UCD GEARY INSTITUTE FOR PUBLIC POLICY
DISCUSSION PAPER SERIES

Commodity Futures Return Predictability and Intertemporal Asset Pricing

John Cotter
Michael Smurfit Graduate Business School, University College Dublin

Emmanuel Eyiah-Donkor
Rennes School of Business

Valerio Potì
Michael Smurfit Graduate Business School, University College Dublin

Geary WP2020/11
November 12, 2020

Commodity Futures Return Predictability and Intertemporal Asset Pricing*

John Cotter[1], Emmanuel Eyiah-Donkor[†2], and Valerio Potì[1]

[1]Michael Smurfit Graduate Business School, University College Dublin
[2]Rennes School of Business

October 9, 2020

Abstract

We find out-of-sample predictability of commodity futures 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, and the combination forecasts of commodity returns have significantly positive predictive power for future economic activity. Two-factor models featuring innovations in each of the combination forecasts and the market factor explain a substantial proportion of the cross-sectional variation of commodity and equity returns. The associated positive risk prices are consistent with the Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973), given how the predictors forecast an increase in future economic activity in the time-series. Overall, combination forecasts act as state variables within the ICAPM, thus resurrecting a central role for macroeconomic risk in determining expected returns.

JEL classification: C22, C53, G11, G12, G13

Keywords: Commodity futures returns; Predictability; Asset allocation; Macroeconomic risk; Intertemporal pricing

*The authors 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, Andrew Detzel, Anna-Maria Fuertes, Richard Levich, Jölle Miffre, Marta Szymanowska (CEMA discussant), Giovanni Pagliardi (EFMA discussant), Robert Korajczyk, and conference participants at the International Symposium of Forecasting 2017 in Cairns, Australia, the Commodity and Energy Markets Annual (CEMA) Meeting 2018 in Rome, Italy, International Association for Applied Econometrics 2019 Annual conference in Nicosia, Cyprus, the 2019 European Financial Management Association Annual (EFMA) Conference in Ponta Delgada, Island of St. Miguel, Portugal, and the 2019 Financial Management Association Annual Meeting in New Orleans.

†Corresponsing author. *E-mail addresses*: john.cotter@ucd.ie (John Cotter); emmanuel.eyiah-donkor@rennes-sb.com (Emmanuel Eyiah-Donkor); valerio.poti@ucd.ie (Valerio Potì).

1. Introduction

Compared to the vast literature on predictability of aggregate stock, bond, and currency returns (see, for example, Cochrane, 2011 and the references therein), the predictability of aggregate returns on commodities has received relatively little attention. This is despite the fact that commodity prices play an important role in explaining fluctuations in macroeconomic activity and help forecast it (Hamilton, 2009). Hence, they are natural candidates as drivers of variation of risk premia and, therefore, as predictors of asset returns. Also, 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 futures excess returns - measured as the return on the S&P Goldman Sachs Commodity Index (S&P GSCI) in excess of the one-month T-bill rate. We consider return-forecasting models that differ in the way they allow for variability of the futures conditional mean return based on 28 potential predictors. In addition to forecasts from individual predictive models, we also consider combination forecast methods that combine forecasts from the individual models.[1] This is important considering that the forecasting literature has identified two important issues that typically plague individual return-forecasting models, leading to their poor out-of-sample performance: parameter instability and model uncertainty (see, for example, Stock and Watson, 1996; Paye and Timmermann, 2006, Rapach and Wohar, 2006). 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 excess return models. As a benchmark model, we consider a simple no-predictability (historical average) return model against which we compare the performance of our time-varying mean models. The data set consists of monthly observations for the period from January 1976 to December 2016.

To measure the statistical performance of the futures excess return forecasts, we use the out-of-sample (OOS) R^2 which measures the proportional reduction in mean squared forecast error (MSFE) attainable by using the time-varying forecasts instead of the historical average forecast. 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. On the other hand, we find positive out-of-sample

[1] A simple diversification argument is often used to justify the use of combination forecast methods. That is, their ability to diversify against model instability and uncertainty, akin to how portfolio diversification reduces the risk of investment portfolios. For an introduction and survey of the literature on combination forecast methods, see Hendry and Clements (2004) and Timmermann (2006).

R^2 values of 0.30%–4.96%, which are statistically significant, for all the combination forecasts. To explain the different performance of the individual and combination forecasts, we test the stability of the forecasting ability of each using the Giacomini and Rossi's (2009) forecast breakdown test and by conducting an in-depth visual comparison of their relative MSFEs. The results suggest that the poor out-of-sample performance of the individual forecasts can be attributed to the instability of the models' parameters used in generating the forecasts. The ability of forecast combinations to diversify against model instability support their superior forecasting performance.

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 predictability when forming optimal portfolios composed of the commodity futures index and risk-free T-bills. We find that the gains in predictive accuracy from combination forecasts translate into higher Sharpe ratios and certainty equivalent return gains for the investor. For example, the investor would be willing to pay a fee of 3.4% per annum to have access to the portfolios generated by the combination forecasts relative to the one generated by the benchmark forecast. In contrast, these findings are reversed for individual forecasts.

We then examine the drivers of commodity futures return predictability. First, we address the sources of predictability by analysing the extent to which futures excess return predictability is related to the business cycle using the NBER-dated business cycle indicator. We find that return predictability is largely concentrated in economic recessions, with R^2 values as high as 18%. Second, we assess whether the ability of the combination forecasts to predict commodity returns stem from their ability to proxy for state variables that drive discount rates. To do so, we follow the advice in Cochrane (2005, 2007) and check whether combination forecasts predict economic activity. If time-varying expected returns are related to time-varying discount rates, then the predictors used to forecast asset returns (in our case commodity returns) should also have forecasting power for future economic activity. In a nutshell, this is because, in dynamic equilibrium, aggregate returns must reflect aggregate economic activity and vice versa. Consistent with this notion, we find that combination forecasts predict increases in economic activity proxied by the smoothed recession probability of Chauvet (1998), Aruoba, Diebold, and Scotti (2009) business condition index, the Chicago Fed national activity index, log growth in industrial production index, change in total capacity utilization, and log growth in total nonfarm payroll employment with R^2 values ranging from 2% to 11% for 1-, 3- and 12-month horizons.

Taken together, these results suggest that the ability of combination forecasts to predict commodity returns boils down to their ability to capture time-varying discount rates. That is, an implication of the ability of combination forecasts to predict both aggregate commodity returns and future economic activity is that they could be valid state variables within Merton's (1973) intertemporal capital asset pricing model (ICAPM).[2] Therefore, as a final contribution, we test this implication by estimating a two-factor models that includes a state variable (combination forecast of commodity returns) next to the market factor, using the cross-section of 23 individual commodity futures and 25 equity portfolios formed on size and book-to-market as test assets. Our results show significant positive risk prices associated with the state variables, consistent in sign with how the same variables forecast an increase in future economic activity. Our results thus imply consistency of the predictability in the time-series and the cross-section of commodity returns, an important implication of the ICAPM of Merton (1973).

Our study is related to a strand of literature that investigates the time-series predictability of commodity returns (see, for example, Hong and Yogo, 2012; Gorton, Hayashi, and Rouwenhorst, 2013; Gargano and Timmermann, 2014; Ahmed and Tsvetanov, 2016, and the references therein). However, we depart from these studies 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. Second, we address the impact of parameter instability and uncertainty about return prediction models by implementing forecast combination methods. In this regard, our approach is similar to Gargano and Timmermann, 2014. While they use commodity spot indexes, we study commodity futures and use a much broader set of forecast combination methods. We are motivated by studies such as Rapach and Wohar (2006) and Welch and Goyal (2008), among others, who show that the poor out-of-sample forecasting performance of predictive models of U.S. stock excess returns might be a direct consequence of structural breaks and uncertainty in the underlying data generating process. Many studies have shown that the use of combination forecasts leads to improved forecast performance in a variety of settings. See, for example, Stock and Watson (2003, 2004) for forecasting inflation and output growth for seven developed countries, and Timmermann (2006) and Rapach, Strauss, and Zhou (2010) for forecasting U.S.

[2]State variables that forecast macroeconomic activity (recession state variables) should command a risk premium since they are of hedging concern to investors. The argument is that besides stocks and bonds, most investors own other non-marketed assets such as labour income. Therefore, aggregate stock returns could be a relatively poor proxy for the return on aggregate wealth, which is the investment opportunity set of ultimate importance to the representative investor (Roll, 1977).

stock excess returns. Third, we address concerns raised by previous studies that find that statistical evidence of return predictability does not always translate into economic significance (Ahmed and Tsvetanov, 2016 for commodity futures returns; Della Corte, Sarno, and Tsiakas, 2009 and Potì, 2018 for exchange rate returns; Thornton and Valente, 2012 and Sarno, Schneider, and Wagner, 2016 for bond returns). We fill this gap by examining the utility gains that accrue to risk-averse investors who exploit predictability of commodity futures excess returns relative to the no-predictability benchmark in a mean-variance optimal asset allocation framework. Finally, our study goes further by linking the ability of combinations of individual predictors to forecast aggregate commodity returns to their ability to predict future economic activity and capture time-varying discount rates. In doing so, we specify, test and offer empirical support for a novel version of the ICAPM of Merton (1973) where the innovations to combination forecasts of commodity returns act as valid state variables.

The rest of the paper is organised as follows. Section 2 details the commodity futures returns data and predictor variables and presents descriptive statistics. It also details the return prediction models we consider, and the framework for evaluating out-of-sample return predictability. In Section 3, we present and discuss in-sample and out-of-sample evidence of commodity futures excess return predictability. Section 4 analyses the link between commodity return predictability, portfolio performance and the business cycle. In this section, we also present our ICAPM-type two factor model and discuss its implications for time-series and cross-sectional return predictability. Section 5 concludes.

