Commodity Futures Return Predictability and Intertemporal Asset Pricing

John Cotter[a,b], Emmanuel Eyiah-Donkor[c,*], Valerio Potì[a]

[a]*Michael Smurfit Graduate Business School, University College Dublin, Co. Dublin, Ireland*
[b]*UCLA Anderson School of Management, Los Angeles, USA*
[c]*Rennes School of Business, 2 Rue Robert D'Arbrissel, 35065 Rennes, France*

Abstract

We find out-of-sample predictability of commodity futures excess returns using combination forecasts 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 significant predictive power for future economic activity. Two-factor models featuring the market factor and the innovations in each of the combination forecasts explain a substantial proportion of the cross-sectional variation of both commodity and equity returns. The associated positive risk premia are consistent with the intertemporal capital asset pricing model (ICAPM), given how the combination forecasts predict an increase in future economic activity and a decline in stock market volatility 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 on commodities.

JEL classification: C22, C32, C53, G10, G11, G12, G13
Keywords: Commodity futures; Return predictability; Asset allocation; Business cycle; Intertemporal asset pricing

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 (Barsky and Kilian, 2002; Kilian,

*Corresponding author. *Telephone number:* +33626756191.
Email addresses: john.cotter@ucd.ie (John Cotter), emmanuel.eyiah-donkor@rennes-sb.com (Emmanuel Eyiah-Donkor), valerio.poti@ucd.ie (Valerio Potì)

This version: 1^{st} *August, 2022*

2008; Alquist et al., 2013). Also, interest in commodities as an alternative investment asset class has grown tremendously in recent years (Fuertes et al., 2010; Erb and Harvey, 2016).

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 - and whether such predictability can be explained in terms of time varying discount rates. We consider combination forecast methods (see, for example, Timmermann, 2006; Baumeister and Kilian, 2015) that combine forecasts from 28 individual predictive models. 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 14 combination forecasts of commodity excess returns ranging from simple averaging schemes of individual forecasts to more sophisticated ones. The data set consists of monthly observations for the period from January 1976 to December 2019.

Our study is related to the literature that investigates the time-series predictability of commodity returns (see Hong and Yogo, 2012; Gorton et al., 2013; Gargano and Timmermann, 2014; Ahmed and Tsvetanov, 2016; Hollstein et al., 2021; Rad et al., 2021; Conlon et al., 2022; and the references therein). For example, Hollstein et al. (2021) study the predictive power of a set of business cycle variables for the return and volatility of individual spot commodities over a century. Rad et al. (2021) examine the statistical and economic significance of individual commodity futures return predictability using a large set of 128 predictive variables using machine learning approaches.

Like Gargano and Timmermann (2014), we address the impact of parameter instability and uncertainty about return prediction models by implementing combination forecast methods. There are, however, important differences between our study and that of Gargano and Timmermann (2014). While we examine both time-series commodity return predictability and the asset pricing implications, their focus is only on the former. Another difference is that while the authors focus on forecasting the return on the Reuters/Jeffries-CRB spot indexes compiled by the Commodity Research Bureau, our interest is in the S&P GSCI total return index. The S&P GSCI is a widely adopted 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 et al., 2000; Jensen and Mercer, 2011; and Erb and Harvey, 2016). It also exhibits substantially different statistical properties from the CRB aggregate spot index. For example, the return correlation between the two indexes is about 45%. The return on the S&P GSCI total return index has average return and volatility closer to the stock market, while the return on the CRB spot index has lower average returns and volatility (see Table 2 of Bakshi et al., 2019 for more details).

Our empirical findings indicate that combination forecasts of commodity returns outperform the benchmark historical average return forecast by generating positive and statistically significant out-of-sample R^2 in the range 0.81%–3.77%. To explain the performance of the combination forecasts, we test their stability by conducting an in-depth visual comparison of their relative MSFEs. The results suggest that the ability of combination forecasts to diversify against the instability of the individual predictive models supports their superior forecasting performance. Also consistent with previous literature (for example, Welch and Goyal, 2008 for equities and Gargano and Timmermann, 2014 for commodity spot), we find that predictability is largely concentrated in economic recessions, with R^2 values as high as 20%.

We depart from previous studies of commodity return predictability in that, for all our forecasts of commodity returns, we check whether the return forecasts predict also macroeconomic activity. Rapach et al. (2010) and Gargano et al. (2017) already did this for equity return predictability and bond return predictability, respectively. We in turn carry out this check for commodity return predictability. This check is important as it allows us to asses whether, like for equities and bonds, we can ascribe at least part of commodity return predictability to time-variation in risk premia.

As argued by Cochrane (2005, 2007, 2011) and as considered by Rapach et al. (2010) in assessing equity return predictability, time-variation of risk premia is a plausible explanation for return predictability only if the predictors forecast changes in future investment opportunities (aggregate stock market return or stock market volatility and economic activity as proxied, for example, by the Aruoba et al. (2009) business condition index, the Chicago Fed national activity index, the log growth of industrial production, the smoothed recession probability of Chauvet (1998), the change in total capacity utilization, and the log growth in total nonfarm payroll employment). In a nutshell, the reason is that, in dynamic equilibrium, aggregate returns must reflect aggregate business conditions and vice versa. Thus the return forecasts should predict investment opportunities. We find that combination forecasts indeed predict an increase in stock market returns and economic activity and a decline in stock market volatility for the 1-, 3-, and 12-month horizons.

Taken together, the results demonstrate that variation in risk premia is a plausible explanation for commodity return predictability and that the superior predictive ability exhibited by combination forecasts is due to their ability to capture variation in risk premia (discount rates). This was shown by Rapach et al. (2010) for equity return predictability and by Gargano et al. (2017) for bond return predictability. Our study for the first time demonstrates that this results extends to commodity return predictability. This is important because, in conjunction with the conclusions drawn by Rapach et al. (2010) and Gargano et al. (2017) for equities and bonds, it suggests that variation in risk premia (discount rates) is a plausible explanation for return predictability for three of the most important asset classes. Casassus and Collin-Dufresne (2005) and, more recently, Fernandez-Perez et al. (2017) provided evidence of time-varying risk premia in commodity markets. Our study is the first to demonstrate its importance in explaining commodity return predictability.

