 1772–1791.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Pagano, P., Pisani, M., 2009. Risk-adjusted forecasts of oil prices. The BE Journal of Macroeconomics 9.

Pesaran, M. H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. The Journal of Finance 50, 1201–1228.

Potì, V., 2018. A new tight and general bound on return predictability. Economics Letters 162, 140–145.

Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121, 46–65.

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, 821–862.

Rapach, D. E., Zhou, G., 2013. Forecasting stock returns. Handbook of Economic Forecasting 2, 328–383.

Sadorsky, P., 1999. Oil price shocks and stock market activity. Energy economics 21, 449–469.

Sadorsky, P., 2002. Time-varying risk premiums in petroleum futures prices. Energy Economics 24, 539–556.

Sarno, L., Schneider, P., Wagner, C., 2016. The economic value of predicting bond risk premia. Journal of Empirical Finance 37, 247–267.

Shibata, R., 1976. Selection of the order of an autoregressive model by akaike's information criterion. Biometrika 63, 117–126.

Stock, J. H., Watson, M. W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335.

Stock, J. H., Watson, M. W., 2002a. Forecasting using principal components from a large number of predictors. Journal of the American statistical association 97, 1167–1179.

Stock, J. H., Watson, M. W., 2002b. Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20, 147–162.

Stock, J. H., Watson, M. W., 2003. Forecasting output and inflation: The role of asset prices. Journal of Economic Literature 41, 788–829.

Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23, 405–430.

Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financial Analysts Journal 68, 54–74.

Thornton, D. L., Valente, G., 2012. Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. The Review of Financial Studies 25, 3141–3168.

Timmermann, A., 2006. Forecast combinations. Handbook of economic forecasting 1, 135–196.

Timmermann, A., 2008. Elusive return predictability. International Journal of Forecasting 24, 1–18.

Timmermann, A., Zhu, Y., 2017. Monitoring forecasting performance. UCSD working paper .

Wachter, J. A., 2006. A consumption-based model of the term structure of interest rates. Journal of Financial Economics 79, 365–399.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.

Ye, M., Zyren, J., Shore, J., 2005. A monthly crude oil spot price forecasting model using relative inventories. International Journal of Forecasting 21, 491–501.

Table 1: Monthly Predictor Variables for Commodity Futures Returns

Predictor	Article(s)	Variable definition and motivation for their consideration
Basis	Fama and French (1987), Hong and Yogo (2012), Gorton et al. (2013).	To construct our monthly aggregate measure of commodity basis, we first collect futures prices of 32 individual commodities. Most of these individual commodities make up the constituents of the S&P GSCI. We then compute the basis for each individual commodity futures as the difference in log prices between two nearest-to-maturity futures prices: $$\text{Basis}_t^i = \frac{\log(f_t^{i,T_1}) - \log(f_t^{i,T_2})}{T_2 - T_1},$$ where f_t^{i,T_1} and f_t^{i,T_1} are the nearby and next-to-nearby futures prices of commodity i, respectively. We then compute the mean basis across commodities for each commodity sector, namely agriculture, energy, livestock, and metals. Finally, the aggregate basis variable is computed as an equally weighted average of the basis across the four commodity sectors similarly to Hong and Yogo (2012). Details of the 32 individual commodities futures are provided in Table A1 of the appendix. The consideration of the basis is motivated by the theory of storage of Brennan (1958) which posits that the benefit of holding the physical commodity (convenience yield) should decline with rising inventory levels. The convenience yield is therefore closely linked to basis since it is benefit that accrues to inventory holders and not to holders of futures contract. The information content of basis could be used as a signal for inventories since commodities with low inventories have higher basis which means higher prior futures prices. As such, basis should be important for forecasting commodity returns.
Log growth of global crude oil production ($PROD$)	Groen and Pesenti (2011), Baumeister and Kilian (2012), Baumeister and Kilian (2014), Baumeister and Kilian (2015)	Log growth in global crude oil production is calculated as log(global crude oil production(t))$-log$(global crude oil production($t-1$)). Data on global crude oil production is downloaded from the database of the U.S. Energy Information Administration. Supply is one the most important determinants of crude oil prices. For example, if crude oil production should drop whiles demand remains constant, prices would be pushed upwards. Considering that energy commodities, and more especially crude oil, are heavily weighted in the S&P GSCI, crude oil production should affect the overall price level of the index. These motivate our consideration of this variables.
Log growth of global crude oil inventory (INV)	Ye, Zyren, and Shore (2005), Groen and Pesenti (2011), Gorton et al. (2013), Kilian and Murphy (2014)	Log growth of global crude oil inventory is defined as log(global crude oil inventory(t))$-log$(global crude oil inventory($t-1$)). The inventory data used in calculating this variables is constructed by multiplying U.S. crude oil inventories by the ratio of OECD petroleum inventories to U.S. petroleum inventories. Petroleum inventories are defined to include both stocks of crude oil and stocks of refined products. te consideration of this variable is motivated by the theories of storage and normal backwardation of Brennan (1958) and Keynes (1930) which posit that the fundamental determinants of expected commodity returns is inventory. For example, rising crude oil inventories should signal speculative demand in the commodities market. Speculators receive compensation for taking long positions since commodity producers hedge the future spot price by taking short positions in the futures market. Also, since the S&P GSCI is more heavily weighted towards energy commodities, and more especially crude oil, we should expect the level of crude oil inventory to partly drive movements in the returns of the index.
Log dividend-price ratio (DP)	Bessembinder and Chan (1992), Gargano and Timmermann (2014)	Log dividend-price ratio is the difference between the log of the 12-month moving sum of the dividends paid on the S&P 500 index and the log price of the S&P 500 index. The consideration of this variable as a predictor for commodity returns is motivated by studies such as Tang and Xiong (2012) and Hamilton and Wu (2015) who show that the commodities market has become more integrated with the stock and bond markets. As such, state variables that drive stock and bond returns should partly be responsible for movements in commodity returns.