2. Data and Methodology

2.1. Return and Predictor Data

Our dataset contains monthly observations for the sample period January 1976 to December 2016. It includes primarily end-of-month total return data on the S&P Goldman Sachs Commodity Index (S&P GSCI),[3] a fully investable commodity index, downloaded from Bloomberg. Our use of this index rather than individual commodities or more special-

[3]Three S&P GSCI indices are published: excess return, total return and spot. The excess return index measures the returns accrued from investing in uncollateralized nearby commodity futures, the total return index measures the returns accrued from investing in fully-collateralized nearby commodity futures, and the spot index measures the level of nearby commodity prices. Thus, the excess return and total return indices provide useful representations of returns available to investors from investing in the S&P GSCI. For more information, see `https://www.goldmansachs.com/gsci/insert.html`.

ized commodity spot indices is because it is designed to resemble the total return on an investable portfolio of commodities, hence it realistically reflects transaction costs. Compared to equally weighted portfolios of individual commodity futures (considered, for example, by Hong and Yogo, 2012; Gorton et al., 2013; Ahmed and Tsvetanov, 2016, and the references therein)[4] its advantage is that, as well as being diversified, its composition resembles realistic portfolios of investment managers seeking to gain exposure to the broad commodity market, viewed as an asset class. We compute excess returns as the log return on S&P GSCI less the log return on a one-month T-bill.[5] Subtraction of the risk-free rate is needed because the total return index measures the returns accrued from investing in fully-collateralized nearby commodity futures.[6]

As predictors, we consider a set of 28 variables. They include commodity market, stock market, treasury market, corporate bond market, currency market, and macroeconomic variables. The commodity and currency market variables include the futures basis (namely, the difference between the futures and the spot price), crude oil production, 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. These are all predictors that have been shown, in the prior studies on commodity return predictability mentioned earlier, to have predictive power for commodity spot and futures returns. For example, Gorton et al. (2013) find that individual commodity futures risk premia are driven by the basis and inventory levels. Their role can be rationalized with the classical theories of storage (Kaldor, 1939; Brennan, 1958) and normal backwardation (Keynes, 1930; Hicks, 1939), which imply that commodity market variables such as basis, influenced by hedging pressure, and inventory should exhibit predictive power for commodity returns.

The stock, treasury, and corporate bond market variables include the dividend-price ratio, the return on the S&P 500 stock market index, the yield on 3-month treasury bill, the term spread default premium, among others. The definition of the predictors,

[4]Due to the high storage, transportation and insurance costs associated with holding physical commodities, individual and institutional investors have traditionally relied on commodity futures to gain exposure to commodities. The S&P GSCI is the benchmark commodity futures tracked by investment vehicles such as commodity-based exchange traded products through which individual and institutional investors gain broad exposure to the commodities market (see, for example, Jensen, Johnson, and Mercer, 2000; Jensen and Mercer, 2011; and Erb and Harvey, 2016).

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

[6]The total return (i.e., the S&P GSCI total return index) is the measure of commodity returns that is completely comparable to returns from a regular investment in the S&P 500 (with dividend reinvestment) or a government bond, while the return on the excess return index is comparable to the return on the S&P 500 above cash.

motivation for their use, and relevant prior commodity return predictability studies are summarized in Table 1.

[Insert Table 1 about here]

Most of these variables are considered in studies of U.S. equity excess return predictability (Welch and Goyal, 2008), treasury and corporate bond excess return predictability including Gargano, Pettenuzzo, and Timmermann (2017) and Lin, Wu, and Zhou (2017), and references therein. Recent evidence such as Tang and Xiong (2012) and Hamilton and Wu (2015) show that the financialization of commodities has led to the commodities market becoming more integrated with capital markets, and explains why the dividend-price ratio of the S&P 500 stock index could be informative about future commodity returns.

Finally, the macroeconomic variables include, among others, inflation, money stock, unemployment rate, industrial production growth, degree of capacity utilization in U.S. manufacturing, and a global real economic activity index. 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 (hence, be captured by) 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.

Panel A of Table 2 presents descriptive statistics of monthly excess returns on the S&P GSCI for the full sample period. We report the number of observations, the mean, standard deviation, minimum and maximum values, first-order autocorrelation and Sharpe ratio of excess returns. The table shows that the mean excess return was 0.14% with volatility close to 6%. The index thus recorded an annualized Sharpe ratio of 0.09. The low mean return and high standard deviation suggests that static positions in commodities would be unattractive as a stand-alone investment strategy on a risk-adjusted basis.

Panel B of Table 2 presents the summary statistics for the predictor variables. The data on the variables used in computing the statistics are in percent except TBL, CTBL, TMS, CTMS, YS, CDFP which are in annualized percent because the t-bill and corporate bond rates data are annualized. Except for the commodity currencies, LTR, DFR, M1, and UNRATE, the other predictors are strongly positively autocorrelated with first-order autocorrelation coefficients between 0.28 and 0.99 encouraging the use of test statistics that are robust to autocorrelation when testing for predictability.

[Insert Table 2 about here]

2.2. Models for Forecasting Returns

This section details the individual and combination forecasting models used in our study of commodity excess-return predictability. The individual forecasting models condition on a single predictive variable and generate predictions, to which we refer as "*individual forecasts*", based on one of the 28 predictors at a time. The combination forecasting models are 16 different combinations of the individual forecasts. We refer to their predictions as "*combination forecasts*". The section also details our procedure for generating (pseudo) out-of-sample forecasts using these models and statistical and economic measures of out-of-sample forecasting performance.

2.2.1. Individual Forecasts

Following much of the literature on return predictability, we model commodity returns as

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1}, \tag{1}$$

where r_{t+1} is the continuously compounded return on the (fully collateralized) commodity futures index in excess of the risk-free rate at time $t+1$, $x_{i,t}$ is one of the 28 predictor variables of interest available at time t, and ε_{t+1} is a zero-mean error term.

We estimate first the individual forecasting models to generate (pseudo) out-of-sample forecasts based on a single predictive variable at a time and then combine these individual forecasts to obtain the combination forecasts (as described in the next sub-section). The out-of-sample forecasts are generated using a recursive (expanding window) estimation scheme as follows. Suppose T observations are available for r_t and $x_{i,t}$. We divide the total sample into two parts: an in-sample period containing the first $n = 167$ observations (February 1976 to December 1989) and an out-of-sample period containing the remaining $P = T - n = 312$ observations (January 1990 to December 2016). The choice of length of the in-sample estimation period enables us to have a sufficiently long out-of-sample forecast 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 first out-of-sample forecast of commodity excess returns based on the $x_{i,t}$ predictor is given by

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

where $\hat{\alpha}_{i,n}$ and $\hat{\beta}_{i,n}$ are the ordinary least squares (OLS) estimates of α_i and β_i in (1), respectively, from regressing $\{r_t\}_{t=2}^{n}$ on a constant and $\{x_{i,t}\}_{t=1}^{n-1}$. The next out-of-sample forecast is given by

$$\hat{r}_{i,n+2} = \hat{\alpha}_{i,n+1} + \hat{\beta}_{i,n+1} x_{i,\,n+1}, \tag{3}$$

where $\hat{\alpha}_{i,n+1}$ and $\hat{\beta}_{i,n+1}$ are the OLS estimates from regressing $\{r_t\}_{t=2}^{n+1}$ on a constant and $\{x_{i,t}\}_{t=1}^{n}$. We proceed in this recursive fashion until the end of the out-of-sample period, generating a time-series of P one-step-ahead out-of-sample forecasts of returns $\{\hat{r}_{i,t+1}\}_{t=n}^{T-1}$.

2.2.2. 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 may not provide enough guidance and so one is likely to ignore potentially important predictors. There are also the important issues of structural breaks resulting from changes of the parameters of the underlying data generating process and estimation uncertainty surrounding return forecasting models, leading to poor out-of-sample performance (see, for example, Paye and Timmermann, 2006; Rapach and Wohar, 2006; Welch and Goyal, 2008).

Combination forecasts have been shown to perform well out-of-sample (see, for example, Stock and Watson, 2004; Timmermann, 2006; Rapach et al., 2010). As suggested by Hendry and Clements (2004), there are several potential explanations for why forecast combination methods might work. The common thread is that they provide a means to diversify against estimation sampling error of the forecasting model parameters and model uncertainty. The latter, in turn, derives from model instability in the presence of possible structural breaks in the data generating process as well as the fact that all models are likely misspecified. Both sets of circumstances are impossible (or very difficult) to model in full and therefore an advantageous course of action is to diversify against the risk of forecast errors to which they give rise.[7]

We consider three types of combination forecasts: 4 simple combination forecasts; 9 performance-based forecasts; and 3 factor model forecasts. Our combination forecasts use individual predictive model forecasts as building blocks and differ in the way weights

[7] Attempting to correct the individual models in any particular way, in a financial portfolio management analogy, would correspond to hedging the forecasting error risk instead of diversifying it.

assigned to the individual forecasts are computed. Generally, it entails (i) estimating a regression of returns on each of the predictors, (ii) forming individual forecasts based on the estimated parameters from each of the regressions, and (iii) combining the individual forecasts to generate a single forecast. Formally, let $\hat{r}_{i,t+1}$ denote the out-of-sample forecast of r_{t+1} computed at time t based on the ith predictor variable as given by (2). A combination forecast at time $t+1$, $\hat{r}^{\text{CF}}_{t+1}$, is a weighted average of the individual out-of-sample forecasts:

$$\hat{r}^{\text{CF}}_{t+1} = \sum_{i=1}^{N} w_{i,t}\hat{r}_{i,t+1}, \tag{4}$$

where $w_{i,t}$ is the weight assigned to the ith forecast with $\sum_{i=1}^{N} w_{i,t} = 1$ and N is the number of individual forecasts.

The first set of combination forecasts we consider use simple averaging schemes: mean, trimmed mean, median, and weighted-mean forecasts. They are very easy to generate and do not take into account the historical performance of the individual forecasts. Stock and Watson (2003, 2004) find that simple combining methods work well in forecasting inflation and output growth for seven developed countries using a large number of potential predictors compared to more sophisticated methods. Rapach et al. (2010) report similar results for forecasting the U.S. stock excess returns. Smith and Wallis (2009) argue that the reason why simple combination methods work better compared to more sophisticated methods is because there is little or no estimation error associated with estimating their combining weights. The mean combination forecast, $\hat{r}^{\text{Mean}}_{t+1}$, is the average of the N (N=28) individual predictive model forecasts that assign equal weights, $w_{i,t} = 1/N, i = 1, ..., N$, to each forecast defined in (2). The trimmed mean forecast, $\hat{r}^{\text{Trimmed mean}}_{t+1}$, sets in (4) $w_{i,t} = 0$ for the lowest and highest forecasts and $w_{i,t} = 1/(N-2)$ for the remaining individual forecasts. Removing the lowest and highest forecasts before combining mitigates the influence of outliers on the forecasts. The median combination forecast, $\hat{r}^{\text{Median}}_{t+1}$, is the sample median of the 28 individual predictive model forecasts. The weighted-mean forecast ($\hat{r}^{\text{Weighted mean}}_{t+1}$) proposed by Bates and Granger (1969) specifies the combination weights to be proportional to the inverse of the estimated residual variance, $\sigma^2_{i,t}$, for the individual predictive regression models given by (1),

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

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 depend inversely on the historical performance of the individual predictive model forecasts over a holdout out-of-sample period,

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

where $\theta \in (0,1)$ is the discount factor.[8] When $\theta < 1$, greater importance is attached to the individual predictive model forecasts with the lowest mean squared forecast error (MSFE). By attaching more weight on the recent forecasting accuracy of the individual predictive models, thereby allowing time-variation in the data generating process of returns, the DMFSE should work relatively well. In the special case where there is no discounting ($\theta = 1$) and forecasts are uncorrelated, this leads to the optimal combination weights proposed by Bates and Granger (1969) given by (5). We consider θ values of 0.9 and 0.7 to examine the impact of discounting forecast further back in time. Rapach et al. (2010) also show that the DMSFE combination forecasts of U.S. stock 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:

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

where $\triangle_{i,t} = \text{AIC}_{i,t} - \max_i(\text{AIC}_{i,t})$ 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 and robust to parameter and model uncertainty. 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 compared to the mean combination forecast.