Moreover, unlike Rapach et al. (2010) and Gargano et al. (2017), we check the plausibility of this explanation for return predictability by checking whether its implications hold in both the time-series and the cross-section of asset returns rather than only in the former. To do so, we specify, test and offer empirical support for a novel version of the ICAPM of Merton (1973) where the innovations in combination forecasts of commodity returns (state variables) act as risk factors capable of explaining variation of average excess returns in a broad cross-section of equity portfolios and individual commodity futures. An implication of the ability of combination forecasts to predict both aggregate commodity returns and future investment opportunities is that they would be valid state variables within Merton's (1973) intertemporal capital asset pricing model (ICAPM).[1] We test this implication by estimating two-factor models, that include innovations in each of the state variables next to the market factor, using the cross-section of 24 individual commodity futures and 25 equity portfolios formed on size and book-to-market as test assets. Our results show significantly positive risk premia associated with innovations in the state variables, consistent in sign with how the same state variables positively forecast aggregate stock market return and economic activity, and negatively predict stock market volatility. Our results thus imply consistency of the predictability in the time-series and the cross-section of asset returns, an important implication of the ICAPM framework of Merton (1973).

A further dimension along which our work contributes to the extant literature is that we address concerns raised by previous studies that statistical evidence of return predictability does not always translate into economic significance (Ahmed and Tsvetanov, 2016 for commodity futures returns; Della Corte et al., 2009 and Potì, 2018 for exchange rate returns; Thornton and Valente, 2012 and Sarno et al., 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. In this exercise, we assume a mean-variance investor with a relative risk aversion of three who forms optimal dynamic portfolios composed of the commodity futures index and risk-free T-bills, rebalancing such portfolios over time based on commodity return forecasts. We find that

[1]State variables that future investment opportunities (recession state variables) should command a risk premium since they are of hedging concern to investors.

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 2.90% per annum to hold the portfolios generated by the combination forecasts relative to the one generated by the benchmark forecast.

Motivated by the debate on in-sample versus out-of-sample predictability (see, for example, Inoue and Kilian, 2005), we evaluate evidence on the predictive ability of our forecasting models both in-sample and (pseudo) out-of-sample. However, since the empirical literature tends to emphasize out-of-sample predictability, to save space we do not tabulate results for in-sample predictability, though they are available upon request. Consequently, in what follows we focus our discussion on estimates and measures of (pseudo) out-of-sample predictability.

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 evidence of commodity futures excess return predictability and discuss its statistical and economic significance. Section 4 analyses the link between commodity return forecasts and the business cycle. In Section 5, we then present our ICAPM-type two factor model and test its implications for time-series and cross-sectional return predictability. Section 6 concludes.

2. Data and econometric methodology

This section describes the data on commodity futures and predictor variables, details the framework for generating the out-of-sample combination forecasts of commodity returns, and the criteria used to evaluate the forecasts.

2.1. Data on commodity futures and predictor variables

Our dataset contains monthly observations for the sample period January 1976 to December 2019. It includes primarily end-of-month total return data on the S&P GSCI,[2]

[2]Three S&P GSCI indexes are published: excess return, total return and spot. The excess return index measures the returns accrued from investing in uncollateralized nearby commodity futures, the

a fully investable commodity futures index, downloaded from Bloomberg. We use this index rather than individual commodities or more specialized commodity spot indexes because it is designed to resemble the total return on an investable portfolio of commodities. 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),[3] 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 return on S&P GSCI less the return on a one-month T-bill.[4] 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.[5]

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 list o predictors along with a brief description is provided below. Further details of the variables including their definition, the motivation for their use, and the relevant prior return predictability studies are summarized in Table A1 of the internet appendix.

The commodity and currency market variables we consider are

- Basis: we first calculate basis for each commodity that makeup the S&P GSCI, and then average the basis across all commodities to get an aggregate measure of basis;
- Global oil inventory (GOI): growth rate in global crude oil inventory. The inventory data used in calculating this variable is constructed by multiplying U.S. crude oil

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/what-we-do/global-markets/business-groups/sts-folder/gsci/components-weights-index-levels.html.

[3]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.

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

[5]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.

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;

- Global oil production (GOP): growth rate in global crude oil production;
- Commodity currencies: growth rate in the exchange rate of the currencies of Australia (AUS), Canada (CAN), New Zealand (NZ), South Africa (SA), and India (IND) 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 are

- Log dividend-price ratio (DP): 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;
- Return on S&P 500 index (S&P500): log return on the S&P 500 index;
- Treasury bill rate (TBL): interest rate on the U.S. 3-month Treasury bill (secondary market);
- Change in Treasury bill rate (CTBL);
- Long term return (LTR): the return on long-term government bonds;
- Term spread (TMS): long term government bond yield minus treasury bill rate;
- Change in term spread (CTMS);
- Yield spread (YS): Aaa-rated bond yield minus one-month T-bill rate;
- Change in default premium (CDFP): change in yield on Baa-rated bond minus yield on long-term government bond;

- Default return spread (DFR): long-term corporate bond returns minus long-term government bond returns.

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 et al. (2017) and Lin et al. (2017), and references therein. Like in these studies, we include these variables among the predictors because they are possible drivers of variation in discount rates. This way, we allow for commodity return predictability driven by variation in discount rates or, equivalently, in risk premia.

Finally, the macroeconomic variables we consider include

- Inflation (INFL): the log growth rate of U.S. consumer price index;
- Money supply (M1): log growth rate of U.S. narrow money;
- Unemployment rate (UNRATE): monthly unemployment rate;
- Industrial production (INDPRO): log growth rate of OECD aggregate industrial production;
- Capacity utilization in U.S. manufacturing: log growth rate of the degree of capacity utilization in U.S. manufacturing;
- Global real economic activity index (REA): the index is constructed from data on global dry cargo ocean shipping freight rates as described in Kilian (2009);
- Chicago Fed National activity index (CFNAI);
- OECD composite leading indicator (CLI), business confidence index (BCI), consumer confidence index (CCI).[6]

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 inflation,

[6]The sources of data for constructing the predictors are as follows: Basis and commodity currencies are from Bloomberg; GOI and GOP are from the U.S. Energy information administration; DP, SP500, TBL, LTR, TMS, YS, CDFP, DFR, and INFL are available on Professor Amit Goyal's website at http://www.hec.unil.ch/agoyal/; M1, UNRATE, CUTIL, REA, and CFNAI are from the St Louis Federal Reserve Economic Data at https://fred.stlouisfed.org/; INDPRO, CLI, BCI, and CCI are obtained from OECD data website, `https://data.oecd.org/`.

money supply, among others, have forecasting power for raw industrials and metals commodity index returns. Because many of these macroeconomic variables are published with some delays, we use two-month lags rather than one-month lagged values for inflation, money supply, industrial production, degree of capacity utilization in U.S. manufacturing and composite leading indicator. This is done to ensure that the data represents publicly available information.

Panel A of Table 1 presents descriptive statistics of monthly excess returns on the S&P GSCI for the full sample period. We report the annualized percent mean, standard deviation, minimum and maximum values, and first-order autocorrelation of excess returns. Panel B of the Table presents the summary statistics for the predictor variables. The majority of the predictors are strongly positively autocorrelated with first-order autocorrelation coefficients between 0.13 and 0.99, thus encouraging the use of test statistics that are robust to autocorrelation when testing for predictability.