Table 1: *continued*

Predictor	Article(s)	Variable definition and motivation for their consideration
S&P 500 index return (*SP500*)	Jones and Kaul (1996), Sadorsky (1999), DeRoon and Nijman (2001)	*SP500* is the log return on the S&P 500 computed as log(S&P 500 (t))-log(S&P 500 (t-1)). S&P 500 is the price level of the S&P 500 stock market index. Jones and Kaul (1996) and Sadorsky (1999) find that the stock market and oil prices tend to move together in the same direction as a response to global aggregate demand factors. Shifts in aggregate demand should therefore influence both corporate profits and the demand for oil. The S&P GSCI is heavily weighted towards energy commodities, particularly crude oil. We should expect the S&P index returns to drive movements in commodity returns. These motivate our consideration of this variable as a predictor for commodity returns.
3-month Treasury bill rate (*TBL*)	Bessembinder and Chan (1992), Sadorsky (2002) Bessembinder (1992), Bjornson and Carter (1997), Hong and Yogo (2012), Gargano and Timmermann (2014)	*TBL* is the yield on U.S. 3-month Treasury bill (secondary market). The following are the motivations for considering this variable. According to the theory of storage, interest rate determines the storage cost of storable commodities. For example, a commodity market participant's expectation of the futures price of a storable commodity will depend on prevailing interest rate and the cost of storage if borrowed funds are used to purchase the commodity. The *TBL* is also negatively correlated with the business-cycle; expected returns are high when business conditions are weak and low when business conditions are strong. If assume market integration, then we should also expect the same variable known to predict stock and bond returns to forecast commodity returns. Again the monetary policy regime of the U.S. could impact commodity prices through currency valuation and interest rates.
Change in 3-month T-bill rate (*CTBL*)	Bessembinder and Chan (1992), Bessembinder (1992), Bessembinder (1993), Hong and Yogo (2012)	*CTBL* is defined as *TBL* (t) - *TBL* (t-1). Similar to the motivations given for considering the 3-month T-bill rate as a candidate predictor, changes in the T-bill rate is also an economic activity variable and therefore tracks changes in business condition.
Long term return (*LTR*)	Gargano and Timmermann (2014)	*LTR* is the return on long-term government bonds. The same motivations stated for considering the log dividend-price ratio predictor applies to the long term return.
Term spread (*TMS*)	Bessembinder and Chan (1992), Bessembinder (1993), Groen and Pesenti (2011), Gargano and Timmermann (2014)	The *TMS* is defined as long term government bond yield minus the yield of T-bills. The *TMS* is an economic activity variable and therefore tracks changes in business condition. It is known to predict returns on stocks and bonds (Fama and French (1989)), and negatively correlated with the business-cycle: expected returns are high when business conditions are weak and low when business conditions are strong. If one assumes that the commodities market is integrated with the stock and bond markets, then we should also expect the term spread to forecast commodity returns. These reasons motivate our consideration of this variable as a predictor for commodity returns.
Change in term spread (*CTMS*)	Bessembinder (1992), Bessembinder (1993)	*CTMS* is defined as *TMS*(t) - *TMS*(t-1). Similar to the motivations given for considering the the term spread as a predictor for commodity returns, changes in the term spread is also an economic activity variable and therefore should tracks changes in business condition.
Yield spread (*YS*)	Fama and French (1989), Bessembinder and Chan (1992), Hong and Yogo (2012)	The yield spread is defined as the yield on Aaa-rated bond minus the yield on the 3-month treasury bill rate. Our consideration of the *YS* is motivated by the fact it is an economic activity variable and therefore should track changes in business condition. It is negatively correlated with the business-cycle (Hong and Yogo (2012)) and therefore we should expect the returns on commodities to be high when business conditions are weak and low when business conditions are strong.
Change in default premium (*CDFP*)	Bessembinder (1992)	Change in default premium is defined as yield on Baa-rated bond minus yield on long-term government bond. What motivates the use of this variable as a predictor for commodity returns is that it is an economic activity variable and therefore tracks changes in business condition. It is also negatively correlated with the business-cycle (Fama and French, 1989). We should therefore expect commodity returns on commodities to be high when business conditions are weak and low when business conditions are strong.
Default return spread (*DFR*)	Bessembinder and Chan (1992), Gargano and Timmermann (2014)	*DFR* is defined as long-term corporate bond returns minus long-term government bond returns. The motivation for considering this variable is the same as the motivation given for considering the log dividend-price ratio predictor variable.

Table 1: *continued*

Predictor	Paper(s)	Variable definition and motivation for their consideration
Inflation (*INFL*)	Bessembinder (1993), Groen and Pesenti (2011), Gargano and Timmermann (2014)	*INFL* is defined as the log growth in U.S. consumer price index. The following motivates its consideration as predictor for commodity returns. It is an economic activity variable and therefore tracks changes in business condition, and also signal fluctuations in economic activity. It is negatively correlated with the business-cycle (Hong and Yogo (2012)). We should expect commodity returns on commodities to be high when business conditions are weak and low when business conditions are strong. Commodity futures prices are also of interest to central banks and policymakers because they provide forecasts for key commodities, and play an important role in explaining fluctuations in and projecting macroeconomic activity.
Money stock (*M1*)	Groen and Pesenti (2011), Gargano and Timmermann (2014)	*M1* is the log growth in log growth in monthly M1 money stock. The motivation for considering this variable is same as the motivations given for considering the log dividend-price ratio predictor.
Unemployment rate (*UNRATE*)	Groen and Pesenti (2011), Gargano and Timmermann (2014)	*UNRATE* is the monthly unemployment rate from the website of the Archival Federal Reserve Bank of St. Louis Economic Data (ALFRED). As measure of economic activity, *UNRATE* variables also signal fluctuations in economic activity similar to given for inflation, the term spread, among others.
Log industrial production (*INDPRO*)	Bessembinder (1993), Bjornson and Carter (1997), Pagano and Pisani (2009), Groen and Pesenti (2011), Gargano and Timmermann (2014)	*INDPRO*) is the monthly log growth in OECD aggregate industrial production obtained from OECD data website, `https://data.oecd.org/`. As a measure of economic activity, *INDPRO* also signal fluctuations in economic activity similar to, for example, the inflation rate, unemployment rate, the term spread predictors.
Log degree of capacity utilization in U.S. manufacturing (*CUTIL*)	Pagano and Pisani (2009), Baumeister and Kilian (2016)	*CUTIL* is the log growth in degree of capacity utilization in U.S. manufacturing. As a measure of economic activity, *CUTIL* also signal fluctuations in economic activity similar to, for example, the inflation rate, unemployment rate, industrial production, the term spread predictors.
Global real economic activity index (*REA*)	Alquist, Kilian, and Vigfusson (2013), Baumeister and Kilian (2014)	The global real activity index is constructed from data on global dry cargo ocean shipping freight rates as described in Kilian (2009). The reason that motivates its consideration as a predictor is that global economic activity drives demand for oil and other industrial commodities in global markets and has has been shown to forecast movement in crude oil returns. This variable is based on dry cargo single voyage ocean freight rates and is explicitly designed to capture shifts in the demand for industrial commodities in global business markets. It exploits the positive correlation between ocean freights rate and economic activity. Commodities are traded globally as such the state of the global economy will partly impact movements in commodity prices.
Chicago Fed National Activity index (*CFNAI*)	Hong and Yogo (2012)	The *CFNAI* is a monthly summary statistic for U.S. economic growth. As a measure of economic activity, the index is designed to gauge overall economic activity and related inflationary pressure. The motivation for considering this variable is that commodity prices form a key component of forming expectations of inflation. High economic activity is also negatively correlated with inflation (Stock and Watson, 1999). Therefore we should expect the index to drive movement in commodity prices.
OECD composite leading indicator (*CLI*), business confidence index (*BCI*), consumer confidence index (*CCI*)	Pagano and Pisani (2009), Groen and Pesenti (2011)	These variables are measures of global economic activity similar to the global index of real economic activity. They are designed to provide signals of turning points in the business cycle and fluctuations in economic activity. This motivates their consideration as predictors for commodity returns.
Commodity currencies: Australia (*AUS*), Canada (*CAN*), New Zealand (*NZ*), South Africa (*SA*) & India (*IND*)	Chen et al. (2010), Gargano and Timmermann (2014), Groen and Pesenti (2011)	These predictors are motivated by the study of Chen et al. (2010) who exploit the notion that changes in commodity currencies are correlated with commodity prices. These countries are major commodity exporters where commodities represent a quarter to one-half of their total export earnings, and also have a sufficiently long history of market-based floating exchange rates. Therefore movements in their exchange rate against the US dollar should be informative future commodity returns.

Notes. This table outlines and defines the predictors that we use, the motivation for their use, and the relevant prior commodity return studies.