Third, we use so-called complete Subset Regression forecasts, a class of combination forecasts recently proposed by Elliott, Gargano, and Timmermann (2013), based on equally-weighted averages of all forecasts predictive regression models that include a

[8]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 weights at this time point.

fixed number of the predictor variables. They show that the complete subset regression forecast controls 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 studying the predictability of U.S. stock excess returns, Elliott et al. (2013) find that subset regression combination forecasts produce more accurate forecasts than approaches based on equally-weighted combinations of forecasts from individual return prediction models or forecasts generated by bagging, ridge regression or Bayesian model averaging. Suppose the number of potential predictors that enter a regression is N. A subset regression combination is then defined by the set of regression models that include a specified number of regressors, $k \leq N$. The $k \leq N$ dimensional subset forecasts are then averaged to generate the forecasts. In our analysis, we use a maximum k value of 7. Given N regressors in full and k regressors chosen for short models, one has to average over the $C_k^N = N!/(k!(N-k)!)$ subset regression forecasts combinations, where ! is the factorial function. As a special case, when $k = 1$, this results in the mean combination forecast. Formally, the Subset Regression forecast is thus given by

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

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

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

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

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

2.3. Measures of Predictability

2.3.1. Statistical Measures

Motivated by the debate on in-sample vs. out-of-sample predictability (see, for example, Inoue and Kilian, 2005), we measure and evaluate evidence on the predictive ability of our forecasting models of commodity excess-returns both in-sample and (pseudo) out-of-sample.

Our measures of in-sample predictability are the t-statistic of the significance of β_i in (1) estimated over the full sample from February 1976 to December 2016 and the coefficient of determination, R^2. Because the error term can be heteroskedastic and autocorrelated, we compute the t-statistic using heteroskedasticity and autocorrelation consistent standard errors à la Newey and West (1987).

Following a popular approach in the return predictability literature, we measure the accuracy of the forecasts generated by the candidate (individual and combination) models relative to a no-predictability benchmark return forecast using the Campbell and Thompson (2008) out-of-sample R^2 statistic, R^2_{oos}, given by:

$$R^2_{\text{oos}} = 1 - \frac{\frac{1}{T-n} \sum_{t=n}^{T-1} \left(r_{t+1} - \hat{r}_{t+1|t} \right)^2}{\frac{1}{T-n} \sum_{t=n}^{T-1} \left(r_{t+1} - \bar{r}_{t+1|t} \right)^2}, \tag{10}$$

where r_{t+1} is the realized log return at time $t+1$, $\hat{r}_{t+1|t}$ is a candidate forecast, and $\bar{r}_{t+1|t}$ is the benchmark forecast generated by the constant expected excess return model that includes only a constant, α, alongside the error term:

$$r_{t+1} = \alpha + \varepsilon_{t+1}, \tag{11}$$

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.

The R^2_{OOS} statistic measures the proportional reduction in mean squared forecast error (MSFE) for a candidate forecast relative to the HA forecast. A positive R^2_{OOS} implies the candidate forecast outperforms the HA forecast as it has a lower MSFE. 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 candidate model forecasts against the one-sided (upper-tailed) alternative hypothesis that HA forecast MSFE is greater than the MSFE of the candidate forecast. Under the null of no-predictability, the HA return forecast is expected to have a lower MSFE.

In addition to carrying out our formal test of significance of the R^2_{OOS}, we assess the stability of our forecasts also by examining their relative MSFEs defined as the ratio of the MSFE of a predictive forecast to the MSFE of the HA benchmark forecast, following Stock and Watson (2003). To this end, we divide our out-of-sample forecast period into two halves and compute the MSFEs over the two periods. Forecasts that are stable should have relative MSFEs less than one in both periods, whereas unstable forecasts will have relative MSFEs less than one in one period and greater than one in another period or greater than one in both periods.

Also, we 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 displays 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.

2.3.2. Economic Measures

We next detail the asset allocation framework that we use to evaluate the risk-adjusted 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. We are motivated by studies such as Della Corte et al. (2009) 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.

Following Campbell and Thompson (2008), we consider a mean-variance investor who monthly allocates her wealth between commodities futures and risk-free T-bills using either the individual (combination) forecasts or HA forecast of futures excess 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. Like Campbell and Thompson (2008), we assume that the investor uses five-year rolling-windows of past returns to estimate the variance of commodity futures excess return. We set the risk aversion coefficient equal to 3 and allow for a moderate portfolio leverage of 50%, similar to other studies such as Campbell and Thompson (2008) and Rapach, Ringgenberg, and Zhou (2016). Since we use futures, we do not need to impose short-sales constraint.

To evaluate the performance of the portfolios generated by the individual and combination forecasts, we first compute the realized average utility or certainty equivalent return (CER) 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 the 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 forecast and the portfolio generated by the HA return forecast. We annualize the CER gain so that it can 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 relative to the portfolio generated by the HA return forecast. Positive values indicate that the time-varying predictability models perform better than 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 portfolio strategies based on the individual and combination forecasts to be more costly to implement compared to the strategy based on the HA return forecast 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 costs, where we set the proportional transaction costs to 20 basis points per dollar of trading. Second, following Della Corte et al. (2009), we calculate the break-even proportional transaction costs, τ^{BE}, that will render the investor indifferent between two competing portfolio strategies 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 individual (combination) and HA portfolio strategies, respectively, and $\mathrm{TO}^{\mathrm{FC}}$ and $\mathrm{TO}^{\mathrm{HA}}$ are their respective average turnover. In comparing a dynamic portfolio based on individual (combination) forecast to that of static strategy based on HA forecast, an investor who faces 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

3.1. In-Sample Predictability

Table 3 reports the OLS estimates of the slope coefficient, β, the associated t-statistic computed using Newey and West (1987) autocorrelation and heteroskedasticity-consistent standard errors, 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 commodity futures excess returns at conventional levels. The R^2 statistics for these 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):

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

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 (1), $\bar{r}$ is the 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 23 out of the 28 predictors. For example, the R^2_{REC} for the DFR, INDPRO, CLI, and BCI quadruples during recessions.

[Insert Table 3 about here]

3.2. Out-of-Sample Predictability

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 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. To circumvent this problem and guard against overfitting, we conduct out-of-sample tests for predictability.

3.2.1. Tests of Out-Of-Sample Predictability

Table 4 summarizes the performance of the individual and the combination forecasts relative to the HA benchmark forecast for one-month ahead forecast of commodity futures returns. The table reports the mean squared forecast error, the R^2_{OOS} as defined in (10) and the MSFE-adjusted statistic. The out-of-sample statistics are based on forecasts generated using the recursive estimation approach detailed in the earlier section. As explained therein, statistical significance of $R^2_{\text{OOS}} > 0$ is assessed using the p-value of the MSFE-adjusted statistic of Clark and West (2007).

In Panel A of Table 4, we report results for the individual predictive regression forecasts. Most of the individual predictors fail to outperform the HA forecast in terms of MSFE, similar to the evidence reported in many of the extant return predictability studies (see, for example, Welch and Goyal, 2008; Gargano and Timmermann, 2014; Neely et al., 2014). Only for twelve out of the 28 predictors do we find a positive R^2_{OOS}, with

MSFEs significantly less than the MSFE of the HA forecast at the 5% level. The exceptions are the impressive R^2_{oos} statistics of 3.80% and 6.92% recorded for CLI and BCI, and these are significantly greater than zero at the 1% level. In comparison, for the four variables that exhibit the second greatest predictive ability, namely CDFP, DFR, CUTIL, and CFNAI, the R^2_{oos} ranges from 1.13% to 3.15%. Pagano and Pisani (2009) find similar results when examining the possibility of forecasting crude oil futures returns with the CLI as a predictor. CLI and BCI variables are indicators of global economic activity and provide early warning of cyclical conditions in the world economy, being combinations of key economic indicators (including opinions about economic activity, financial and monetary data, labour market statistics, information on production, stocks and orders on finished goods, and foreign trade[9]). The rather good performance of these two variables suggests a possible link between forecasting ability for commodity returns and ability to forecast economic activity, which we shall investigate later.

Panel B of Table 4 report results for the combination forecasts. In this case, the findings are much more supportive of predictability. The R^2_{oos} generated by each of the combination forecasts is in some cases impressive, ranging from 0.30% for the Mean combination forecast to 4.96% for the PC (IC = BIC) combination forecasts, and all 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.

[Insert Table 4 about here]

3.2.2. Tests of Stability of Return Forecasts

As earlier hypothesized, parameter instability of the individual predictive models may explain their poor out-of-sample performance as reported in Table 4, and why forecast combinations work well to improve forecast accuracy.

We assess the stability of our forecasts by examining their relative MSFEs. We divide our out-of-sample forecasts period in two parts: the first period from January 1990 to December 2003 and the second period from January 2004 to December 2016, and compute the relative MSFE of each model (i.e., the model MSFE relative to the MSFE of the constant-mean model) over the two periods. Forecasts that are stable should have relative MSFEs less than one in both periods, whereas unstable forecasts will have relative MSFEs

[9] See the following link from the website of the OECD: `https://www.oecd.org/sdd/leading-indicators/45430429.pdf`.

less than one in one period and greater than one in another period or greater than one in both periods. Figure 1 displays the scatterplot of the logarithm of the relative MSFEs of the forecasts based on all the individual predictors in the first period (x-axis) versus the second period (y-axis). In the scatterplot, a point represents the pair of log relative MSFEs for each of the predictors. If the forecasts were stable, we should expect the points to be scattered around the third (southwest) quadrant. The figure shows that many of the points are rather scattered in the first (northeast) and second (southeast) quadrants, indicating the poor performance of the individual forecasts in one period and vice versa.

The scatterplot of the log relative MSFE for the combination forecasts is displayed in Figure 2. As can be seen from the figure, the combination forecasts show considerable stability. All the points plot in the third (southwest) quandrant, indicating improved forecasting performance compared to the HA benchmark forecast in both out-of-sample periods.