[Insert Table 1 about here]

2.2. Combination forecast methods

Combination forecasts have been shown to perform well out-of-sample (see, for example, Stock and Watson, 2004 for output gap; Rapach et al., 2010 for equities; Baumeister et al.,2014 and Garratt et al., 2019 for crude oil). As suggested by Hendry and Clements (2004), there are several potential explanations for this. The common thread is that they provide a means to diversify against sampling error of the forecasting model parameters and model uncertainty. Sampling error of the model parameters, while unavoidable in any estimation exercise, is especially a concern when the estimated model is to be used to forecast financial asset returns, due to the high volatility of the variable to be forecast. Returns on investments involving commodities are likely to be no exception. Model uncertainty derives instead from possible model instability in the presence of 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 to.[7]

Our combination forecasts use individual forecasts as building blocks and differ in the way weights assigned to the individual forecasts are computed. We refer to their predictions as "*combination forecasts*". Generally, our approach to forming a combination forecast entails (i) estimating a regression of returns on each of the predictors, (ii) forming individual forecasts based on the estimated parameters from each of these 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 (3). A combination forecast at time $t+1$, $\hat{r}^{\mathrm{CF}}_{t+1}$, is a weighted average of the individual out-of-sample forecasts:

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

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.

To generate the individual predictive forecasts that enter (1), we estimate the following predictive regression models:

$$r_{t+1} = \alpha_i + \beta_i x_{i,\,t} + \varepsilon_{t+1}, \tag{2}$$

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 our 28 predictors at time t, and ε_{t+1} is a zero-mean error term. We generate out-of-sample forecasts using a recursive (expanding window) estimation scheme as follows. Let T observations be available for r_t and $x_{i,\,t}$. We divide the total sample into two parts: an in-sample estimation period containing the first $n = 168$ observations (January 1976 to December 1989) and an out-of-sample period containing the remaining $P = T - n = 360$ observations

[7]See Table A3 of the internet appendix where we document the poor out-of-sample performance of majority of the individual predictive model forecasts consistent with evidence for equities in Welch and Goyal (2008) and for commodities spot indexes in Gargano and Timmermann (2014).

(January 1990 to December 2019). Hansen and Timmermann (2012) warn 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. We thus choose the length of the in-sample estimation period so as to allow for a sufficiently long out-of-sample forecast evaluation period. The first out-of-sample forecast of commodity excess returns based on the $x_{i,t}$ predictor is given by

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

where $\hat{\alpha}_{i,n}$ and $\hat{\beta}_{i,n}$ are the ordinary least squares (OLS) estimates of α_i and β_i in (2), 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

$$\hat{r}_{i,n+2} = \hat{\alpha}_{i,n+1} + \hat{\beta}_{i,n+1} x_{i,\,n+1}, \tag{4}$$

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 $\{\hat{r}_{i,t+1}\}_{t=n}^{T-1}$.

We consider three types of combination forecasts: four simple combination forecasts, seven performance-based forecasts, and three factor model 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, $\hat{r}_{t+1}^{\text{Mean}}$, 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 (3). The trimmed mean forecast, $\hat{r}_{t+1}^{\text{Trimmed mean}}$, sets in (1) $w_{i,t} = 0$ 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, $\hat{r}_{t+1}^{\text{Median}}$, is the sample median of the 28 individual predictive model forecasts. The weighted-mean forecast ($\hat{r}_{t+1}^{\text{Weighted mean}}$) proposed by Bates and Granger (1969) specifies the combination weights to be proportional to the inverse of the estimated residual variance, $\sigma_{i,t}^2$, for the individual predictive regression models given by (2),

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

The second set of combination forecasts 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,

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

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). It attaches more weight to the recent forecasting accuracy of the individual predictive models, thereby allowing time-variation in the data generating process of returns. 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

[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.

further back in time. Rapach et al. (2010) 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 et al. (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 forecast, with its firm information-theoretic rationale, tents to be 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 et al. (2013), based on equally-weighted averages of all forecasts predictive regression models that include a fixed number of the predictor variables. This approach 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. Given N potential predictors, a subset regression combination is 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 set $k = 5$. One has then to average over the $C_k^N = N!/(k!(N-k)!)$ subset regression combination forecasts, 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 $\dim(x_{i,t}) = k$.

Our final combination forecasts follows Stock and Watson (2002a,b). We generate out-of-sample forecasts by estimating a predictive regression model based on a diffusion index that assumes a latent factor structure:

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

where $F_{k,t}$ is the kth principal component extracted from our 28 predictor variables. Diffusion indexes provide a convenient way of extracting common factors from a large number of potential predictor variables. Neely et al. (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. Forecast evaluation

2.3.1. Statistical performance measures

Our first measure of predictability is the widely used Campbell and Thompson (2008) out-of-sample R^2 statistic, R^2_{OS}, given by:

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

where r_{t+1} is the realized log return at time $t+1$, $\hat{r}_{t+1|t}$ is a combination 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).[9] We refer to it as the historical average (HA) model.[10] We evaluate the statistical significance of the R^2_{OS} statistic using the p-value of the *MSFE-adjusted* statistic of Clark and West (2007). The statistic tests the null hypothesis that the HA forecast MSFE is less than or equal to the competing combination forecast MSFE against the one-sided (upper-tailed) alternative hypothesis that the HA forecast MSFE is greater than the combination forecast MSFE. Under the null of no-predictability, the HA return forecast is expected to have a lower MSFE.[11]

Following Stock and Watson (2003), we assess the stability of our forecasts by examining their relative MSFEs, defined as the ratio of the MSFE of a predictive forecast to the MSFE of the HA benchmark forecast. 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 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.

2.3.2. Economic performance measures

A limitation to the R^2_{OS} measure is that it does not explicitly take into account the risk that an investor would have to bear over the out-of-sample period. Also, studies such as Della Corte et al. (2009) and Potì (2018) for exchange rate returns and Thornton and

[9] 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)

[10] This means 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_{OS} statistic measures the proportional reduction in mean squared forecast error (MSFE) for a combination forecast relative to the HA forecast. A positive R^2_{OS} implies the combination forecast outperforms the HA forecast as it has a lower MSFE.

[11] None of the currently available tests of equal predictive accuracy, including the test of Clark and West (2007), are suitable for testing pseudo out-of-sample forecasting performance. In line with the purpose of this paper, our aim is not to compare (pseudo) out-of-sample forecasts but rather to compare a null model that, in a reduced form representation of the data generating process, includes the given predictive variable to a null model that does not include it. The Clark and West (2007) test is suitable for this purpose and hence the reason why we report statistical significance based on the p-values of the test. See Diebold (2015) and Kilian (2015) for a discussion of the difference between comparing forecasts and comparing models.