Table 2: Summary Statistics for Returns and Predictor Variables

Variable	Obs	Mean	Standard deviation	Min	Max	Auto correlation	Sharpe ratio
Panel A: Index							
S&P GSCI	491	0.143	5.57	−28.29	22.31	0.16	0.089
Panel B: Predictor Variables							
Panel B1: Predictors from the commodity predictability literature							
Basis	491	0.52	1.05	−3.83	4.04	0.70	
INV	491	100.70	4.98	86.40	120.95	0.83	
PROD	491	0.09	1.46	−9.49	6.50	−0.07	
Panel B2: Predictors from the equity and bond risk premium predictability literature							
DP	491	−365.60	43.90	−452.36	−275.33	0.99	
SP500	491	0.63	4.30	−24.54	12.38	0.04	
TBL	491	4.68	3.58	0.01	16.30	0.99	
CTBL	491	−0.01	0.46	−4.62	2.61	0.36	
LTR	491	0.73	3.19	−11.24	15.23	0.05	
TMS	491	2.21	1.45	−3.65	4.55	0.95	
CTMS	491	0.00	0.47	−3.28	4.23	0.10	
YS	491	2.99	1.52	−2.28	5.93	0.97	
CDFP	491	0.00	0.30	−1.20	1.39	−0.12	
DFR	491	0.00	1.48	−9.75	7.37	−0.03	
INFL	491	0.30	0.37	−1.92	1.52	0.62	
Panel B3: Predictors related to economic activity							
M1	491	0.49	0.87	−3.20	4.93	0.12	
UNRATE	491	−0.73	17.23	−70.00	60.00	0.12	
INDPRO	491	0.17	0.61	−3.98	2.01	0.27	
CUTIL	491	0.00	0.76	−3.55	2.53	0.28	
REA	491	−0.02	55.19	−163.74	187.66	0.96	
CFNAI	491	−3.51	92.67	−466.00	273.00	0.62	
CLI	491	0.00	0.15	−0.78	0.60	0.96	
BCI	491	0.00	0.16	−0.85	0.52	0.88	
CCI	491	0.00	0.13	−0.44	0.45	0.82	
Panel B4: Exchange rates of major commodity exporting countries							
AUS	491	−0.11	3.30	−18.68	9.92	0.03	
CAN	491	−0.06	2.00	−13.03	8.85	−0.06	
NZ	491	−0.08	3.49	−24.89	18.01	−0.03	
SA	491	−0.56	4.22	−24.82	14.05	0.02	
IND	491	−0.41	2.11	−19.89	7.05	0.05	

Notes. This table reports the summary statistics of the returns on the S&P GSCI and the 31 predictor variables. We report the number of observations (Obs), the mean, standard deviation, minimum and maximum values, first-order autocorrelation and the annualized Sharpe ratio. All values are in percent. The sample period is from February 1976 to December 2016.

Table 3: In-sample Evaluation of Commodity Return Predictability

Predictor Variable	Slope Coefficient	t-stats	R^2 (%)	R^2_{EXP} (%)	R^2_{REC} (%)
Basis	−0.12	−0.49	0.05	−0.12	0.47
INV	−0.07	−1.27	0.43	1.12	−1.19
PROD	−0.04	−0.22	0.01	−0.04	0.13
DP	0.00	−0.12	0.00	0.02	−0.03
SP500	0.02	0.26	0.02	−0.28	0.75
TBL	0.02	0.33	0.02	−0.05	0.19
CTBL	0.27	0.47	0.05	0.02	0.13
LTR	−0.22	−2.44**	1.52	0.91	2.94
TMS	0.02	0.10	0.00	0.03	−0.06
CTMS	0.70	1.33	0.34	0.41	0.19
YS	−0.12	−0.68	0.11	−0.21	0.86
CDFP	−3.22	−3.56***	2.97	−0.04	10.07
DFR	0.79	2.61***	4.30	0.15	14.12
INFL	0.54	0.63	0.13	−0.36	1.27
M1	−0.45	−1.33	0.49	−0.55	2.94
UNRATE	−0.02	−0.97	0.26	−0.37	1.74
INDPRO	1.01	1.96**	1.20	−0.69	5.67
CUTIL	1.05	2.18**	2.04	−0.67	8.46
REA	0.00	0.82	0.22	0.36	−0.10
CFNAI	0.01	2.34**	2.29	0.45	6.62
CLI	7.00	3.33***	3.69	0.07	12.23
BCI	9.06	3.81***	6.35	1.17	18.57
CCI	4.91	1.86*	1.22	0.55	2.79
AUS	0.08	0.77	0.21	−0.93	2.91
CAN	0.08	0.51	0.09	−0.44	1.34
NZ	0.07	0.76	0.16	−0.60	1.96
SA	0.05	0.70	0.14	−0.52	1.70
IND	0.19	1.37	0.53	−0.38	2.69

Notes. This table reports the in-sample estimation results for the bivariate predictive regression model of log commodity excess returns and the predictor variables individually. The immediate right of slope coefficients report the Newey and West (1987) heteroskedasticity-consistent t-statistics. The R^2 statistics are computed for the full sample period 1976:02-2016:12. The R^2_{EXP} (%) (R^2_{REC} (%)) statistics in the last two columns are computed separately for the National Bureau of Economic Research (NBER)-dated business cycle expansions (recessions). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Statistical Evaluation of Commodity Return Forecasts

Predictor	MSFE	R^2_{OOS} (%)	MSFE-adjusted	Predictor	MSFE	R^2_{OOS} (%)	MSFE-adjusted
HA	38.46			HA	38.46		
Panel A: Individual predictive forecasts				Panel B: Combination forecasts			
Basis	38.69	−0.59	−0.53	Mean	38.00	1.19	2.51***
INV	38.44	0.05	0.50	Median	38.34	0.30	2.20**
PROD	38.57	−0.28	−1.20	Trimmed mean	38.07	1.01	2.52***
DP	38.72	−0.68	−0.91	Weighted mean	38.00	1.21	2.51***
SP500	38.72	−0.67	−0.23	DMSFE, 0.9	37.97	1.26	2.49***
TBL	38.95	−1.27	−1.33	DMSFE, 0.7	37.96	1.30	2.48***
CTBL	38.47	−0.03	−0.47	ABMA	38.01	1.16	2.51***
LTR	38.03	1.13	1.58*	Subset (k=2)	37.66	2.08	2.50***
TMS	38.57	−0.27	−0.85	Subset (k=3)	37.40	2.75	2.49***
CTMS	38.50	−0.10	0.05	Subset (k=4)	37.20	3.27	2.49***
YS	38.52	−0.15	−0.61	Subset (k=5)	37.06	3.64	2.48***
CDFP	37.52	2.44	2.07**	Subset (k=6)	36.96	3.90	2.46***
DFR	37.25	3.15	1.83**	Subset (k=7)	36.90	4.07	2.44***
INFL	38.63	−0.44	−0.67	PC (ic=AIC)	36.64	4.73	2.49***
M1	38.43	0.07	0.74	PC (ic=BIC)	37.01	3.76	2.29**
UNRATE	38.55	−0.24	−0.09	PC (ic=R2)	36.55	4.96	2.50***
INDPRO	38.23	0.61	1.25				
CUTIL	37.73	1.89	1.85**				
REA	38.68	−0.57	−0.48				
CFNAI	37.66	2.08	1.56*				
CLI	37.00	3.80	2.43***				
BCI	35.80	6.92	2.86***				
CCI	38.26	0.52	1.06				
AUS	38.62	−0.42	−0.42				
CAN	38.76	−0.77	−0.93				
NZ	38.63	−0.45	−0.35				
SA	38.65	−0.49	−0.66				
IND	38.37	0.23	0.90				

Notes. This table reports out-of-sample results for the individual and combination forecasts of log excess commodity returns. HA is the historical average benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OOS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the HA forecast. Statistical significance for the R^2_{OOS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the HA forecast MSFE is less than or equal to the competing forecast MSFE against the alternative hypothesis that the HA forecast MSFE is greater than or equal to the competing forecast MSFE. Results are reported for the full out-of-sample forecast evaluation period 1990:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Results of Forecast Breakdown Test

Predictor	t-stats	p-value	Predictor	t-stats	p-value
Basis	3.23	0.001	M1	3.07	0.002
INV	3.12	0.002	UNRATE	3.11	0.002
PROD	3.09	0.002	INDPRO	3.13	0.002
DP	3.15	0.002	CUTIL	3.11	0.002
SP500	3.00	0.003	REA	3.07	0.002
TBL	3.15	0.002	CFNAI	3.11	0.002
CTBL	3.10	0.002	CLI	3.10	0.002
LTR	3.04	0.002	BCI	3.21	0.001
TMS	3.10	0.002	CCI	3.25	0.001
CTMS	3.08	0.002	AUS	3.09	0.002
YS	3.09	0.002	CAN	3.14	0.002
CDFP	3.03	0.002	NZ	3.17	0.002
DFR	3.33	0.001	SA	3.24	0.001
INFL	3.07	0.002	IND	3.08	0.002

Notes. This table reports the t-statistics and associated p-values for the forecast breakdown tests of Giacomini and Rossi (2009) using a quadratic loss function. Similarly to our out-of-sample forecasting tests, we use a recursive window estimation approach where the step-ahead forecast starts in January 1990 till the end of the sample December 2016. p-values lower than 0.1,0.05 and 0.01 denotes significance at the 10%, 5% and 1%, respectively.