[Insert Figure 1 about here]

[Insert Figure 2 about here]

As a further analysis of the instability of the individual predictive models, Table 5 reports the Giacomini and Rossi (2009) t-statistic and the associated p-value for testing the stability of forecasting ability of the individual predictive models (that is, the hypothesis that a model's out-of-sample MSFE is equal to its in-sample MSFE). The out-of-sample forecasts used in the test are generated using the same recursive estimation approach as before, and the test statistics are computed similarly using a quadratic loss function. As presented in the table, the null hypothesis is rejected at the 1% level for all the individual predictive models. These results provide further evidence of parameter instability of the individual predictive models, and help explain their poor out-of-sample performance.

[Insert Table 5 about here]

3.2.3. Economic Evaluation of Return Forecasts

Table 6 presents results for 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 relative 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 dynamic portfolio strategy based on individual (combination) forecasts to the average turnover of the static portfolio strategy generated by the HA return forecast. 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 portfolio strategy. We will refer to CER gain of 2% or more as economically significant.

From Column 6 of Panel A in Table 6, we see that for almost all the individual predictors the CER gain is negative, consistent with the R^2_{oos} limited statistical significance reported in Table 4. Positive values well above the 2% threshold are documented for only four predictors, namely CDFP, CUTIL, CLI and BCI, consistent with 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. In contrast to the results for the individual forecasts, positive CER gains are realized for all combination forecasts. The Subset ($k = 2, ..., 7$) and the PC, $\textsc{ic} = R^2$ forecasts all record CER gains well above the 2% level. These findings are consistent, in terms of greater economic value of combination forecasts compared to the individual ones, with the greater predictive ability of the former compared to the latter suggested by the statistical tests reported in Table 4.

Consistent with these findings, Sharpe ratios of the commodity portfolios generated by the individual (combination) predictive model forecasts are lower (higher), for almost all predictors, than for the portfolio that relies on the historical average return forecast. These results are also consistent with the poor (strong) statistical performance of the individual (combination) forecasts.

[Insert Table 6 about here]

Column 7 of Panel A in Table 6 reports the CER gains net of transaction costs for the individual predictive model forecasts. From the table, we observe that, just as in the case without transaction costs, CER losses are recorded for almost all the portfolio strategies based on the individual predictive model forecasts. On the other hand, accounting for the effect of transaction costs does not erode the performance of the portfolios based on the combination forecasts as they continue to deliver positive CER gains net of transaction costs well above the 2% level. 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. 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 relative magnitude of Sharpe ratios for individual and combination forecasts is consistent with the relative magnitude of CER gains when both are calculated net of transaction costs.

4. Return Predictability and the Business Cycle

We conduct further analysis to shed light on the economic drivers of commodity return predictability by investigating the link between return forecasts and the real economy.

4.1. Statistical Performance of Return Forecasts and the Business Cycle

Studies such as Rapach et al. (2010), Henkel, Martin, and Nardari (2011) for US stock returns, and Gargano and Timmermann (2014) for commodity spot index returns show that return predictability 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, as done earlier for our analysis of in-sample predictability, we use the same R^2 statistics for business cycle expansions (EXP) and recessions (REC), given in (15), but this time we compute them based on the errors of the out-of-sample forecasts.

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 reports 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. The DFR, CLI and BCI predictors, however, continue to show significant levels of predictability with significantly greater than zero R^2 at the 5% level during recessions and expansions. The lack of out-of-sample predictive ability of all predictors but DFR, CLI and BCI is in sharp contrast to our earlier in-sample findings where we documented that commodity futures return predictability was stronger in recessions relative to expansions for 23 out of the 28 predictors.

Panel B of Table 7 reports results for the combination forecasts. None of the forecasts are statistically greater than zero R^2_{oos} during expansions. However, during business cycle recessions, the combination forecasts deliver R^2_{oos} values ranging from 0.73% to 18.18%

which are statistically greater than zero at the 5% level and, in some cases, at levels close to or at the 1% threshold. These results show that (out-of-sample) predictability from combination forecasts is stronger in recessions relative to expansions and are supportive of the findings in Gargano and Timmermann (2014).

[Insert Table 7 about here]

To sum up, whilst almost all the individual forecasts of futures returns fail to outperform the benchmark forecast in both recessions and expansions, predictability from combination forecasts is stronger in recessions compared to expansions. Therefore, these results suggest that commodity predictability is a phenomenon largely associated with recessions which can be captured by combination forecasts rather than by individual forecasts.

4.2. Economic Performance of Return Forecasts and the Business Cycle

We now examine whether the performance of the portfolios that can be generated using the commodity return forecasts is 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.

We are motivated by the fact that, in the foregoing analysis, we found evidence of countercyclical commodity return predictability. The key questions are then whether commodity return predictability is genuinely countercyclical and, if so, whether it is related to time-variation in risk premia (discount rates). Asset pricing models featuring habit persistence such as Campbell and Cochrane (1999) suggest that risk premia move countercyclically and, due to a reduced surplus consumption ratio, the Sharpe ratio of the aggregate stock market should be higher during recessions than in expansions. Wachter (2006) derives implications for bond risk premia and the term structure of interest rates in a setting with habit persistence. If risk premia vary 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 (net of transaction costs of 20 basis points) 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 in magnitude. Only a few individual predictors, notably CDFP, CUTIL, CLI and BCI, have better performance in recessions compared to expansions. In contrast, as shown in Panel B, the Sharpe ratios of portfolios based on all the combination forecasts are substantially higher in recessions relative to expansions. This provides strong support for the suggestion of Campbell and Cochrane (1999) that risk premia move countercyclically and this gives rise to return predictability.

Table 8 also reports estimates 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 benefit concentrated in the recessionary phases of the business-cycle relative to expansions, especially for all the combination forecasts. In contrast, the results for the individual predictive model forecasts are mixed. This is not surprising considering their poor performance shown earlier.

[Insert Table 8 about here]

Our results taken together show that, as suggested by the fact that commodity return predictability (captured by combination forecasts) tracks business conditions, the performance of strategies that exploit it is indeed higher when these are entered when business conditions are weak and vice-versa, consistent with the possibility that commodity return predictability derives from time-variation of risk premia.

4.3. State Variables, Economic Activity, and Risk Premia

Cochrane (2007) suggests that it is more likely that return predictability is due to time-variation in risk premia (rather than time-varying mean-reverting mispricing) if the predictors used to forecast returns display predictive power for future economic activity. In this case, according to the intertemporal capital asset pricing model (ICAPM) framework of Merton (1973),[10] the predictors could be considered proxies for state variables that capture changes in the investment opportunity set. Moreover, since the sensitivity of risk-premia to state variables depends on risk aversion and the latter decreases with wealth, risk premia should be more reactive to changes in the state variables during reces-

[10]See, for example, Campbell (1996), Cochrane (2005), Maio and Santa-Clara (2012), among others.

sions. Therefore, if predictability is due to time-varying risk premia, it should be stronger during recessions.

In our context, this means that, if certain variables or combinations thereof forecast commodity returns because they proxy for state variables in the ICAPM sense, they should predict economic activity too, and their forecasting power should be stronger during recessions. If this was empirically the case, the implication would be that combination forecasts capture time-variation in commodity futures risk premia driven by macroeconomic risk whereas individual forecasts fail to do so. This, in turn, would imply that combination forecasts outperform individual forecasts because the former proxy for state variables that drive variation in commodity risk premia (hence in their expected returns) whereas the latter do not.

We test this explanation of the performance of forecast combinations by examining whether combination forecasts of commodity returns can be used to forecast future economic activity, hence whether they can be treated as valid state variables within the intertemporal capital asset pricing model (ICAPM) framework of Merton (1973). This amounts to testing an implication of the ICAPM for the relation between predictability of commodity returns and predictability of economic activity. It is the first of two implications of the ICAPM for commodity (excess-)returns and their predictability that we test.

For a given factor model to be consistent with the ICAPM, the factor risk premia should obey sign restrictions if the factors are state variables that forecast aggregate commodity returns and macroeconomic activity in time-series regressions (see applications in Maio and Santa-Clara, 2012 and Boons, 2016). If a state variable forecasts an increase in future expected aggregate commodity returns and economic activity, the risk price associated with the state variable in the cross-section should also be positive, and vice versa. This is the second implication of the ICAPM for commodity (excess-)returns and their predictability that we test.

We test in turn each one of the two implications of the ICAPM outlined above in the two sub-sections that follow.

4.3.1. Predicting Future Economic Activity

To test whether a state variable forecasts future economic activity, we run the following univariate predictive regression,

$$y_{t+h} = \alpha_i + \beta_i z_{i,t} + \varepsilon_{t+h}, \tag{16}$$

where $y_{t+h} = y_{t+1} + ... + y_{t+h}$ is the economic activity variable, h is the forecast horizon corresponding to 1-, 3-, or 12-months, and $z_{i,t} = \hat{r}^{\text{CF}}_{t+1}$ is one of the 16 combination forecast of commodity returns (state variable) at a time. It is well known that because of the use of overlapping data, the error term ε_{t+h} is serially correlated thereby distorting statistical inference. We address this issue by computing heteroskedasticity and autocorrelation consistent t-statistic for the test of $\beta_i = 0$. As proxies for y_t, we use the smoothed recession probability (SRP) of Chauvet (1998), Aruoba et al. (2009) business condition index (ADSI), the Chicago Fed national activity index (CFNAI), log growth in industrial production index (IP), change in total capacity utilization (TCU), and log growth in total nonfarm payroll employment (PAYEMS). Similar variables are used in studies such as Nieto and Rubio (2014), Lin et al. (2017), Choi, Mueller, and Vedolin (2017), and Maio and Philip (2018), among others. The data on SRP, CFNAI, IP, TCU, and PAYEMS are obtained from the St. Louis FED database (FRED) whereas ADSI is obtained from the Federal Reserve Bank of Philadelphia database (ALFRED). IP and TCU are lagged by one-month to account for delays in the release of such data thus ensuring the data represents publicly available information.

Table 9 reports estimation results of the predictive regression in (16) for the 1-, 3-, and 12-month horizons. From the table, we can see that all the combination forecasts of commodity returns display strong predictive content for future economic activity at all horizons. They forecast significantly increases in ADSI, CFNAI, IP, and TCU, and a decline in SRP. For all state variables, t-stats for the significance of β_i and the R^2 values rise from 1-month horizon to the 3-months horizon, and then falls for the 12-months horizon. For example, the mean combination forecast of commodity returns significantly forecast increases in changes in IP with R^2 values that increase from 11.81% for $h = 1$ to 21.78% for $h = 3$, and then falls to 7.5% for $h = 12$. Of all the economic activity variables, only the results for the PAYEMS amount to mixed findings about the predictive power of future economic activity.