Valente (2012) and Sarno et al. (2016) for bond return predictability find that statistical evidence of return predictability does not always translate into economic significance. To address these concerns, we follow Campbell and Thompson (2008) and consider a mean-variance investor who monthly allocates her wealth between commodities futures and risk-free T-bills using the combination forecasts or HA forecasts 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 to 3 and allow for a moderate portfolio leverage of 50%, similar to other studies such as Campbell and Thompson (2008) and Rapach et al. (2016). Since we use futures, we do not need to impose short-sales constraint.

To evaluate the performance of the portfolios generated by the 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 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 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 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

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

where $\bar{r}_p^{\text{CF}}$ and $\bar{r}_p^{\text{HA}}$ are the portfolio mean returns of the combination and HA portfolio strategies, respectively, and TO^{CF} and TO^{HA} are their respective average turnover. In comparing a dynamic portfolio based on a combination forecast to that of a 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.

As a final performance measure, we estimate time-series regressions of the portfolio excess returns generated by each of the combination forecast, $r_{p,t+1}^{\text{CF}}$, onto the excess returns on the S&P GSCI, r_{t+1}^{SPGSCI}, as

$$r_{p,t+1}^{\text{CF}} = \alpha + \beta r_{t+1}^{\text{SPGSCI}} + e_{p,t+1} \tag{15}$$

A failure to reject the null hypothesis $H_0 : \alpha = 0$ will indicate no excess return gain from return predictability. Conversely, a positive and significant α will demonstrate statistically and quantify economically the gains from return predictability. We also estimate an augmented version of the regression in (15) that includes a recession dummy proxied by the NBER-dated recession indicator. Similarly, a positive and significant α will demonstrate statistically and economically the gains from return predictability in periods of recessions versus expansions.

3. Empirical results

3.1. Out-of-sample forecasting results

Table 2 summarizes the performance of the combination forecasts relative to the HA benchmark forecast for one-month ahead forecast of commodity futures returns. The table reports the MSFE, R^2_{OS} as defined in (10) and *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 $R^2_{OS} > 0$ is assessed using the p-value of the *MSFE-adjusted* statistic of Clark and West (2007).

From the table, we can see that, except the median combination forecast, all combination forecasts of commodity futures returns outperform the HA forecast in terms of MSFE and supportive of return predictability. The R^2_{OS} generated by each of the combination forecasts is in some cases impressive, ranging from 0.81% for the Trimmed mean combination forecast to 3.77% for the PC (IC = BIC) combination forecast. All the combination forecasts have R^2_{OS} significantly greater than zero at the 1% level except the Trimmed mean combination forecast which has R^2_{OS} significantly greater than zero at the 5% level and the Median combination forecast which is not statistically significant.[12]

[Insert Table 2 about here]

Table 3 presents results for the economic significance of return predictability as measured by the CER gains, Sharpe ratios, portfolio turnover ratios, break-even transaction

[12] Our predictability findings are robust to using an alternative commodity futures index, the Bloomberg commodity total return index. These results are available upon request.

costs, and alphas. The mean-variance investor's relative risk-aversion coefficient is set equal to 3, a moderate portfolio leverage of 50% is allowed, and transaction costs are set to 20 basis points.

From Column 6 of the table, we see that positive annualized CER gains are realized for all combination forecasts except the PC (IC = AIC) combination forecast. The Subset ($k = 2, ..., 5$) and the PC (IC = BIC) combination forecasts all record annualized CER gains well above the 2% level. These findings are consistent, in terms of the extent to which they point to a greater economic value of combination forecasts compared to the HA forecast suggested by the statistical tests reported in Table 2. Consistent with these findings, annualized Sharpe ratios of the commodity portfolios generated by the combination forecasts are higher than for the portfolio that relies on the historical average return forecast. These results are also consistent with the strong statistical performance of the combination forecasts.

Column 7 of Table 3 reports the annualized CER gains net of transaction costs. Accounting for the effect of transaction costs does not erode the performance of the portfolios based on the majority of the combination forecasts. They continue to deliver positive CER gains net of transaction costs. The break-even transaction costs values are also much higher than the actual proportional transaction cost meaning that investors would prefer the portfolios based on the combination forecasts. This performance comes at the cost of a somewhat higher average turnover. For example, the Subset ($k = 5$) combination forecast deliver annualized CER gains net of costs of 2.10% compared to 2.90% without transaction costs. The relative magnitude of Sharpe ratios for combination forecasts is consistent with the relative magnitude of CER gains when both are calculated net of transaction costs.

The final column of Table 3 reports the annualized alpha estimates and Newey and West (1987) adjusted t-statistics (in parentheses) from regressions of the portfolio excess returns generated by each of the combination forecasts onto the excess returns on the S&P GSCI. The results show positive and statistically significant annualized alphas for all combination forecasts. The annualized alphas range from 3.90% (t-statistic of 38.71) for the median combination forecast to 7.25% (t-statistic of 9.39) for the PC (IC = BIC) combination forecast. This analysis confirms the statistical and economic gains from return predictability.

[Insert Table 3 about here]

3.2. Stability of out-of-sample return forecasts

As earlier hypothesized, combination forecasts work well to improve forecast accuracy as reported in Table 2 because they diversify against parameter instability of individual predictive models. We assess the stability of our combination 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 2004 and the second period from January 2005 to December 2019, and compute the relative MSFE of each model (that is, 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 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 combination forecasts 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 combination forecast. If the forecasts were stable, we should expect the points to be scattered around the third (southwest) quadrant. As can be seen from the figure, all points plot in the third (southwest) quandrant and show considerable stability, indicating improved forecasting performance compared to the HA benchmark forecast in both out-of-sample periods.

[Insert Figure 1 about here]

4. Out-of-sample return forecasts and business cycles

4.1. Statistical performance of return forecasts in expansions and recessions

Studies such as Rapach et al. (2010), Henkel et al. (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, we compute the following version of the conventional R^2 statistic for business cycle expansions (EXP) and recessions (REC):

$$R_{\mathrm{C}}^2 = 1 - \frac{\sum_{t=1}^{T} I_t^c \hat{\varepsilon}_{i,t}^2}{\sum_{t=1}^{T} I_t^c (r_t - \bar{r})^2} \quad \text{for } c = \text{EXP}, \text{ REC}, \tag{16}$$

where I_t^{EXP} (I_t^{REC}) is an indicator function that takes a value of one when month $t-1$ 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 (2), $\bar{r}$ is the full sample mean of r_t, and T is the full sample observations. We compute R^2 statistics for business cycle expansions (EXP) and recessions (REC), for both in-sample and out-of-sample forecasts, using in (16) in-sample and out-of-sample forecast errors, respectively.