Table 6: Economic Evaluation of Commodity Return Forecasts

Strategy	μ_p	σ_p	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}
HA benchmark	0.002	0.09	0.02	0.02				
				Panel A: Individual predictive forecasts				
Basis	−0.006	0.13	−0.05	−0.08	−1.97	−2.40	12	—
INV	−0.001	0.11	−0.01	−0.05	−0.82	−1.18	10	—
PROD	−0.011	0.08	−0.14	−0.16	−1.04	−1.23	6	—
DP	0.002	0.14	0.01	0.00	−1.93	−2.05	4	—
SP500	0.013	0.15	0.09	0.02	−1.23	−2.27	29	—
TBL	−0.003	0.16	−0.02	−0.02	−2.93	−2.98	2	—
CTBL	0.000	0.09	0.00	−0.01	−0.14	−0.18	2	—
LTR	0.033	0.15	0.22	0.12	0.73	−0.66	37	45
TMS	−0.003	0.09	−0.04	−0.05	−0.51	−0.55	2	—
CTMS	0.002	0.12	0.02	−0.03	−0.79	−1.36	16	—
YS	−0.001	0.09	−0.01	−0.02	−0.27	−0.31	2	—
CDFP	0.064	0.16	0.40	0.30	3.65	2.07	43	77
DFR	0.101	0.28	0.37	0.30	−0.24	−2.06	49	—
INFL	−0.002	0.11	−0.02	−0.05	−0.99	−1.24	8	—
M1	0.011	0.13	0.09	0.03	−0.43	−1.19	21	—
UNRATE	−0.009	0.11	−0.08	−0.13	−1.70	−2.19	14	—
INDPRO	0.015	0.12	0.13	0.05	0.34	−0.54	24	30
CUTIL	0.033	0.12	0.28	0.22	2.17	1.57	17	103
REA	−0.003	0.13	−0.02	−0.04	−1.91	−2.09	6	—
CFNAI	0.035	0.14	0.24	0.20	1.44	0.84	17	109
CLI	0.097	0.18	0.54	0.52	5.83	5.59	8	689
BCI	0.123	0.21	0.59	0.56	6.74	6.35	12	557
CCI	−0.006	0.11	−0.05	−0.09	−1.34	−1.73	11	—
AUS	−0.004	0.09	−0.04	−0.10	−0.66	−1.16	14	—
CAN	−0.012	0.09	−0.12	−0.18	−1.51	−2.03	14	—
NZ	0.005	0.12	0.04	−0.03	−0.65	−1.41	21	—
SA	−0.003	0.10	−0.03	−0.09	−0.72	−1.26	15	—
IND	0.006	0.09	0.07	0.02	0.40	0.07	10	23
				Panel B: Combination forecasts				
Mean	0.019	0.09	0.21	0.18	1.66	1.46	6	106
Median	0.006	0.09	0.06	0.05	0.44	0.35	3	22
Trimmed mean	0.017	0.09	0.19	0.16	1.46	1.28	6	93
Weighted mean	0.020	0.09	0.21	0.19	1.69	1.48	6	108
DMSFE ($\theta = 0.9$)	0.020	0.09	0.22	0.19	1.75	1.54	6	112
DMSFE ($\theta = 0.7$)	0.021	0.09	0.23	0.20	1.82	1.61	7	117
ABMA	0.019	0.09	0.21	0.18	1.64	1.44	6	104
Subset (k = 2)	0.033	0.11	0.31	0.27	2.60	2.21	11	192
Subset (k = 3)	0.046	0.13	0.36	0.31	3.16	2.61	16	267
Subset (k = 4)	0.054	0.14	0.38	0.33	3.46	2.75	19	322
Subset (k = 5)	0.062	0.16	0.39	0.34	3.50	2.66	23	367
Subset (k = 6)	0.068	0.17	0.39	0.34	3.35	2.39	26	405
Subset (k = 7)	0.073	0.19	0.39	0.33	3.07	1.99	29	436
PC (IC = AIC)	0.106	0.25	0.43	0.35	2.56	0.73	49	636
PC (IC = BIC)	0.080	0.21	0.39	0.35	2.59	1.80	22	480
PC (IC = R^2)	0.115	0.25	0.45	0.38	2.91	1.03	50	691

Notes. This table reports portfolio performance results for a mean-variance investor with relative risk aversion of three who monthly allocates his wealth between commodities and risk-free T-bills using either the HA benchmark forecast (static portfolio strategy) or the individual predictive regression (combination) forecasts (dynamic portfolio strategy). The forecasts in Panel A are based on one each of the 28 predictor variables. The forecasts in Panel B are based on 28 predictors using the different combination methods outlined in Section 2.1.2. For each portfolio strategy, we report the annualized mean realized return (μ_p), annualized realized volatility (σ_p), annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized utility gain (net of cost), $\triangle$ ($\triangle_\tau$), the portfolio management fee that the investor would be willing to pay in order to have access to the dynamic strategy relative to the static strategy, the turnover ratio (TO) ratio, the ratio of the average turnover of the dynamic strategy relative to that of the static strategy, and the break-even transaction costs, τ^{BE}, that will render the investor indifferent between the dynamic and static portfolio strategies. We set proportional transaction costs of 20bps per dollar of trading. Since we use commodity futures, we avoid short sales restrictions but limit leverage to 50% of wealth to avoid excessive risk taking. Results are reported for the full out-of-sample forecast evaluation period 1996:01-2016:12.

Table 7: Statistical Evaluation of Commodity Return Forecasts during Business-cycles

	Expansion			Recession		
Predictor	MSFE	R^2_{OOS} (%)	MSFE-adjusted	MSFE	R^2_{OOS} (%)	MSFE-adjusted
HA	29.32			109.37		
Panel A: Individual predictive model forecasts						
Basis	29.65	−1.14	−1.11	108.75	0.56	0.72
INV	29.32	0.01	0.41	109.23	0.13	0.29
PROD	29.34	−0.07	−0.49	110.17	−0.73	−1.10
DP	29.66	−1.15	−1.15	109.04	0.30	0.89
SP500	29.15	0.57	1.25	112.93	−3.25	−1.32
TBL	29.55	−0.80	−0.39	111.82	−2.24	−1.54
CTBL	29.35	−0.12	−1.81	109.20	0.16	1.27
LTR	29.36	−0.14	0.70	105.24	3.78	1.55*
TMS	29.32	0.00	0.24	110.30	−0.85	−1.81
CTMS	29.36	−0.15	−0.02	109.36	0.01	0.08
YS	29.32	−0.02	0.06	109.84	−0.43	−0.89
CDFP	29.43	−0.37	0.54	100.31	8.28	2.12**
DFR	29.18	0.48	1.56*	99.86	8.70	1.32
INFL	29.46	−0.48	−0.92	109.76	−0.36	−0.21
M1	29.76	−1.50	−0.62	105.71	3.35	2.18**
UNRATE	29.47	−0.52	−0.94	109.00	0.34	0.37
INDPRO	29.30	0.06	0.58	107.47	1.74	1.10
CUTIL	29.50	−0.61	−0.02	101.61	7.10	2.14**
REA	29.44	−0.40	−0.22	110.40	−0.94	−0.44
CFNAI	29.42	−0.35	−0.10	101.58	7.12	1.64*
CLI	29.41	−0.32	1.28	95.86	12.35	2.21**
BCI	29.06	0.89	1.94**	88.08	19.46	2.50**
CCI	29.32	0.00	0.63	107.63	1.59	0.85
AUS	29.42	−0.36	−0.27	109.97	−0.55	−0.32
CAN	29.48	−0.55	−0.47	110.73	−1.24	−0.85
NZ	29.80	−1.64	−2.08	107.16	2.02	2.05**
SA	29.38	−0.20	0.05	110.57	−1.09	−1.24
IND	29.31	0.03	0.49	108.67	0.64	0.88
Panel B: Combination forecasts						
Mean	29.26	0.21	0.88	105.85	3.22	2.49**
Median	29.29	0.10	0.82	108.57	0.73	2.26**
Trimmed mean	29.27	0.15	0.74	106.32	2.79	2.58***
Weighted mean	29.26	0.21	0.88	105.79	3.27	2.50**
DMSFE ($\theta = 0.9$)	29.26	0.20	0.83	105.56	3.48	2.48**
DMSFE ($\theta = 0.7$)	29.26	0.20	0.82	105.43	3.61	2.48**
ABMA	29.26	0.20	0.87	105.91	3.16	2.49***
Subset (k = 2)	29.24	0.28	0.85	103.00	5.82	2.50**
Subset (k = 3)	29.24	0.25	0.84	100.67	7.95	2.50**
Subset (k = 4)	29.27	0.17	0.82	98.75	9.71	2.49**
Subset (k = 5)	29.31	0.03	0.81	97.19	11.14	2.49**
Subset (k = 6)	29.36	−0.14	0.78	95.94	12.28	2.48**
Subset (k = 7)	29.42	−0.34	0.76	94.89	13.24	2.47**
PC (IC = AIC)	29.71	−1.32	0.86	90.43	17.32	2.48**
PC (IC = BIC)	29.73	−1.39	0.50	93.53	14.49	2.40**
PC (IC = R^2)	29.73	−1.40	0.85	89.49	18.18	2.50**