[Insert Table 9 about here]

4.3.2. How Is Exposure to State Variable Risk Priced?

If the combination forecasts of commodity returns are truly state variables, they should be priced in the cross-section of asset returns. Specifically, since combination forecasts predict increases in future economic activity, they should be priced with a positive risk premium. To test this implication, we estimate a simple two-factor asset pricing model representing an implication of Merton's (1973) ICAPM:

$$E(R_i) = \lambda_M \beta_{i,M} + \lambda_z \beta_{i,z} + e_i, \qquad i = 1, ..., N, \tag{17}$$

where $E(R_i)$ is the expected excess return on test asset i and $\beta_{i,M}$ and $\beta_{i,z}$ are the asset loadings on the market factor and the state variable, respectively, λ_M is the risk premium for exposure to the market factor, and λ_z is the risk premium for exposure to innovations in state variable z. The proxy for the market factor is the CRSP value-weighted return minus the return on the 1-month Treasury bill downloaded from Professor Kenneth French's Data library.

Following Campbell (1996), the time-series dynamics for each state variable (combination forecast of commodity returns) is specified as a first-order vector autoregressive VAR(1) process,

$$F_{t+1} = A_0 + A_1 F_t + \tilde{Z}_{t+1}. \tag{18}$$

The first element of the state vector $F_t = (R_{M,t}, z_t)$ is the excess return of the market portfolio and the second element is the state variable, where $z_{i,t} = \hat{r}^{\text{CF}}_{t+1}$ is one of the 16 combination forecast of commodity returns at a time. The corresponding elements in the error vector $\tilde{Z}_t$ represent the innovations to state variables that is used as a risk factor in (17).

As test assets, we use the monthly returns on 23 individual commodity futures covering the energy, grains, oilseeds, livestock, metals and softs categories obtained from Thomson Reuters Datastream, and 25 portfolios of CRSP NYSE/AMEX/NASDAQ stocks formed on size and book-to-market also downloaded from Professor Kenneth French's Data library. The latter dataset is widely used in cross-sectional asset pricing tests. Our justification for expanding the set of test assets to include stocks is inline with the ICAPM which posits that state variables in the model must be state variables for all assets, and not just commodities. The sample period is January 1990 to December 2016, and is dictated by the availability of commodity futures data. Futures returns are measured as the logarithmic price changes of the front-end contracts up to one month before maturity; the positions are then rolled to the second-nearest contract, and so on. We use individual commodities as opposed to commodity futures portfolios which is standard in the commodity pricing literature (see for, example, Yang, 2013; Szymanowska, Roon, Nijman, and Goorbergh, 2014 and Bakshi, Gao, and Rossi, 2019). Our reasons for using individual commodities are two-fold: first, the cross-section of commodities is small which limits the number of portfolios that can be formed. This could distort the risk premia estimates. Second, Ang, Liu, and Schwarz (2010) argue that the use of individual assets can lead to more efficiency gains in the estimated risk prices than portfolios because the wider dispersions should make up for the noise in the estimated exposures.

To test our two-factor model in (17), we run cross-sectional regressions using the two-step approach following Cochrane (2005). In the first step, the factor loadings are estimated from a multivariate time-series regression for each of the 23 individual commodity futures or an extended set which includes the 25 equity portfolios:

$$R_{i,t+1} = \delta_i + \beta_{i,M} R_{M,t+1} + \beta_{i,z} \tilde{z}_{t+1} + \varepsilon_{i,t+1}, \qquad t = 1, ..., T \text{ for each } i, \tag{19}$$

where $R_{i,t+1}$ is the excess return on test asset i at time $t+1$, $R_{M,t+1}$ is the excess return on the market portfolio at time $t+1$, and $\tilde{z}_{t+1}$ is the innovation in the state variable z at time $t+1$. In the second step, we estimate OLS cross-sectional regressions across the test assets of average returns on the estimated betas in the first step,

$$\bar{R}_i = \lambda_0 + \lambda_M \hat{\beta}_{i,M} + \lambda_z \hat{\beta}_{i,z} + \alpha_i, \qquad i = 1, ..., N, \tag{20}$$

where $(\hat{\lambda}_M, \hat{\lambda}_z)$ and $\hat{\alpha}_i$ are the estimated risk prices and pricing errors, respectively, and $\hat{\lambda}_0$ is the zero-beta rate.[11]

We test the significance of the risk prices using t-statistics computed with GMM standard errors that correct for heteroskedasticity and autocorrelation in the error term, and errors-in-variables bias in the cross-sectional regression (see Cochrane, 2005). This adjustment is important considering that both the betas and the innovations to the state variables are estimated, and are thus subject to estimation error. We also assess the fit of the model based on each state variable by computing the cross-sectional OLS coefficient of determination (see Campbell and Vuolteenaho, 2004; Fernandez-Perez, Fuertes, and Miffre, 2017; Maio and Philip, 2018; among others),

$$R^2_{\text{OLS}} = 1 - \frac{\text{Var}_N(\hat{\alpha}_i)}{\text{Var}_N(\bar{R}_i)}, \tag{21}$$

where $\text{Var}_N(\cdot)$ is the cross-sectional variance, and R^2_{OLS} represents the fraction of the cross-sectional variance of average excess returns on the test assets that is explained by the factor loadings associated with the model. When an intercept is not included in the

[11] We also report results for a version of the model that does not include an intercept in the cross-sectional regression as dictated by the asset pricing model. Essentially, if a given model is correctly specified, the intercept in the cross-sectional regression should be equal to zero.

cross-sectional regression, the R^2 measure can assume negative values.

Table 10 presents results for monthly cross-sectional regressions using the 23 individual commodities as test assets. The table reports the unconditional monthly average risk premiums and associated t-statistics, and the cross-sectional OLS R^2 for two-factor model specifications that include innovations in each of the combination forecasts of commodity returns state variables next to the market portfolio. Panel A of the table shows that the risk premiums of all the state variables are positive and statistically significant. The results also indicate that the state variables have good explanatory power for the cross-section of commodity returns as indicated by the R^2_{OLS} ranging from 9% for the mean combination forecast to 28% for the DMSFE combination forecast. However, all the risk premium estimates for the market portfolio are negative and statistically insignificant. The results for the models that restricts the intercept to zero are presented in Panel B. There are clear differences with the results in Panel A. Risk premiums for the market factor, although now significant, increase substantially to levels that more than double the typical values of the sample average of the market portfolio excess returns. Risk premiums for the state variables also almost double. Unsurprisingly, however, the fit of the model also deteriorates substantially with R^2_{OLS} values as low as about 4%. In summary, the results based on the individual commodity test assets indicate a potential estimation inconsistency, akin to sample selection bias, resulting from the use of a restricted universe of test assets.

[Insert Table 10 about here]

Table 11 presents results for the extended set of test assets that include, in addition to the 23 individual commodities, the 25 equity portfolios formed on size and book-to-market. The results for the models that include an intercept in the cross-sectional regressions reported in Panel A indicate very similar findings to those reported in Panel A of Table 10. Although the risk premiums for the market are now positive, however, they are not statistically significant based on a one-tailed t-test. We also observe a considerably high explanatory power of over 74% for all choices of state variable. Turning to the models that restrict the intercept to zero, we make the following two observations. First, the estimated market risk premiums are all statistically significant with typical values that are close to the sample average of the market portfolio excess returns. Second, the estimated risk premiums are positive and significant risk premiums for all the combination forecast state variables and the R^2_{OLS} is considerably high, values of about 70%.

[Insert Table 11 about here]

Overall, our results are consistent in sign with how the state variables (all our 16 combination forecasts of commodity returns) forecast both positive economic activity in the time-series and returns in the cross-section, an implication of the ICAPM. This is especially the case for the models estimated using the extended set of test assets (that include the individual commodities and the equity portfolios) restricting the intercept to zero as dictated by the asset pricing model.

5. Conclusion

This paper provides a comprehensive study on aggregate commodity futures return predictability using a large set of predictors, including commodity, stock, corporate bond and treasury market, and macroeconomic variables. Our analysis considers individual predictive regression models and forecast combination methods that account for both parameter instability and model uncertainty.

We find that almost all of the individual predictive regression model forecasts of futures excess returns fail to outperform the benchmark historical average return forecast both statistically and economically. Forecast stability analysis using relative mean squared forecast errors and a forecast breakdown test show strong evidence of instability in the relation between commodity futures excess returns and the individual predictors, and provides an explanation for their inconsistent out-of-sample performance. Combination forecasts, on the other hand, are very stable over our sample and perform significantly better than the historical average forecast both statistically and economically. The superior forecasting performance of the combination forecasts can be attributed to their ability to diversify against instability and uncertainty associated with the individual predictive models.

We also find that the sources of predictability of combination forecasts for commodity futures returns have links to the real economy. Commodity return predictability is 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. Importantly, combination forecasts display significant predictive power at horizons ranging from 1-month to 12-months for macroeconomic activity proxied by the smoothed recession probability of Chauvet (1998), Aruoba et al. (2009) business condition index, the Chicago Fed national activity index, log growth in industrial production index, change in total capacity utilization, or log growth in total nonfarm payroll employment.

Finally, we provide evidence consistent with Merton's (1973) ICAPM framework that the combination forecasts of commodity returns are valid state variables that command significant positive risk premiums in the cross-section of individual commodity futures and equity portfolio returns. Our results also establish an important implication of the ICAPM, namely the positive sign of the risk prices associated with the state variables ability to forecast an increase in future economic activity. These results provide a further explanation for the significant out-of-sample performance of the combination forecasts, implying it derives from an ability of the combination forecasts of picking up time variation in the expected compensation for macroeconomic risks in commodity futures returns.

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.

Ang, A., Liu, J., Schwarz, K., 2010. Using stocks or portfolios in tests of factor models. Journal of Financial and Quantitative Analysis pp. 1–74.

Aruoba, S. B., Diebold, F. X., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business & Economic Statistics 27, 417–427.

Bakshi, G., Gao, X., Rossi, A. G., 2019. Understanding the sources of risk underlying the cross section of commodity returns. Management Science 65, 619–641.

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.

Boons, M., 2016. State variables, macroeconomic activity, and the cross section of individual stocks. Journal of Financial Economics 119, 489–511.

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

Campbell, J. Y., 1996. Understanding risk and return. Journal of Political economy 104, 298–345.

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.

Campbell, J. Y., Vuolteenaho, T., 2004. Bad beta, good beta. American Economic Review 94, 1249–1275.

Chauvet, M., 1998. An econometric characterization of business cycle dynamics with factor structure and regime switching. International economic review pp. 969–996.

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.

Choi, H., Mueller, P., Vedolin, A., 2017. Bond variance risk premiums. Review of Finance 21, 987–1022.

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., 2005. Asset pricing: Revised edition. Princeton university press.

Cochrane, J. H., 2007. Financial markets and the real economy, in rajnish mehra, ed. 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.

Della Corte, P., Sarno, L., Tsiakas, I., 2009. 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.

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

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.

Fernandez-Perez, A., Fuertes, A.-M., Miffre, J., 2017. Commodity markets, long-run predictability, and intertemporal pricing. Review of Finance 21, 1159–1188.