While, to save space, we do not report the results for the in-sample forecasts (see Table A2 of the internet appendix), we report those for the out-of-sample forecasts in Table 4. The table 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}) for the forecast combinations. None of the forecasts exhibit statistically greater than zero R^2_{OS} during expansions. However, during recessions, each combination forecasts delivers R^2_{OS} values ranging from 0.36% to 15.05% which are statistically greater than zero at the 1% level and, in some cases, at levels close to or beyond 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 of Gargano and Timmermann (2014).

To sum up, predictability from forecast combinations is stronger in recessions compared to expansions and outperform the benchmark HA forecast. Therefore, these results suggest that commodity predictability is a phenomenon largely associated with recessions which can be captured by combination forecasts.

[Insert Table 4 about here]

4.2. Economic performance of return forecasts in expansions and recessions

We now examine whether the performance of the portfolios that can be generated using the combination forecasts of commodity return 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 market portfolio 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 5 reports annualized Sharpe ratios (net of transaction costs of 20 basis points) computed separately for expansions and recessions. 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. As shown in the table, 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 5 also reports estimates of economic significance as measured by out-of-sample annualized CER gains (net of proportional transaction costs of 20 basis points) separately for business cycle expansions and recessions. Overall, the above 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 for all the combination forecasts.

The final column of Table 5 reports the annualized alpha estimates and Newey and

West (1987) adjusted t-statistics (in parentheses) from regressions of the portfolio excess returns generated by each of the combination forecasts onto the excess returns on the S&P GSCI and a recession dummy proxied by the NBER-dated recession indicator. The results show positive and statistically significant annualized alphas for all combination forecasts. The annualized alphas range from 3.82% (t-statistic of 36.41) for the median combination forecast to 8.56% (t-statistic of 14.29) for the PC (IC = BIC) combination forecast. This analysis confirms our earlier findings that the statistical and economic gains from return predictability are stronger for business cycle recessions.

These results taken together are consistent with the possibility that commodity return predictability derives from time-variation of risk premia. An alternative possibility is that, as suggested by Cujean and Hasler (2017), the greater predictability during recessions stems from a greater dispersion of beliefs among investors than during expansions. The mechanism advocated by these authors, in a nutshell, is that the greater dispersion of beliefs renders the beliefs of the representative investor more reactive to past returns. Since these two explanations are both plausible, we now undertake further investigation to assess their relative merits.

[Insert Table 5 about here]

5. State variables, investment opportunities, and intertemporal pricing

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 also forecast changes in future economic activity. In this case, according to the ICAPM framework of Merton (1973),[13] the predictors could be considered proxies for state variables that capture changes in future investment opportunities driven by aggregate stock market return, stock market volatility returns, or economic activity.

Also, if combination forecasts forecast commodity returns because they proxy for state variables in the ICAPM sense, they should predict changes in future investment

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

opportunities, 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. This, in turn, would imply that combination forecasts outperform because they proxy for state variables that drive variation in commodity risk premia and hence in their expected returns.

We test this explanation of the performance of combination forecasts by examining whether combination forecasts of commodity returns have predictive power for future investment opportunities, hence whether they can be treated as valid state variables within the ICAPM framework. This amounts to testing an implication of the ICAPM for the relation between predictability of commodity returns and predictability of future investment opportunities, a necessary condition for a state variable to be priced in the cross-section. 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 (other than the market) should obey sign restrictions if the factors are state variables that predict future investment opportunities in time-series regressions (see applications in Maio and Santa-Clara, 2012 and Boons, 2016). As such, if a state variable forecasts an increase in expected aggregate stock market returns or economic activity and a decline in stock market volatility, its innovation should earn a positive risk premium in the cross-section.

The intuition is that if an asset covaries positively with future expected market returns (because it covaries positively with a state variable that forecasts an increase in future aggregate market returns), it offers high returns when market returns are expected to be higher. Such an asset does not provide a hedge for adverse changes in future market returns, and hence, a risk-averse investor will require a positive risk premium to invest in such an asset, implying a positive price of risk for the hedging factor (innovations in the state variable). A similar argument applies to assets that covary with economic activity and stock market volatility. 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 subsections that follow.

5.1. Predicting investment opportunities

To test whether a state variable forecasts future aggregate stock market return or stock market volatility, we estimate the following pair of predictive regressions,

$$r_{M,t+1:t+h} = \alpha_i + \beta_i z_{i,t} + \varepsilon_{t+1:t+h}, \tag{17}$$

$$\sigma^2_{M,t+1:t+h} = \alpha_i + \beta_i z_{i,t} + \epsilon_{t+1:t+h}. \tag{18}$$

The dependent variable in (17) is the continuously compounded excess return on the market portfolio from month $t+1$ to $t+h$, $r_{M,t+1:t+h} = r_{M,t+1} + ... + r_{M,t+h}$. The dependent variable in (18) is the sum of log monthly realized market excess return variance from month $t+1$ to $t+h$, $\sigma^2_{M,t+1:t+h} = \sigma^2_{M,t+1} + ... + \sigma^2_{M,t+h}$, where $\sigma^2_{M,t+1}$ is computed as the sum of squared daily market excess returns in month $t+1$. The proxy for $r_{M,t+1}$ is the CRSP value-weighted market return in excess of one-month T-bill rate. The data is downloaded from Professor Kenneth French's data library. h is the forecast horizon corresponding to 1, 3, and 12 months ahead and $z_{i,t} = \hat{r}^{\text{CF}}_{t+1}$ is one of the 14 combination forecasts of commodity returns (state variable) at a time. To evaluate the statistical significance of the regression coefficients, we use Newey and West (1987) t- statistics with $h-1$ lags, allowing us to correct for the serial correlation in the residuals caused by the use of overlapping data.

The results for the predictive regressions for stock market excess returns in (17) are presented in Panel A of Table 6. The slope coefficient estimates, β, indicate that at the 1 and 3 month horizons, except the median combination forecast, all state variables have positive predictive power for market excess returns. Of these, only the PC (IC=AIC) and PC (IC=BIC) are marginally statistically significant at the 10% level. The weak forecasting power of the state variables for excess market returns is also captured by the very low R^2 values. At the 12-month horizon, eleven of the fourteen state variables forecast negative market excess return, although the associated slope estimates are not significant. The above results are in agreement with Maio and Santa-Clara (2012) and Cooper and Maio (2019) who document similar findings.

The results for the predictive regressions for stock market volatility in (18) are presented in Panel B of Table 6. We can see there is stronger evidence of predictability for future stock market volatility across the majority of the state variables. Almost all the state variables have negative and statistically significant predictive power for stock market volatility at all horizons. The exceptions are PC (IC=AIC), PC (IC=BIC), and PC (IC=R^2) which forecast a decline in stock market volatility with statistically insignificant t-statistics at the 12-month horizon.