Notes. This table reports out-of-sample results for the individual and combination forecasts of log excess commodity returns using the NBER-dated recession indicator. HA is the historical average benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OOS} statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the HA forecast. Statistical significance for the R^2_{OOS} statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the HA forecast MSFE is less than or equal to the competing forecast MSFE against the alternative hypothesis that the HA forecast MSFE is greater than or equal to the competing forecast MSFE. Results are reported for the full out-of-sample evaluation period 1990:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Economic Evaluation of Commodity Return Forecasts during the Business-cycle

	Expansion						Recession					
Strategy	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}
HA benchmark	0.01	0.01					0.06	0.06				
Panel A: Individual predictive model forecasts												
Basis	−0.12	−0.16	−2.31	−2.74	11	—	0.22	0.20	0.59	0.16	8	168
INV	−0.03	−0.07	−0.97	−1.32	9	—	0.09	0.06	0.33	−0.15	9	18
PROD	−0.01	−0.05	−0.08	−0.26	5	—	−0.65	−0.67	−8.37	−8.61	5	—
DP	−0.05	−0.06	−2.33	−2.47	4	—	0.28	0.27	1.18	1.09	3	1064
SP500	0.28	0.18	1.94	0.92	25	57	−0.34	−0.38	−25.90	−27.04	19	—
TBL	0.02	0.01	−1.47	−1.52	2	—	−0.14	−0.15	−14.41	−14.54	3	—
CTBL	−0.03	−0.04	−0.22	−0.26	2	—	0.09	0.09	0.44	0.41	1	342
LTR	0.07	−0.04	−0.86	−2.24	34	—	0.79	0.74	13.38	12.02	22	312
TMS	0.01	0.00	0.04	0.00	2	3	−0.22	−0.23	−4.72	−4.77	2	—
CTMS	−0.04	−0.11	−0.55	−1.11	14	—	0.17	0.14	−2.76	−3.35	11	—
YS	0.00	−0.01	−0.05	−0.08	2	—	−0.07	−0.08	−2.03	−2.09	2	—
CDFP	0.12	−0.02	0.13	−1.46	39	17	1.41	1.37	32.84	31.43	25	582
DFR	0.22	0.12	0.12	−1.66	43	42	0.91	0.87	−1.10	−3.12	36	—
INFL	−0.04	−0.07	−0.72	−0.95	7	—	0.02	−0.01	−3.12	−3.54	8	—
M1	−0.07	−0.14	−2.13	−2.90	19	—	0.79	0.75	13.09	12.32	13	404
UNRATE	−0.12	−0.18	−1.20	−1.68	13	—	−0.01	−0.04	−5.73	−6.21	8	—
INDPRO	0.05	−0.06	0.03	−0.83	22	8	0.38	0.34	2.67	1.85	14	211
CUTIL	0.00	−0.07	−0.54	−1.13	15	—	1.26	1.23	23.99	23.41	10	915
REA	−0.06	−0.08	−1.36	−1.53	5	—	0.07	0.05	−6.36	−6.62	5	—
CFNAI	−0.02	−0.09	−0.64	−1.23	15	—	0.96	0.93	18.44	17.72	13	824
CLI	0.34	0.32	2.66	2.45	6	337	1.34	1.32	33.19	32.55	15	1284
BCI	0.41	0.37	3.48	3.08	11	226	1.41	1.40	36.36	35.91	14	1720
CCI	0.04	0.00	−0.25	−0.60	9	—	−0.44	−0.47	−9.75	−10.34	10	—
AUS	0.02	−0.04	0.04	−0.41	12	3	−0.26	−0.30	−6.15	−6.87	12	—
CAN	−0.08	−0.14	−0.69	−1.16	12	—	−0.30	−0.34	−7.90	−8.57	12	—
NZ	−0.16	−0.25	−1.99	−2.76	19	—	0.65	0.63	9.97	9.41	10	501
SA	0.08	0.01	0.26	−0.29	14	21	−0.45	−0.49	−8.24	−8.67	8	—
IND	0.03	−0.02	0.11	−0.22	9	10	0.20	0.18	2.65	2.46	4	207

Table 7 continued

	Expansion						Recession					
Strategy	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}
HA benchmark	0.01	0.01					0.06	0.06				
Panel B: Combination forecasts												
Mean	0.07	0.04	0.44	0.24	6	29	0.74	0.73	11.31	11.11	4	1193
Median	0.03	0.02	0.19	0.10	3	9	0.19	0.18	2.38	2.31	2	197
Trimmed mean	0.07	0.04	0.39	0.21	5	25	0.66	0.65	9.92	9.75	4	1047
Weighted mean	0.08	0.04	0.45	0.25	6	29	0.75	0.74	11.48	11.28	4	1214
DMSFE ($\theta = 0.9$)	0.07	0.04	0.43	0.23	6	28	0.79	0.77	12.16	11.94	4	1293
DMSFE ($\theta = 0.7$)	0.08	0.04	0.44	0.24	6	29	0.82	0.80	12.72	12.50	5	1356
ABMA	0.07	0.04	0.43	0.24	6	28	0.73	0.72	11.14	10.95	4	1172
Subset (k = 2)	0.10	0.04	0.50	0.11	10	42	1.08	1.06	19.37	18.98	7	2304
Subset (k = 3)	0.11	0.04	0.49	−0.06	14	53	1.21	1.19	24.87	24.29	10	3277
Subset (k = 4)	0.12	0.04	0.42	−0.28	18	63	1.29	1.26	28.40	27.66	13	3965
Subset (k = 5)	0.12	0.04	0.25	−0.58	21	70	1.31	1.28	30.48	29.60	15	4551
Subset (k = 6)	0.11	0.03	0.02	−0.94	24	74	1.31	1.28	31.38	30.35	17	5058
Subset (k = 7)	0.11	0.02	−0.28	−1.35	26	—	1.30	1.27	31.63	30.46	19	5511
PC (IC = AIC)	0.15	0.04	−0.99	−2.91	46	—	1.38	1.36	34.65	33.47	23	7491
PC (IC = BIC)	0.09	0.04	−1.36	−2.18	20	—	1.42	1.40	36.49	35.86	14	6132
PC (IC = R^2)	0.16	0.05	−0.79	−2.74	47	—	1.44	1.41	37.19	35.77	26	8210

Notes. This table reports portfolio performance results for a mean-variance investor with relative risk aversion of three who monthly allocates his wealth between commodities and risk-free T-bills using either the HA benchmark forecast (static portfolio strategy) or the individual predictive regression (combination) forecasts (dynamic portfolio strategy). The forecasts in Panel A are based on one each of the 28 predictor variables. The forecasts in Panel B are based on 28 predictors using the different combination methods outlined in Section 2.1.2. For each portfolio strategy, we report the annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized utility gain (net of cost), $\triangle$ ($\triangle_\tau$), the portfolio management fee that the investor would be willing to pay in order to have access to the dynamic strategy relative to the static strategy, the turnover ratio (TO) ratio, the ratio of the average turnover of the dynamic strategy relative to that of the static strategy, and the break-even transaction costs, τ^{BE}, that will render the investor indifferent between the dynamic and static portfolio strategies. We set proportional transaction costs of 20bps per dollar of trading. Since we use commodity futures, we avoid short sales restrictions but limit leverage to 50% of wealth to avoid excessive risk taking. Results are reported separately for NBER-dated business-cycle expansions and recessions. The out-of-sample forecast evaluation period is 1990:01-2016:12.