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 65, 508–540.

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.

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 .

Hendry, D. F., Clements, M. P., 2004. Pooling of forecasts. The Econometrics Journal 7, 1–31.

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.

Inoue, A., Kilian, L., 2005. In-sample or out-of-sample tests of predictability: Which one should we use? Econometric Reviews 23, 371–402.

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.

Jensen, G. R., Mercer, J. M., 2011. Commodities as an investment. Research Foundation Reviews 6, 1–33.

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

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.

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

Maio, P., Philip, D., 2018. Economic activity and momentum profits: further evidence. Journal of Banking & Finance 88, 466–482.

Maio, P., Santa-Clara, P., 2012. Multifactor models and their consistency with the icapm. Journal of Financial Economics 106, 586–613.

Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica: Journal of the Econometric Society pp. 867–887.

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.

Nieto, B., Rubio, G., 2014. Volatility bounds, size, and real activity prediction. Review of Finance 18, 373–415.

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

Paye, B. S., Timmermann, A., 2006. Instability of return prediction models. Journal of Empirical Finance 13, 274–315.

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., Wohar, M. E., 2006. Structural breaks and predictive regression models of aggregate us stock returns. Journal of Financial Econometrics 4, 238–274.

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

Roll, R., 1977. A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. Journal of financial economics 4, 129–176.

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.

Smith, J., Wallis, K. F., 2009. A simple explanation of the forecast combination puzzle. Oxford Bulletin of Economics and Statistics 71, 331–355.

Stock, J. H., Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30.

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.

Szymanowska, M., Roon, F., Nijman, T., Goorbergh, R., 2014. An anatomy of commodity futures risk premia. The Journal of Finance 69, 453–482.

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.

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.

Yang, F., 2013. Investment shocks and the commodity basis spread. Journal of Financial Economics 110, 164–184.

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 similar to Hong and Yogo (2012). 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 $log(\text{global crude oil production}(t)) - log(\text{global crude oil production}(t-1))$. 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 $log(\text{global crude oil inventory}(t)) - log(\text{global crude oil inventory}(t-1))$. 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.

Table 1: *continued*

Predictor	Article(s)	Variable definition and motivation for their consideration
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.
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.

Table 1: *continued*

Predictor	Article(s)	Variable definition and motivation for their consideration
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.
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 policy-makers 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 the 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.

Table 1: *continued*

Predictor	Paper(s)	Variable definition and motivation for their consideration
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, Rogoff, and Rossi (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 for 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: Returns							
S&P GSCI	491	0.14	5.57	−28.29	22.31	0.16	0.09
Panel B: Predictor Variables							
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	
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	
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	
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 28 predictors. 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 except TBL, CTBL, TMS, CTMS, YS, CDFP which are in annualized percent. The sample period is from February 1976 to December 2016.

Table 3: Full-Sample Predictive Regression Estimates

Predictor Variable	β	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 OLS estimation results for the bivariate predictive regression model of log commodity excess returns and the predictor variables individually. The immediate right of slope coefficients, *beta*, report the t-statistics calculated using Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors. R^2 is the coefficient of determination. 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). The sample period is February 1976 to December 2016. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Out-of-Sample Performance of 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 (Pane A) and combination (Panel B) 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 one-sided (upper-tailed) alternative hypothesis that the HA forecast MSFE is greater than or equal to the competing forecast MSFE. The out-of-sample forecast evaluation period is January 1900 to December 2016. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Tests of Stability of Return Forecasts

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. The null hypothesis is that a model that displays good forecasting performance in one sample period will continue to do so in other sample periods. That is, the out-of-sample MSFE of a model is equal to its in-sample MSFE. 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 in 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 Performance of 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 each of the 28 predictor variables. The forecasts in Panel B are based on 28 predictors using the different forecast combination methods outlined in Section 2.2.2. 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. The out-of-sample forecast evaluation period is January 1990 to December 2016.

Table 7: Statistical Performance of Return Forecasts in Expansions and Recessions

	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 (Panel A) and combination (Panel B) 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. The out-of-sample evaluation period is January 1990 to December 2016. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Economic Performance of Return Forecasts in Expansions and Recessions

	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 8: 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 each of the 28 predictor variables. The forecasts in Panel B are based on 28 predictors using the different combination methods outlined in Section 2.2.2. For each portfolio strategy, we report the annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized utility gain or certainty equivalent return 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 evaluation period is January 1990 to December 2016.

Table 9: Predicting Economic Activity with Combination Forecasts of Commodity Returns

	Forecast horizon: 1-month			Forecast horizon: 3-months			Forecast horizon: 12-months		
Combination forecast	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
y = SRP									
Mean	−0.28	−1.76	6.62	−0.86	−1.49	7.45	−1.44	−1.19	1.65
Median	0.05	0.21	0.06	0.17	0.20	0.09	2.18	0.88	1.08
Trimmed mean	−0.29	−1.68	6.06	−0.90	−1.38	6.63	−1.35	−0.96	1.18
Weighted mean	−0.28	−1.77	6.74	−0.86	−1.51	7.61	−1.44	−1.22	1.71
DMSFE ($\theta = 0.9$)	−0.27	−1.83	7.28	−0.86	−1.57	8.25	−1.41	−1.28	1.82
DMSFE ($\theta = 0.7$)	−0.27	−1.85	7.59	−0.86	−1.60	8.66	−1.39	−1.26	1.81
ABMA	−0.28	−1.74	6.50	−0.87	−1.46	7.29	−1.43	−1.15	1.59
Subset (k = 2)	−0.20	−2.34	10.41	−0.63	−2.07	11.84	−1.25	−2.14	3.81
Subset (k = 3)	−0.15	−2.52	11.60	−0.48	−2.25	13.19	−0.98	−2.41	4.48
Subset (k = 4)	−0.12	−2.57	11.97	−0.39	−2.30	13.56	−0.79	−2.45	4.60
Subset (k = 5)	−0.11	−2.60	12.16	−0.33	−2.33	13.73	−0.67	−2.46	4.59
Subset (k = 6)	−0.09	−2.60	12.12	−0.29	−2.31	13.61	−0.57	−2.39	4.42
Subset (k = 7)	−0.08	−2.58	11.98	−0.26	−2.29	13.42	−0.51	−2.32	4.23
PC (IC = AIC)	−0.06	−2.14	9.88	−0.19	−2.33	13.38	−0.52	−2.55	8.22
PC (IC = BIC)	−0.11	−4.30	23.64	−0.33	−3.91	27.43	−0.86	−3.24	15.25
PC (IC = R^2)	−0.05	−2.05	9.14	−0.18	−2.24	12.52	−0.52	−2.58	8.45
y = ADSI									
Mean	1.42	2.93	15.07	3.90	2.35	13.84	7.47	2.45	4.07
Median	1.03	1.46	2.53	2.54	1.03	1.83	1.97	0.24	0.08
Trimmed mean	1.57	3.01	15.38	4.26	2.39	13.60	7.99	2.31	3.80
Weighted mean	1.40	2.94	15.14	3.85	2.36	13.91	7.38	2.49	4.10
DMSFE ($\theta = 0.9$)	1.36	3.00	15.77	3.75	2.41	14.44	7.00	2.55	4.07
DMSFE ($\theta = 0.7$)	1.36	3.05	16.33	3.74	2.45	14.86	6.82	2.52	4.01
AMBA	1.43	2.93	15.00	3.95	2.34	13.76	7.56	2.41	4.04
Subset (k = 2)	0.92	3.44	18.92	2.57	2.81	17.92	5.30	3.59	6.24
Subset (k = 3)	0.68	3.59	20.02	1.92	2.94	19.10	3.99	3.83	6.85
Subset (k = 4)	0.55	3.63	20.39	1.55	2.97	19.46	3.23	3.87	7.00
Subset (k = 5)	0.47	3.67	20.67	1.32	3.00	19.75	2.73	3.90	7.04
Subset (k = 6)	0.41	3.68	20.70	1.16	3.00	19.74	2.38	3.86	6.94
Subset (k = 7)	0.37	3.67	20.66	1.04	3.00	19.69	2.13	3.81	6.82
PC (IC = AIC)	0.23	2.73	14.79	0.69	2.39	15.59	1.74	2.92	8.37
PC (IC = BIC)	0.37	4.28	25.26	1.07	3.50	26.01	2.61	3.20	12.87
PC (IC = R^2)	0.23	2.68	14.27	0.68	2.36	15.28	1.75	2.98	8.63
y = CFNAI									
Mean	0.91	2.76	12.71	2.68	2.47	16.65	5.29	2.67	5.32
Median	0.69	1.50	2.30	1.84	1.16	2.46	1.78	0.34	0.17
Trimmed mean	1.02	2.86	13.14	2.91	2.51	16.28	5.61	2.52	4.89
Weighted mean	0.90	2.76	12.77	2.64	2.48	16.72	5.22	2.71	5.35
DMSFE ($\theta = 0.9$)	0.87	2.81	13.28	2.57	2.53	17.32	4.96	2.78	5.32
DMSFE ($\theta = 0.7$)	0.88	2.86	13.78	2.57	2.60	18.00	4.87	2.79	5.33
ABMA	0.92	2.76	12.65	2.71	2.46	16.57	5.36	2.64	5.29
Subset (k = 2)	0.59	3.16	15.67	1.74	2.91	21.01	3.66	3.69	7.77
Subset (k = 3)	0.43	3.27	16.54	1.29	3.03	22.22	2.74	3.87	8.43
Subset (k = 4)	0.35	3.31	16.83	1.04	3.06	22.55	2.21	3.90	8.56
Subset (k = 5)	0.30	3.34	17.08	0.89	3.09	22.82	1.87	3.93	8.57
Subset (k = 6)	0.26	3.35	17.09	0.78	3.08	22.75	1.63	3.91	8.43
Subset (k = 7)	0.24	3.35	17.05	0.70	3.07	22.66	1.45	3.88	8.27
PC (IC = AIC)	0.15	2.50	12.52	0.47	2.48	18.42	1.19	2.91	10.27
PC (IC = BIC)	0.23	3.88	20.59	0.71	3.49	29.43	1.77	3.11	15.37
PC (IC = R^2)	0.14	2.40	11.60	0.45	2.40	17.50	1.18	2.91	10.10