[Insert Table 6 about here]

Considering the mixed findings from the predictive regressions for stock market returns documented previously, we also investigate whether the state variables predict future economic activity. Our motivation for this exercise relies on Roll (1977). The argument is that besides stocks and bonds, most investors own other non-marketed assets, such as labour income, which are related to economic activity. Therefore, aggregate stock market 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. Hence, economic activity is likely to be a better proxy for the unobservable return on the aggregate wealth portfolio. To assess the predictor power of the state variables for economic activity, we estimate the following regressions,

$$y_{t+1:t+h} = \alpha_i + \beta_i z_{i,t} + e_{t+1:t+h}, \tag{19}$$

where $y_{t+1:t+h} = y_{t+1} + ... + y_{t+h}$ is the economic activity variable. As proxies for y_t, we use the Aruoba et al. (2009) business condition index (ADSI), the Chicago Fed national activity index (CFNAI), the log growth of U.S. industrial production (IP), the smoothed recession probability (SRP) of Chauvet (1998), the change in total index capacity utilization (TCU), and the 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 et al. (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 an additional one-month to account for delays in the release of such data.

Table 7 reports estimation results of the predictive regression in (19) for the forecasting horizons of 1, 3, and 12 months ahead. From the table, we can see that all the state variables display strong predictive content for future economic activity at all horizons. They forecast significant increases in ADSI, CFNAI, IP, and TCU, and a decline in SRP. For all state variables, t-statistics for the significance of β and the R^2 values rise from the 1-month to 3-month horizons, and then falls for the 12-month horizon. For example, the PC (IC = BIC) state variable predict significant increases in ADSI with R^2 values that rise from 21.84% for the 1-month horizon to 26.35% for the 3-month horizon, and then falls to 13.06% for the 12-month horizon. Of all the economic activity variables, only the results for the PAYEMS amount to mixed findings about the predictive power of the state variables.

[Insert Table 7 about here]

Finally, we test whether the S&P GSCI excess return has predictive power for economic activity. We run predictive regressions similar to (19) of business condition variables on the S&P GSCI excess return at the 1-month horizon. Finding that the S&P GSCI excess returns does not predict the business cycle would reinforce our claim that return predictability is important and that combination forecasts of excess returns on the S&P GSCI (because they significantly predict business conditions as shown in Table 7) are needed to capture this predictability. The results from this exercise reported in Table 8 show that the S&P GSCI excess return does not have predictive power for a battery of business condition variables, including the return on the return on the S&P 500 index, dividend yield on the S&P 500, Term spread, Default spread, Unemployment rate, the Smoothed recession probability of Chauvet (1998), the Aruoba et al. (2009) business condition index, log growth in industrial production index, change in total capacity utilization, and the log growth in total nonfarm payroll employment.

[Insert Table 8 about here]

To sum up, the results in this subsection indicate that the candidate state variables have significant predictive power for future stock market volatility and economic activity for horizon ranging from 1 to 12 months. While there is evidence that state variables positively predict stock market excess returns for the 1- and 3-month horizons, the evidence is mixed for the 12-month horizon. These results are therefore consistent with the predictions of the ICAPM framework.

5.2. 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, and as shown in the previous subsection, since combination forecasts predict a significant increase in future economic activity and a significant decline in stock market volatility, its innovation should be priced with a positive risk premium in the cross-section to be consistent with the ICAPM framework. To test this implication, we estimate two-factor asset pricing models that include each individual state variable next to the market portfolio. We assume that asset excess returns are governed by the discrete-time version of the ICAPM

$$E(R_i) = \lambda_M \beta_{i,M} + \lambda_z \beta_{i,z}, \ \forall i, \tag{20}$$

where $E(R_i)$ is the expected excess return on test asset i, λ_M and λ_z denote the risk premia associated with the market factor and the innovations in state variable z (one of the 14 combination forecast of commodity returns at a time), respectively, whereas $\beta_{i,M}$ and $\beta_{i,z}$ represent the assets' loadings with respect to the market factor and the innovations in the state variable, respectively. The betas are the slope coefficients from the return-generating process

$$R_{i,t+1} = \alpha_i + \beta_{i,M} R_{M,t+1} + \beta_{i,z} u_{z,t+1} + \varepsilon_{i,t+1}, \;\; \forall i, \tag{21}$$

where $R_{i,t+1}$ is the excess return on test asset i at the end of month $t+1$, $R_{M,t+1}$ is the excess return on the market portfolio at the end of month $t+1$, and $u_{z,t+1}$ is the innovation in state variable z at end of month $t+1$. The market factor is the raw monthly return on the Center for Research in Security Prices (CRSP) value-weighted stock market portfolio in excess of the one-month Treasury bill rate. The data is downloaded from Professor Kenneth French's Data library.

Following Campbell (1996) and Petkova (2006), the time-series dynamics for each state variable (combination forecast of commodity returns) is specified as a first-order vector autoregressive, VAR(1), process

$$\begin{pmatrix} R_{M,t+1} - \mu_M \\ z_{t+1} - \mu_z \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} R_{M,t} - \mu_M \\ z_t - \mu_z \end{pmatrix} + \begin{pmatrix} u_{M,t+1} \\ u_{z,t+1} \end{pmatrix}, \tag{22}$$

where $\boldsymbol{A}$ is the companion matrix of the VAR, and μ_M and μ_z are the sample averages of the market portfolio excess return and state variable, respectively. The estimated residuals, $\hat{u}_{z,t+1}$, represent the innovations in the state variable which we use as our proxy for the intertemporal risk factor in (21).

To test our two-factor ICAPM in (20), we run cross-sectional regressions using the two-step Fama and MacBeth (1973) procedure. In the first step, full-sample factor loadings are estimated from the following time-series regression for each test asset:

$$R_{i,t+1} = \alpha_i + \beta_{i,M} R_{M,t+1} + \beta_{i,\hat{u}_z} \hat{u}_{z,t+1} + \varepsilon_{i,t+1}, \;\; \forall i, \tag{23}$$

where $\hat{u}_{z,t+1}$ represent the estimated innovations in the state variable from (22). In the second step, we estimate a single OLS cross-sectional regression by relating the average excess returns of all test assets to their estimated betas in the first step,

$$\bar{R}_i = \lambda_0 + \lambda_M \hat{\beta}_{i,M} + \lambda_{\hat{u}_z} \hat{\beta}_{i,z} + \alpha_i, \ \forall t, \tag{24}$$

to obtain estimates for the zero-beta rate, λ_0, vector of factor risk premia, (λ_M, λ_z), and pricing errors, α_i. $\bar{R}_i$ represents the average excess return for asset i.