Table 9: Macroeconomic Variable Out-of-Sample Forecasting Results for Individual Forecasts

	Economic activity variables					Economic activity variables			
Predictor	INDPRO	CFNAI	TBL	DFY	Predictor	INDPRO	CFNAI	TBL	DFY
Basis	0.40	−0.36	−0.57	−0.33	Mean	7.27***	3.87***	13.08***	6.76***
INV	1.08**	−0.49	0.66**	−0.25	Median	5.82***	2.33***	11.49***	5.78***
PROD	0.50*	0.00	−0.23	−0.09	Trimmed mean	1.46***	0.35***	2.77***	2.16***
DP	−2.55	−0.94	−6.46	−2.16	Weighted mean	8.35***	4.86***	13.94***	7.06***
SP500	1.98**	1.13*	1.63**	9.85***	DMSFE, 0.9	10.78***	4.21***	17.28***	8.29***
TBL	−2.59	−0.91	−17.66	−4.66	DMSFE, 0.7	10.42***	4.43***	15.32***	7.93***
CTBL	3.28***	−0.24	20.52***	−3.02	ABMA	6.36***	3.08***	12.25***	6.47***
LTR	0.29	−0.68	−78.68***	2.32**	Subset (k=2)	13.03***	7.74***	19.49***	11.86***
TMS	−1.43	−0.20	−6.23	−1.22	Subset (k=3)	17.69***	11.55***	21.17***	15.73***
CTMS	0.70	0.27	1.29*	−2.86	Subset (k=4)	21.47***	15.21***	19.51***	18.73***
YS	−0.56	−0.46	−1.50	−1.71	Subset (k=5)	24.53***	18.72***	15.64***	20.91***
CDFP	0.80	0.90	−30.36***	16.25***	Subset (k=6)	27.03***	21.94***	10.31***	22.63***
DFR	−0.85	−3.92	2.32***	25.18***	Subset (k=7)	29.05***	24.96***	4.06***	23.94***
INFL	0.09	−0.34	−10.03	−4.05	PC (ic=AIC)	33.12***	0.28**	−57.11***	25.23***
M1	0.80	−0.84	−18.74	0.20	PC (ic=BIC)	33.12***	0.94**	−56.06***	25.97***
UNRATE	2.86**	−0.33	2.97***	−1.10	PC (ic=R2)	31.53***	1.14***	−59.44***	25.16***
INDPRO	14.16***	−1.05	2.17***	−1.92					
CUTIL	11.28***	21.83***	−1.68***	−1.94					
REA	−1.40	−0.55	−7.20	−0.07					
CFNAI	17.35***	13.21***	−2.94***	−1.62					
CLI	34.95***	−0.75	−17.24***	5.43**					
BCI	29.82***	0.76	−73.03***	18.78***					
CCI	9.71***	−0.21**	−25.42***	9.76***					
AUS	−0.20	−1.41	−0.26	10.60***					
CAN	−0.64	−0.55	−0.74	10.05***					
NZ	0.62	−0.48	−0.62	5.97***					
SA	−0.38	−0.25	0.57*	4.60***					
IND	−0.82	−2.96	−6.21	4.04***					

Notes. This table reports out-of-sample results for the individual forecast of macroeconomic activity variables: growth in industrial production, growth in consumer price index, changes in 3-month T-bill rate, changes in chicago Fed National Activity index, and changes in default yield spread. HA is the historical average benchmark forecast. MSFE is the mean squared forecast error. The $R^2_{\rm oos}$ statistic measures the proportional reduction in MSFE for the competing forecasts given in the first column relative to the HA forecast. Statistical significance for the $R^2_{\rm oos}$ statistic is based on the p-value for the Clark and West (2007) MSFE-adjusted statistic. This statistic tests the null hypothesis that the HA forecast MSFE is less than or equal to the competing forecast MSFE against the alternative hypothesis that the HA forecast MSFE is greater than or equal to the competing forecast MSFE. Results are reported for the full out-of-sample forecast evaluation period 1990:01-2016:12. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Statistical Evaluation of Monitoring Commodity Return Forecasts Performance

	Panel A: Individual Predictive model forecast							
	CPI				M2			
Predictor	t_0	t_1	R^2	GW t-stat	t_0	t_1	R^2	GW t-stat
Basis	0.12	−1.46	0.66	3.91	−2.27**	2.14**	1.40	5.98
INV	1.60	−2.56	2.00	1.48	−1.16	1.53	0.72	3.42
PROD	−0.24	−1.21	0.45	1.59	−1.39	0.80	0.20	1.86
DP	−1.75*	1.16	0.42	1.68	−0.89	0.09	0.00	4.18
SP500	−4.92***	7.09***	13.55	0.84	0.40	−1.09	0.37	0.52
TBL	−11.03***	14.92***	40.95	3.71	0.05	−1.61	0.80	4.59
CTBL	1.69*	−3.52***	3.71	4.02	−2.25**	2.44**	1.82	2.20
LTR	−0.89	2.49**	1.90	0.98	1.22	−0.95	0.28	1.05
TMS	−3.14***	3.79***	4.27	3.14	1.45	−2.71***	2.23	2.80
CTMS	−3.30***	5.14***	7.60	0.13	2.21**	−3.01***	2.74	0.04
YS	−0.50	−0.12	0.00	1.73	−0.62	0.22	0.02	2.23
CDFP	2.19**	−1.74*	0.93	2.22	−0.94	2.35**	1.69	2.29
DFR	3.72***	−4.98***	7.17	0.96	−0.04	0.72	0.16	1.07
INFL	−3.79***	5.10***	7.48	1.87	−0.24	−0.38	0.04	1.63
M1	−0.47	0.88	0.24	2.36	−1.44	1.91*	1.13	3.38
UNRATE	1.70*	−3.37***	3.42	2.09	−1.57	1.67*	0.86	3.94
INDPRO	2.69***	−3.71***	4.11	2.86	−2.24**	3.31***	3.30	3.14
CUTIL	5.02***	−6.32***	11.05	2.24	−0.52	1.81*	1.01	2.38
REA	−4.66***	6.54***	11.76	1.04	−0.46	−0.07	0.00	0.74
CFNAI	6.19***	−8.68***	19.01	2.07	−2.01*	3.43***	3.54	4.27
CLI	5.60***	−7.15***	13.75	2.91	−2.39**	4.28***	5.39	5.02
BCI	5.79***	−6.41***	11.33	5.22	−1.34	3.57***	3.82	5.68
CCI	2.21**	−3.06***	2.83	2.34	−0.01	0.37	0.04	1.34
AUS	−1.32	1.22	0.46	0.77	−0.23	−0.29	0.03	0.62
CAN	−1.82*	1.29	0.52	2.12	1.32	−2.74	2.29	1.81
NZ	0.80	−2.32**	1.65	1.80	−2.02*	1.97*	1.20	1.26
SA	−2.58**	2.98***	2.69	4.74	−0.06	−0.69	0.15	1.00
IND	2.05**	−2.67**	2.17	1.23	−2.07**	3.10***	2.90	0.88