Table 9: continued

	Forecast horizon: 1-month			Forecast horizon: 3-months			Forecast horizon: 12-months		
Combination foreacst	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
$y =$ IP									
Mean	1.05	3.70	11.81	3.03	3.70	21.78	5.56	3.73	7.50
Median	1.05	2.55	3.75	2.77	1.91	5.65	3.97	0.81	1.09
Trimmed mean	1.19	3.94	12.51	3.36	3.79	22.06	5.96	3.50	7.05
Weighted mean	1.03	3.69	11.77	3.00	3.72	21.90	5.50	3.79	7.57
DMSFE ($\theta = 0.9$)	0.99	3.68	12.02	2.91	3.88	22.62	5.22	3.88	7.53
DMSFE ($\theta = 0.7$)	0.98	3.57	11.99	2.86	3.88	22.65	5.01	3.77	7.19
ABMA	1.06	3.72	11.83	3.07	3.67	21.64	5.62	3.66	7.42
Subset (k = 2)	0.65	3.88	13.61	1.92	4.44	25.97	3.72	5.39	10.20
Subset (k = 3)	0.48	3.87	13.95	1.41	4.62	27.04	2.76	5.73	10.86
Subset (k = 4)	0.38	3.85	13.96	1.14	4.67	27.34	2.22	5.80	10.99
Subset (k = 5)	0.32	3.82	13.86	0.96	4.70	27.46	1.87	5.83	10.93
Subset (k = 6)	0.28	3.80	13.78	0.85	4.69	27.41	1.63	5.77	10.75
Subset (k = 7)	0.25	3.76	13.54	0.76	4.66	27.25	1.45	5.69	10.54
PC (IC = AIC)	0.14	2.54	7.40	0.48	3.24	19.90	1.19	4.49	13.01
PC (IC = BIC)	0.25	4.10	16.03	0.73	5.34	31.17	1.65	4.41	17.10
PC (IC = R^2)	0.13	2.48	7.18	0.47	3.13	18.91	1.17	4.55	12.75
$y =$ TCU									
Mean	0.81	4.16	12.12	2.31	4.42	22.71	4.19	3.63	8.28
Median	0.79	2.67	3.67	2.03	2.06	5.43	2.27	0.80	0.69
Trimmed mean	0.90	4.39	12.45	2.52	4.42	22.25	4.27	3.47	7.03
Weighted mean	0.80	4.15	12.13	2.29	4.47	22.92	4.17	3.65	8.46
DMSFE ($\theta = 0.9$)	0.77	4.18	12.43	2.22	4.74	23.76	3.99	3.57	8.55
DMSFE ($\theta = 0.7$)	0.75	4.03	12.25	2.17	4.70	23.41	3.74	3.54	7.79
AMBA	0.82	4.16	12.11	2.34	4.37	22.49	4.21	3.61	8.09
Subset (k = 2)	0.51	4.36	14.12	1.47	5.50	27.43	2.85	4.40	11.64
Subset (k = 3)	0.37	4.34	14.42	1.08	5.75	28.46	2.11	4.50	12.35
Subset (k = 4)	0.30	4.30	14.34	0.87	5.77	28.59	1.69	4.51	12.37
Subset (k = 5)	0.25	4.24	14.12	0.73	5.78	28.47	1.41	4.49	12.09
Subset (k = 6)	0.22	4.20	13.92	0.64	5.69	28.16	1.22	4.44	11.68
Subset (k = 7)	0.19	4.14	13.57	0.57	5.61	27.75	1.08	4.39	11.25
PC (IC = AIC)	0.11	3.03	8.77	0.39	4.18	23.48	1.01	5.61	18.19
PC (IC = BIC)	0.21	5.07	19.32	0.61	8.21	38.74	1.43	4.40	24.81
PC (IC = R^2)	0.11	2.92	8.30	0.37	3.97	21.82	0.96	5.65	16.87
$y =$ PAYEMS									
Mean	0.18	1.54	5.22	0.54	1.25	6.13	1.49	1.26	3.42
Median	0.09	0.64	0.39	0.20	0.38	0.26	−0.16	−0.08	0.01
Trimmed mean	0.21	1.66	5.71	0.61	1.32	6.52	1.65	1.26	3.40
Weighted mean	0.18	1.52	5.19	0.53	1.24	6.11	1.47	1.27	3.43
DMSFE ($\theta = 0.9$)	0.17	1.52	5.39	0.51	1.24	6.38	1.41	1.29	3.49
DMSFE ($\theta = 0.7$)	0.18	1.60	5.97	0.53	1.31	7.02	1.43	1.35	3.70
AMBA	0.18	1.55	5.24	0.54	1.25	6.14	1.52	1.26	3.41
Subset (k = 2)	0.12	1.78	7.06	0.37	1.48	8.54	1.09	1.72	5.58
Subset (k = 3)	0.09	1.86	7.67	0.28	1.55	9.34	0.83	1.85	6.23
Subset (k = 4)	0.07	1.89	7.91	0.22	1.58	9.62	0.67	1.88	6.38
Subset (k = 5)	0.06	1.92	8.16	0.19	1.61	9.90	0.57	1.90	6.47
Subset (k = 6)	0.06	1.94	8.26	0.17	1.62	9.98	0.50	1.90	6.40
Subset (k = 7)	0.05	1.95	8.32	0.15	1.63	10.01	0.45	1.89	6.29
PC (IC = AIC)	0.03	1.34	4.39	0.09	1.20	5.97	0.33	1.69	6.37
PC (IC = BIC)	0.05	2.30	10.91	0.16	1.96	14.29	0.59	2.33	13.66
PC (IC = R^2)	0.03	1.32	4.25	0.09	1.20	5.97	0.33	1.73	6.52

Notes. This table reports estimation results for the bivariate predictive regression $y_{t+h} = \alpha_i + \beta_i z_{i,t} + \varepsilon_{t+1}$, where $y_{t+h} = y_{t+1} + ... + y_{t+h}$ is the economic activity variable, h is the forecast horizon corresponding to 1-, 3-, or 12-months, and $z_{i,t} = \hat{r}^{\text{CF}}_{t+1}$ is one of the 16 combination forecast of commodity returns (state variable) at a time. y_{t+h} is the smoothed recession probability (SRP) of Chauvet (1998), Aruoba et al. (2009) business condition index (ADSI), the Chicago Fed national activity index (CFNAI), log growth in industrial production index (IP), change in total capacity utilization (TCU), or log growth in total nonfarm payroll employment (PAYEMS). The forecast horizons are 1-, 3-, and 12-months. To the immediate right of β is the t-statistics calculated using the Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors. R^2 is the coefficient of determination. The sample period is January 1976 to December 2016.

Table 10: Factor Risk Premia Estimates from ICAPM: Individual Commodity Test Assets

Model	λ_0	t_0	λ_M	t_M	λ_z	t_z	R^2_{OLS}
Panel A: Cross-sectional regressions							
Mkt + Mean	0.12	2.88	−0.07	−0.08	0.07	2.67	26.87
Mkt + Median	0.11	2.60	−0.01	−0.01	0.03	1.56	8.68
Mkt + Trimmed mean	0.12	2.86	0.02	0.02	0.06	2.63	25.02
Mkt + Weighted mean	0.12	2.89	−0.07	−0.08	0.08	2.67	26.86
Mkt + DMSFE ($\theta = 0.9$)	0.12	2.89	−0.12	−0.12	0.08	2.67	27.07
Mkt + DMSFE ($\theta = 0.7$)	0.12	2.92	−0.26	−0.27	0.08	2.67	27.64
Mkt + ABMA	0.12	2.88	−0.07	−0.08	0.07	2.67	26.87
Mkt + Subset (k = 2)	0.12	3.00	−0.08	−0.08	0.14	2.68	27.05
Mkt + Subset (k = 3)	0.12	3.08	−0.08	−0.08	0.19	2.70	26.99
Mkt + Subset (k = 4)	0.13	3.14	−0.07	−0.07	0.24	2.71	26.66
Mkt + Subset (k = 5)	0.13	3.19	−0.05	−0.05	0.29	2.75	26.68
Mkt + Subset (k = 6)	0.13	3.22	−0.04	−0.04	0.34	2.81	27.14
Mkt + Subset (k = 7)	0.13	3.30	−0.05	−0.05	0.37	2.86	26.79
Mkt + PC (IC = AIC)	0.11	2.80	0.02	0.03	0.47	2.19	20.94
Mkt + PC (IC = BIC)	0.10	2.69	−0.35	−0.35	0.37	2.12	19.61
Mkt + PC (IC = R^2)	0.12	2.95	−0.06	−0.06	0.52	2.42	24.94
Panel B: Restricted zero-beta rate ($\lambda_0 = 0$) cross-sectional regressions							
Mkt + Mean			1.76	1.75	0.12	4.22	8.64
Mkt + Median			1.24	1.28	0.06	3.50	−4.60
Mkt + Trimmed mean			1.88	1.85	0.10	4.23	6.86
Mkt + Weighted mean			1.76	1.75	0.12	4.22	8.60
Mkt + DMSFE ($\theta = 0.9$)			1.70	1.67	0.13	4.04	8.82
Mkt + DMSFE ($\theta = 0.7$)			1.53	1.51	0.13	4.04	9.13
Mkt + ABMA			1.76	1.74	0.12	4.23	8.68
Mkt + Subset (k = 2)			1.88	1.83	0.22	4.00	7.32
Mkt + Subset (k = 3)			1.96	1.93	0.31	4.16	6.28
Mkt + Subset (k = 4)			2.03	2.00	0.38	4.12	5.13
Mkt + Subset (k = 5)			2.11	2.06	0.45	4.10	4.38
Mkt + Subset (k = 6)			2.15	2.10	0.51	4.10	4.45
Mkt + Subset (k = 7)			2.19	2.20	0.56	4.00	3.51
Mkt + PC (IC = AIC)			1.79	1.76	0.86	3.80	3.58
Mkt + PC (IC = BIC)			0.95	0.96	0.70	3.89	5.26
Mkt + PC (IC = R^2)			1.84	1.82	0.89	4.05	5.97

Notes. This table reports the risk premiums (in %) for each state variables (z), the 16 combination forecasts of commodity returns, next to the market (Mkt) portfolio based on the OLS cross-sectional regression:

$$\bar{R}_i = \lambda_0 + \lambda_M \hat{\beta}_{i,M} + \lambda_z \hat{\beta}_{i,z} + \alpha_i, \qquad i = 1, ..., N.$$

The testing assets (R_is) are the excess returns on the 23 individual commodity futures. Panel A reports results for a regression of returns on exposures to the VAR(1) innovations in the state variables. Panel B restricts the intercept to zero. λ_0, λ_M and λ_z denote the zero beta rate, and risk premium for the market and each state variables, respectively. Next to the risk premium estimates are displayed t-statistics based on GMM standard errors corrected for heteroskedasticity, autocorrelation, and errors-in-variable bias. R^2_{OLS} represents the fraction of the cross-sectional variance of average excess returns on the test assets that is explained by the factor loadings associated with the model. The sample period is January 1990 to December 2016