The betas in the time-series regression in (23) are estimated and therefore represent generated regressors in (24). We therefore test the significance of the factor risk premia estimates using t-statistics computed with GMM standard errors that correct for errors-in-variables bias, heteroskedasticity and autocorrelation in the error term (see Cochrane, 2005). The number of lags to include in the computation of the standard errors is based on the Bartlett kernel with Newey and West (1994) optimal bandwidth selection.

Although our focus is the factor risk premia estimates, we also assess the fit of the models based on each state variable by computing the cross-sectional OLS coefficient of determination (see Campbell and Vuolteenaho, 2004; Fernandez-Perez et al., 2017; Maio and Philip, 2018; among others),

$$R^2_{\text{OLS}} = 1 - \frac{\text{Var}_N(\hat{\alpha}_i)}{\text{Var}_N(\bar{R}_i)}, \tag{25}$$

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 cross-sectional regression, R^2_{OLS} can assume negative values.[14]

As test assets, we use the monthly returns on 24 individual commodity futures obtained from Bloomberg 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 for a total of 49 test assets. The sample period is January 1990 to December

[14]Negative R^2_{OLS} estimates indicate that the regression including the risk factors perform worse in terms of explaining cross-sectional variation in average excess returns than a simple regression model that predicts constant expected returns.

2019 and matches the out-of-sample forecast evaluation period. The commodities cover twelve agricultural products (cocoa, coffee C, corn, cotton no.2, frozen concentrated orange juice, oats, rough rice, soybean meal, soybean oil, soybeans, sugar no.11, wheat), four energies (gasoil, heating oil, WTI crude oil, Brent crude oil), three livestock (feeder cattle, lean hogs, live cattle), four metals (gold, silver, palladium, platinum), and lumber. Returns are computed for each commodity using the front-end contracts up to one month before the maturity date; the positions are then rolled to the second-nearest contract, and so on. We use individual commodities as opposed to commodity futures portfolios because the cross-section of commodities is small, which limits the number of portfolios that can be formed. The 25 portfolios are widely used in cross-sectional asset pricing tests and represent one of the most challenging set of portfolios in the asset pricing literature (see Petkova, 2006). Our justification for expanding the set of test assets to include stocks is accord with the ICAPM which posits that state variables in the model must be state variables for all assets, and not just equities or commodities.

Table 9 presents results for monthly cross-sectional regressions of (24) using the 49 test assets (24 individual commodities and the 25 equity portfolios formed on size and book-to-market). The table reports the unconditional monthly percent 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 (Mkt).

The results for the cross-sectional regressions that restricts the intercept, λ_0, to zero as dictated by the asset pricing model in (20) are reported in Panel A of Table 9. We can see that all risk premia estimates associated with innovations in each of the fourteen state variables are positive and statistically significant at the 1% or 5% levels. The positive risk premia estimates are consistent with their corresponding positive (negative) slope coefficient estimates in the predictive regressions for stock market returns and economic activity (stock market volatility) analysed in the previous subsection. The loadings on the market portfolio also represent a significant risk factor in the cross-sectional tests. Estimates of the market risk premium of about 7.5% are typical values that are close to the sample average market portfolio excess return. All models, with innovations in each state variable next to the market portfolio, have reasonable explanatory power for the cross-section of the 49 testing asset returns with R^2 estimates ranging from 23.42% to 32.12%.

In terms of explanatory power of the CAPM, we have the usual results that it cannot explain the cross-section of asset returns as indicated by the negative R^2 estimate of -57.83%. We draw a similar conclusion to that of the CAPM for the model that includes the innovations in excess return on the S&P GSCI next to the market portfolio (Mkt + S&P GSCI) with R^2 estimate of -65.56%. The risk premium estimate for innovations in the S&P GSCI excess return is negative and insignificant, an indication that combination forecasts of commodity excess return is needed to explain the cross-section of returns on the 49 testing assets.

As a robustness check, we also report results for a version of the two-factor model which includes an intercept in the pricing equation. Essentially, if a given model is correctly specified, the intercept in the pricing equation should be statistically indistinguishable from zero. Also, the estimates of the factor risk premia and the cross-sectional R^2 should not be too different from the corresponding estimates in the model without intercept.

The results for the cross-sectional regressions which include an intercept term in (20) are reported in Panel B of Table 9. Since the dependent variables in the cross-sectional regressions are excess returns, the intercept, λ_0, should be zero. We fail to reject this hypothesis for the CAPM and the Mkt + S&P GSCI model. For example, the CAPM intercept of 0.48% per month (approximately 5% per annum) is statistically significant at the 5% and are economically meaningful. The results show evidence of misspecification for these two models.

For the ICAPM models based on the innovations in each of the fourteen combination forecasts state variables, all the intercepts are not statistically significant. The estimates of the risk premia for all state variables, although insignificant, are similar in magnitude to those from cross-sectional tests that restrict the intercept to zero reported in Panel A of Table 9. The results further reveal that although loadings on the market portfolio does not represent a significant factor in the cross-section of returns, their estimated market risk premiums are typical values that are close to the sample average of the market portfolio excess returns and its presence in the models improve their explanatory power.

Overall, models that include both the excess return of the market portfolio and innovations in the state variables (all our 14 combination forecasts of commodity returns) cannot be rejected over the 1990 to 2019 period. The results show that assets' covariances with state variable surprise factors are important in the cross section of average returns. Our results further demonstrate consistency in terms of sign between how the state variables predict an increase in future stock market returns and economic activity and a decline in stock market volatility in the time-series and command positive risk premiums in the cross-section of returns, an implication of the ICAPM. This is the case whether we restrict the intercept to zero in cross-sectional tests as dictated by the asset pricing model in (20) or include an intercept term.

[Insert Table 9 about here]

6. Conclusions

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 14 combination forecasts of commodity futures returns constructed from forecasts of individual predictive models and take into account both parameter instability and model uncertainty of the individual predictive models.

We find that all combination forecasts of commodity returns outperform a benchmark historical average return forecast both statistically and economically in out-of-sample predictability tests. Forecast stability analysis using relative mean squared forecast errors show strong evidence of stability of combination forecasts commodity futures excess returns. The superior forecasting performance of the combination forecasts can therefore be attributed to their ability to diversify against parameter instability and model 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 for stock market excess returns, stock market volatility, and economic activity (proxied by the Aruoba et al. (2009) business condition index, the Chicago Fed national activity index, log growth in industrial production index, the smoothed recession probability of Chauvet (1998), change in total capacity utilization, or log growth in total nonfarm payroll employment) at horizons ranging from 1 to 12 months.

Finally, we provide evidence consistent with the ICAPM framework of Merton (1973) 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 premia associated with the state variables that predict an increase in future stock market returns and economic activity and a decline in stock market volatility. These results provide an explanation for the significant out-of-sample performance of the combination forecasts, implying it derives from the ability of the combination forecasts to pickup 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. volume 2, pp. 427–507.

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.