Table 9 continued

	Panel A: Individual Predictive model forecast							
	VIX				Macroeconomic uncertainty			
Predictor	t_0	t_1	R^2	GW t-stat	t_0	t_1	R^2	GW t-stat
Basis	−1.91*	1.67*	0.86	3.91	−2.55**	2.45**	1.83	5.98
INV	−2.19**	2.37**	1.72	1.48	−0.23	0.24	0.02	3.42
PROD	2.11**	−2.73***	2.26	1.59	1.37	−1.54	0.73	1.86
DP	−0.94	0.51	0.08	1.68	−0.77	0.60	0.11	4.18
SP500	2.09**	−2.51**	1.92	0.84	5.16***	−5.30***	8.06	0.52
TBL	2.95***	−3.91***	4.54	3.71	4.36***	−4.66***	6.33	4.59
CTBL	−2.32**	2.28**	1.59	4.02	−3.83***	3.79***	4.29	2.20
LTR	−1.93*	2.36**	1.71	0.98	−2.22**	2.34**	1.68	1.05
TMS	3.94***	−4.63***	6.25	3.14	4.50***	−4.68***	6.39	2.80
CTMS	1.23	−1.40	0.61	0.13	1.74*	−1.78*	0.98	0.04
YS	2.66**	−3.12***	2.95	1.73	2.86***	−2.98***	2.70	2.23
CDFP	−3.07***	3.83***	4.37	2.22	−5.24***	5.48***	8.55	2.29
DFR	−2.36**	2.85***	2.47	0.96	−2.85***	2.99***	2.71	1.07
INFL	1.99*	−2.46**	1.84	1.87	1.63	−1.75*	0.95	1.63
M1	−1.87*	2.03**	1.27	2.36	−3.59***	3.63***	3.94	3.38
UNRATE	−3.58***	3.67***	4.03	2.09	−3.21***	3.18***	3.05	3.94
INDPRO	−3.03***	3.44***	3.56	2.86	−3.83***	3.94***	4.62	3.14
CUTIL	−6.27***	7.28***	14.18	2.24	−8.24***	8.52***	18.44	2.38
REA	1.71*	−2.15**	1.42	1.04	3.34***	−3.48***	3.64	0.74
CFNAI	−5.68***	6.50***	11.63	2.07	−6.50***	6.71***	12.29	4.27
CLI	−5.89***	6.90***	12.92	2.91	−7.49***	7.77***	15.82	5.02
BCI	−6.18***	7.54***	15.06	5.22	−9.51***	9.93***	23.51	5.68
CCI	−2.04**	2.35**	1.69	2.34	−3.85***	3.94***	4.61	1.34
AUS	3.22***	−3.73***	4.15	0.77	1.18	−1.29	0.52	0.62
CAN	2.26**	−2.92***	2.59	2.12	3.17***	−3.37***	3.43	1.81
NZ	−2.74***	2.63***	2.12	1.80	−3.00***	2.92***	2.59	1.26
SA	3.92***	−4.57***	6.10	4.74	2.65***	−2.81***	2.39	1.00
IND	−0.54	0.79	0.19	1.23	−2.18**	2.27**	1.58	0.88

Table 9 continued

	Panel B: Combination forecast							
	CPI				M2			
Predictor	t_0	t_1	R^2	GW t-stat	t_0	t_1	R^2	GW t-stat
Mean	5.12***	−5.20***	7.76	5.77	−1.04	3.26***	3.21	6.72**
Median	3.45***	−2.83***	2.44	4.58	−1.24	3.34***	3.36	6.81**
Trimmed mean	4.76***	−4.59***	6.17	5.99	−0.95	3.17***	3.03	6.77**
Weighted mean	5.09***	−5.15***	7.64	5.78	−1.04	3.27***	3.22	6.73**
DMSFE ($\theta = 0.9$)	5.04***	−5.11***	7.51	5.62	−1.10	3.32***	3.32	6.64**
DMSFE ($\theta = 0.7$)	5.12***	−5.25***	7.91	5.56	−1.14	3.36***	3.40	6.62**
ABMA	5.15***	−5.24***	7.88	5.76	−1.03	3.26***	3.21	6.72**
Subset (k = 2)	4.84***	−4.90***	6.96	5.18	−1.10	3.25***	3.19	6.47**
Subset (k = 3)	4.68***	−4.79***	6.67	4.69	−1.17	3.26***	3.21	6.28
Subset (k = 4)	4.56***	−4.74***	6.53	4.28	−1.24	3.27***	3.22	6.13
Subset (k = 5)	4.44***	−4.67***	6.36	3.92	−1.31	3.28***	3.23	6.01
Subset (k = 6)	4.34***	−4.63***	6.26	3.62	−1.38	3.29***	3.26	5.92
Subset (k = 7)	4.24***	−4.60***	6.18	3.37	−1.45	3.30***	3.28	5.87
PC (IC = AIC)	4.16***	−4.85***	6.82	2.56	−1.22	2.76***	2.32	4.70
PC (IC = BIC)	4.44***	−5.39***	8.30	2.61	−1.79	3.45***	3.57	4.56
PC (IC = R^2)	4.49***	−5.35***	8.19	2.61	−1.13	2.66***	2.15	5.04

	VIX				Macroeconomic uncertainty			
Predictor	t_0	t_1	R^2	GW t-stat	t_0	t_1	R^2	GW t-stat
Mean	−6.34***	7.77***	15.82	5.77	−9.39***	9.83***	23.15	6.72
Median	−3.99***	5.11***	7.53	4.58	−6.81***	7.18***	13.83	6.81
Trimmed mean	−6.39***	7.83***	16.04	5.99	−9.42***	9.86***	23.26	6.77
Weighted mean	−6.33***	7.75***	15.76	5.78	−9.38***	9.82***	23.10	6.73
DMSFE ($\theta = 0.9$)	−6.34***	7.75***	15.76	5.62	−9.41***	9.84***	23.18	6.64
DMSFE ($\theta = 0.7$)	−6.41***	7.83***	16.03	5.56	−9.44***	9.88***	23.32	6.62
ABMA	−6.36***	7.78***	15.88	5.76	−9.40***	9.84***	23.19	6.72
Subset (k = 2)	−6.24***	7.61***	15.28	5.18	−9.13***	9.55***	22.11	6.47
Subset (k = 3)	−6.14***	7.46***	14.76	4.69	−8.88***	9.28***	21.16	6.28
Subset (k = 4)	−6.05***	7.31***	14.29	4.28	−8.66***	9.04***	20.29	6.13
Subset (k = 5)	−5.95***	7.17***	13.79	3.92	−8.44***	8.80***	19.44	6.01
Subset (k = 6)	−5.85***	7.02***	13.31	3.62	−8.23***	8.58***	18.65	5.92
Subset (k = 7)	−5.77***	6.89***	12.90	3.37	−8.03***	8.36***	17.89	5.87
PC (IC = AIC)	−5.53***	6.52***	11.69	2.56	−7.63***	7.91***	16.30	4.70
PC (IC = BIC)	−5.98***	6.97***	13.16	2.61	−8.62***	8.91***	19.81	4.56
PC (IC = R^2)	−5.53***	6.52***	11.70	2.61	−7.68***	7.96***	16.49	5.04

Notes. This table reports te results of the conditional test of equal predictive ability of the HA forecast (benchmark, B) relative to the individual and combination forecasts (alternative, A) of log excess commodity. Using the loss differential, $\triangle L_{t+1} = e^2_{B,t+1} - e^2_{A,t+1}$ we test the null $\mathbb{E}[\triangle L_{t+1}|z_t] = 0$ against the two-sided alternative using each of the monitoring instrument growth in consumer price index (CPI), US money stcok (M2), the VIX, and the macroeconomic uncertainty measure of Jurado et al. (2015). The numbers shown are t-statistics of θ_0 and θ_1 from the regression $\triangle L_{t+1} = \theta_0 + \theta_1 z_t + \varepsilon$, the R^2 from this regression and the Giacomini and White (2006) (GW t-stat) test of conditional predictability. In all cases the ne Results are reported for the full out-of-sample forecast evaluation period 1990:01-2016:12. **, and *** indicate significance at the 5%, and 1% levels, respectively.