Table 11: Factor Risk Premia Estimates from ICAPM: Extended Set of Test Assets

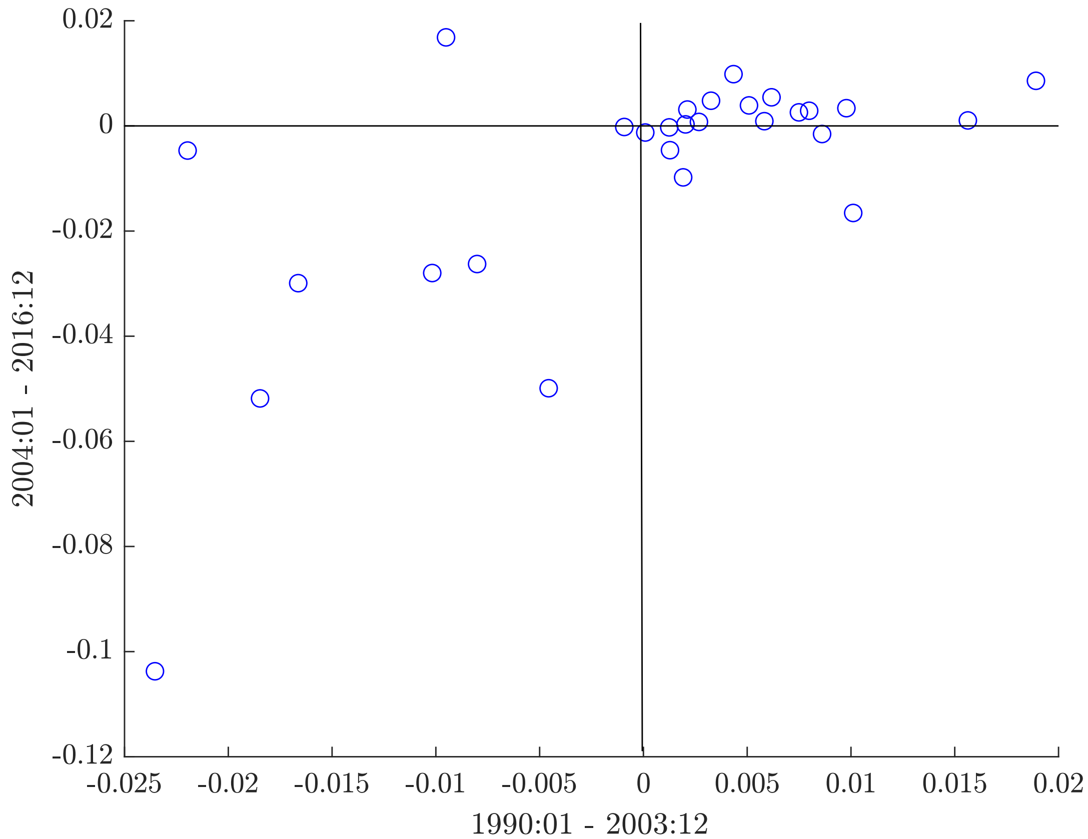
Model	λ_0	t_0	λ_M	t_M	λ_z	t_z	R^2_{OLS}
Panel A: Cross-sectional regressions							
Mkt + Mean	0.13	2.95	0.34	1.23	0.06	2.37	75.00
Mkt + Median	0.10	2.17	0.38	1.40	0.03	2.32	73.61
Mkt + Trimmed mean	0.13	2.93	0.33	1.23	0.05	2.41	74.95
Mkt + Weighted mean	0.13	2.95	0.34	1.24	0.06	2.36	74.98
Mkt + DMSFE ($\theta = 0.9$)	0.12	2.93	0.34	1.25	0.06	2.34	74.89
Mkt + DMSFE ($\theta = 0.7$)	0.12	2.87	0.35	1.27	0.06	2.34	74.92
Mkt + ABMA	0.13	2.95	0.34	1.23	0.06	2.37	75.01
Mkt + Subset (k = 2)	0.13	3.11	0.34	1.25	0.10	2.23	74.98
Mkt + Subset (k = 3)	0.13	3.16	0.34	1.24	0.14	2.22	74.89
Mkt + Subset (k = 4)	0.13	3.20	0.33	1.24	0.18	2.22	74.80
Mkt + Subset (k = 5)	0.14	3.27	0.33	1.25	0.22	2.14	74.70
Mkt + Subset (k = 6)	0.14	3.29	0.33	1.24	0.25	2.15	74.70
Mkt + Subset (k = 7)	0.14	3.32	0.33	1.24	0.28	2.14	74.62
Mkt + PC (IC = AIC)	0.12	3.00	0.35	1.29	0.40	2.18	74.34
Mkt + PC (IC = BIC)	0.11	2.65	0.39	1.44	0.27	1.95	74.62
Mkt + PC (IC = R^2)	0.13	3.12	0.34	1.29	0.41	2.19	74.14
Panel B: Restricted zero-beta rate ($\lambda_0 = 0$) cross-sectional regressions							
Mkt + Mean			0.43	1.65	0.06	2.71	71.08
Mkt + Median			0.44	1.67	0.04	3.05	71.19
Mkt + Trimmed mean			0.43	1.64	0.06	2.76	71.05
Mkt + Weighted mean			0.43	1.65	0.06	2.71	71.06
Mkt + DMSFE ($\theta = 0.9$)			0.44	1.65	0.07	2.70	71.05
Mkt + DMSFE ($\theta = 0.7$)			0.44	1.67	0.07	2.59	71.25
Mkt + ABMA			0.43	1.65	0.06	2.72	71.10
Mkt + Subset (k = 2)			0.44	1.69	0.11	2.52	70.86
Mkt + Subset (k = 3)			0.44	1.71	0.16	2.40	70.64
Mkt + Subset (k = 4)			0.44	1.72	0.20	2.37	70.42
Mkt + Subset (k = 5)			0.45	1.73	0.24	2.35	70.20
Mkt + Subset (k = 6)			0.45	1.75	0.27	2.28	70.11
Mkt + Subset (k = 7)			0.45	1.76	0.30	2.25	69.94
Mkt + PC (IC = AIC)			0.44	1.72	0.46	2.48	70.54
Mkt + PC (IC = BIC)			0.47	1.79	0.32	2.40	71.56
Mkt + PC (IC = R^2)			0.44	1.74	0.46	2.44	70.05

Notes. This table reports the risk premiums (in %) for each state variables (z), the 16 combination forecasts of commodity returns, next to the market (Mkt) portfolio based on the OLS cross-sectional regression:

$$\bar{R}_i = \lambda_0 + \lambda_M \hat{\beta}_{i,M} + \lambda_z \hat{\beta}_{i,z} + \alpha_i, \qquad i = 1, ..., N.$$

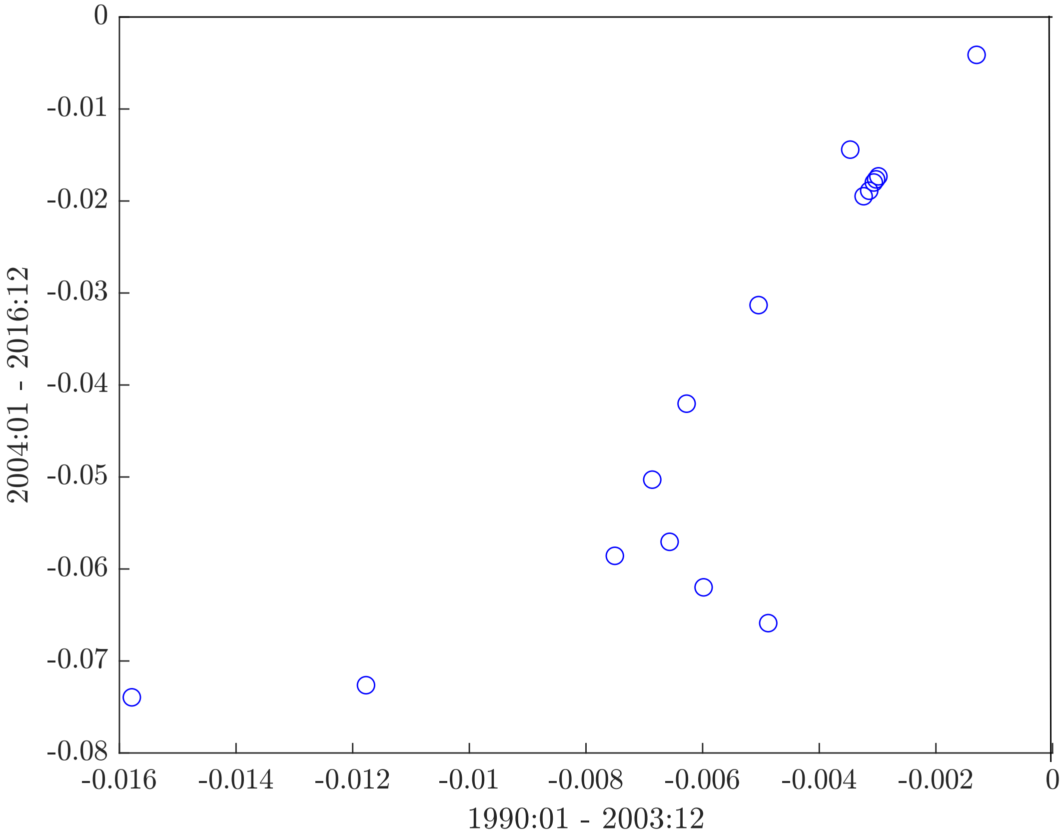
The testing assets (R_is) are the excess returns on the 23 individual commodity futures and the 25 equity portfolios formed on size and book-to-market. Panel A reports results for a regression of returns on exposures to the VAR(1) innovations in the state variables. Panel B restricts the intercept to zero. λ_0, λ_M and λ_z denote the zero beta rate, and risk premium for the market and each state variables, respectively. Next to the risk premium estimates are displayed t-statistics based on GMM standard errors corrected for heteroskedasticity, autocorrelation, and errors-in-variable bias. R^2_{OLS} represents the fraction of the cross-sectional variance of average excess returns on the test assets that is explained by the factor loadings associated with the model. The sample period is January 1990 to December 2016

Figure 1: Log Relative Mean Squared Forecast Error of Individual Predictive Model Forecasts



Notes. This figure plots the relative mean squared forecast error (MSFE) of the individual predictive forecasts of commodity futures returns. The log relative MSFE is defined as the lof of the ratio of the MSFEs of the predictive forecast to the MSFE of the historical average benchmark forecast. The out-of-sample period is from January 1990 to December 2016.

Figure 2: Log Relative Mean Squared Forecast Error of Combination Forecasts



Notes. This figure plots the log relative mean squared forecast error (MSFE) of the forecast combination of commodity futures returns. The relative MSFE is defined as the log of the ratio of the MSFEs of the predictive forecast to the MSFE of the historical average benchmark forecast. The out-of-sample period is from January 1990 to December 2016.