Table 11: Economic Evaluation of Monitoring Commodity Return Forecasts

Strategy	μ_p	σ_p	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}
HA benchmark	−0.003	0.06	−0.05	−0.05				
Panel A: Individual predictive forecasts								
Basis	−0.007	0.07	−0.10	−0.13	−0.49	−0.74	17	—
INV	−0.002	0.09	−0.03	−0.07	−0.47	−0.83	24	—
PROD	−0.004	0.06	−0.06	−0.06	−0.04	−0.04	1	—
DP	−0.007	0.06	−0.11	−0.12	−0.38	−0.41	3	—
SP500	−0.024	0.13	−0.18	−0.23	−4.02	−4.66	43	—
TBL	−0.005	0.06	−0.08	−0.08	−0.16	−0.18	2	—
CTBL	−0.003	0.06	−0.05	−0.05	0.01	0.01	1	39
LTR	0.011	0.13	0.09	−0.01	−0.58	−1.90	85	—
TMS	−0.011	0.07	−0.15	−0.16	−0.85	−0.90	4	—
CTMS	−0.007	0.06	−0.11	−0.13	−0.29	−0.38	6	—
YS	−0.004	0.06	−0.07	−0.07	−0.07	−0.09	2	—
CDFP	0.053	0.14	0.38	0.30	3.39	2.19	79	92
DFR	0.104	0.29	0.36	0.31	−1.07	−2.44	94	—
INFL	−0.013	0.09	−0.16	−0.16	−1.49	−1.53	4	—
M1	0.022	0.08	0.27	0.21	2.16	1.63	36	95
UNRATE	−0.001	0.07	−0.01	−0.04	0.07	−0.12	13	26
INDPRO	0.030	0.10	0.31	0.27	2.52	2.12	27	166
CUTIL	0.015	0.10	0.15	0.08	0.92	0.16	50	49
REA	−0.007	0.09	−0.08	−0.09	−0.88	−0.94	5	—
CFNAI	0.062	0.14	0.44	0.40	4.16	3.57	40	219
CLI	0.132	0.20	0.66	0.64	8.09	7.68	31	576
BCI	0.169	0.24	0.71	0.68	9.27	8.69	43	527
CCI	0.005	0.09	0.06	0.01	0.13	−0.31	30	39
AUS	−0.012	0.07	−0.17	−0.19	−1.04	−1.19	11	—
CAN	−0.009	0.07	−0.12	−0.15	−0.76	−0.94	13	—
NZ	−0.004	0.06	−0.07	−0.08	−0.15	−0.21	5	—
SA	−0.003	0.07	−0.04	−0.06	−0.07	−0.22	11	—
IND	0.008	0.07	0.11	0.06	0.95	0.60	23	65
Panel B: Combination forecasts								
Mean	0.020	0.06	0.33	0.30	2.35	2.17	13	134
Median	0.002	0.06	0.04	0.02	0.62	0.54	6	33
Trimmed mean	0.016	0.06	0.27	0.24	1.99	1.84	11	112
Weighted mean	0.020	0.06	0.33	0.30	2.39	2.21	13	136
DMSFE ($\theta = 0.9$)	0.021	0.06	0.35	0.31	2.49	2.31	13	143
DMSFE ($\theta = 0.7$)	0.023	0.06	0.36	0.33	2.61	2.42	13	151
ABMA	0.019	0.06	0.32	0.29	2.32	2.14	12	132
Subset (k = 2)	0.039	0.08	0.51	0.46	3.97	3.63	23	248
Subset (k = 3)	0.058	0.10	0.58	0.53	5.20	4.71	33	354
Subset (k = 4)	0.073	0.12	0.60	0.54	5.98	5.35	41	443
Subset (k = 5)	0.085	0.14	0.59	0.54	6.32	5.58	49	512
Subset (k = 6)	0.092	0.16	0.56	0.51	6.11	5.27	55	555
Subset (k = 7)	0.110	0.18	0.62	0.57	7.23	6.35	59	662
PC (IC = AIC)	0.074	0.21	0.35	0.28	1.69	0.22	96	450
PC (IC = BIC)	0.081	0.18	0.45	0.39	4.19	3.18	66	489
PC (IC = R^2)	0.121	0.25	0.48	0.42	3.37	1.89	99	724

Notes. This table reports portfolio performance results for a mean-variance investor with relative risk aversion of three who monthly allocates his wealth between commodities and risk-free T-bills using either the HA benchmark forecast (static portfolio strategy) or the individual predictive regression (combination) forecasts (dynamic portfolio strategy). The forecasts in Panel A are based on one each of the 28 predictor variables. The forecasts in Panel B are based on 28 predictors using the different combination methods outlined in Section 2.1.2 using the forecast monitoring decision rule in Giacomini and White (2006). The monitoring instrument we use is broad money, M2. For each portfolio strategy, we report the annualized mean realized return (μ_p), annualized realized volatility (σ_p), annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized utility gain (net of cost), $\triangle$ ($\triangle_\tau$), the portfolio management fee that the investor would be willing to pay in order to have access to the dynamic strategy relative to the static strategy, the turnover ratio (TO) ratio, the ratio of the average turnover of the dynamic strategy relative to that of the static strategy, and the break-even transaction costs, τ^{BE}, that will render the investor indifferent between the dynamic and static portfolio strategies. We set proportional transaction costs of 20bps per dollar of trading. Since we use commodity futures, we avoid short sales restrictions but limit leverage to 50% of wealth to avoid excessive risk taking. Results are reported for the full out-of-sample forecast evaluation period 1996:01-2016:12.

A. Appendix

A.1. Construction of the basis predictor

In this appendix, we lists all the 32 individual commodities grouped by sectors, the exchanges they are traded on, the corresponding Blomberg tickers, and the corresponding code in the Commitment of Traders report used in constructing the basis predictor. The commodity futures are traded on the Chicago Board of Trade (CBOT), the Chicago Mercantile Exchange (CME), the London Metal Exchange (LME), Intercontinental Exchange (ICE), the New York Commodities Exchange (COMEX), and the New York Mercantile Exchange (NYMEX). NA means data is not available.

Table A1: Individual Commodity Futures Data

Sector/Commodity	Exchange	Bloomberg Ticker	CFTC code(s)
Agriculture			
Corn	CBOT	C	002601, 002602
Rough Rice	CBOT	RR	039601, 039781
Soybean Meal	CBOT	SM	026603
Soybean Oil	CBOT	BO	007601
Soybeans	CBOT	S	005601, 005602
Wheat	CBOT	W	001601, 001602
Ethanol	CME	DL	025601
Lumber	CME	LB	058641, 058643
Cocoa	ICE	CC	073732
Coffee	ICE	KC	083731
Cotton	ICE	CT	033661
Orange Juice	ICE	JO	040701
Sugar	ICE	SB	080732
Energy			
Brent Crude Oil	ICE	CO	—
Gasoil	ICE	QS	—
Gasoline	NYMEX	HU/XB	11659
Heating Oil	NYMEX	HO	022651
Natural Gas	NYMEX	NG	023651
WTI Crude Oil	NYMEX	CL	067651
Livestock			
Feeder Cattle	CME	FC	061641
Lean Hogs	CME	LH	054641, 054642
Live Cattle	CME	LC	057642
Metals			
Palladium	COMEX	PA	075651
Platinum	COMEX	PL	076651
Aluminium	LME	LA	NA
Copper	LME	LP	085691, 085692
Lead	LME	LL	NA
Nickel	LME	LN	NA
Tin	LME	LT	NA
Zinc	LME	LX	NA
Gold	COMEX	GC	061641
Silver	COMEX	SI	084691