Barsky, R.B., Kilian, L., 2002. Do we really know that oil caused the great stagflation? a monetary alternative. NBER Macroeconomics Annual 2001 16, 137–183.

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

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., Lee, T.K., 2014. Are there gains from pooling real-time oil price forecasts? Energy Economics 46, S33–S43.

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

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.

Casassus, J., Collin-Dufresne, P., 2005. Stochastic convenience yield implied from commodity futures and interest rates. The Journal of Finance 60, 2283–2331.

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

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.

Conlon, T., Cotter, J., Eyiah-Donkor, E., 2022. The illusion of oil return predictability: The choice of data matters! Journal of Banking & Finance 134, 106331.

Cooper, I., Maio, P., 2019. Asset growth, profitability, and investment opportunities. Management Science 65, 3988–4010.

Cujean, J., Hasler, M., 2017. Why does return predictability concentrate in bad times? The Journal of Finance 72, 2717–2758.

Della Corte, P., Sarno, L., Tsiakas, I., 2009. An economic evaluation of empirical exchange rate models. The review of financial studies 22, 3491–3530.

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.

Diebold, F.X., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. Journal of Business & Economic Statistics 33, 1–1.

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., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

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.

Garratt, A., Vahey, S.P., Zhang, Y., 2019. Real-time forecast combinations for the oil price. Journal of Applied Econometrics 34, 456–462.

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

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.

Hollstein, F., Prokopczuk, M., Tharann, B., Simen, C.W., 2021. Predictability in commodity markets: Evidence from more than a century. Journal of Commodity Markets 24, 100171.

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.

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., 2008. The economic effects of energy price shocks. Journal of economic literature 46, 871–909.

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., 2015. Comment. Journal of Business & Economic Statistics 33, 13–17.

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 , 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.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61, 631–653.

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

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

Petkova, R., 2006. Do the fama–french factors proxy for innovations in predictive variables? The Journal of Finance 61, 581–612.

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

Rad, H., Low, R.K.Y., Miffre, J., Faff, R.W., 2021. The commodity risk premium and neural networks. Available at SSRN 3816170 .

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.

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., 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.

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.

Table 1: Summary statistics for returns and predictor variables

Variable	Mean	Standard deviation	Min	Max	Auto-correlation
Panel A: Excess returns					
S&P GSCI	1.70	19.03	−28.28	22.34	0.16
Panel B: Predictor variables					
Basis	−7.82	3.46	−4.15	5.85	0.68
GOI	5.01	59.54	−72.61	73.70	−0.48
GOP	1.00	4.93	−9.49	6.50	−0.08
DP	−4 411.30	149.16	−452.36	−275.33	0.99
SP500	8.13	14.79	−24.54	12.38	0.03
TBL	4.47	3.54	0.01	16.30	0.99
CTBL	−0.01	0.45	−4.62	2.61	0.36
LTR	8.62	10.91	−11.24	15.23	0.04
TMS	2.13	1.45	−3.65	4.55	0.95
CTMS	0.00	0.46	−3.28	4.23	0.10
YS	7.03	2.58	2.82	14.38	0.99
CDFP	0.00	0.29	−1.20	1.39	−0.12
DFR	0.23	5.12	−9.75	7.37	−0.03
INFL	3.51	1.26	−1.92	1.52	0.62
M1	5.96	2.72	−3.37	5.81	0.13
UNRATE	74.90	5.64	3.50	10.80	0.99
INDPRO	1.97	2.13	−3.87	1.98	0.23
CUTIL	0.02	2.59	−3.46	2.59	0.29
REA	−14.30	187.83	−162.01	188.91	0.96
CFNAI	−17.02	197.96	−291.00	171.00	0.51
CLI	−0.04	0.60	−0.85	0.72	0.96
BCI	−0.01	0.64	−1.03	0.55	0.89
CCI	0.01	0.50	−0.52	0.52	0.83
AUS	−1.32	11.25	−19.55	9.90	0.02
CAN	−0.56	6.96	−13.01	8.94	−0.05
NZ	−0.98	12.02	−24.85	18.00	−0.02
SA	−6.31	14.63	−24.82	14.05	0.01
IND	−4.71	7.32	−19.89	7.00	0.04

Notes. This table reports the summary statistics of the excess returns on the S&P GSCI and the 28 predictors. We report the mean, standard deviation, minimum and maximum values, and the first-order autocorrelation. All values are in annualized percent. The sample period is from January 1976 to December 2019.

Table 2: Statistical performance of return forecasts

Predictor	MSFE	R^2_{OS} (%)	*MSFE-adjusted*
HA forecast	36.53		
Combination forecast			
Mean	36.16	1.01	2.40***
Median	36.50	0.08	0.74
Trimmed mean	36.24	0.81	2.31**
Weighted mean	36.15	1.04	2.41***
DMSFE, 0.9	36.12	1.13	2.41***
DMSFE, 0.7	36.09	1.21	2.43***
ABMA	36.17	0.98	2.38***
Subset (k=2)	35.87	1.81	2.42***
Subset (k=3)	35.65	2.41	2.43***
Subset (k=4)	35.48	2.89	2.44***
Subset (k=5)	35.35	3.22	2.44***
PC (ic=AIC)	35.24	3.54	2.43***
PC (ic=BIC)	35.15	3.77	2.38***
PC (ic=R2)	35.35	3.24	2.46***

Notes. This table reports out-of-sample forecasting results for the combination forecasts of log excess commodity returns. HA is the historical average benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the combination forecasts given in the first column relative to the HA forecast. Statistical significance for the R^2_{OS} 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 combination forecast MSFE against the one-sided (upper-tailed) alternative hypothesis that the HA forecast MSFE is greater than the combination forecast MSFE. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The out-of-sample forecast evaluation period is January 1990 to December 2019.

Table 3: Economic performance of return forecasts

Strategy	μ_p	σ_p	SR	SR_τ	$\triangle$	$\triangle_\tau$	TO	τ^{BE}	α
HA forecast	0.17	8.30	0.02	0.02					
Combination forecast									
Mean	1.88	10.12	0.19	0.16	1.21	1.01	7	82	4.12 (25.71)
Median	0.57	9.59	0.06	0.05	0.05	−0.04	4	19	3.90 (38.71)
Trimmed mean	1.63	10.11	0.16	0.14	0.96	0.78	6	69	4.08 (27.92)
Weighted mean	1.92	10.14	0.19	0.17	1.25	1.04	7	84	4.12 (25.32)
DMSFE ($\theta = 0.9$)	2.04	10.18	0.20	0.18	1.35	1.15	7	89	4.11 (23.91)
DMSFE ($\theta = 0.7$)	2.18	10.26	0.21	0.19	1.47	1.26	7	96	4.07 (23.46)
ABMA	1.84	10.10	0.18	0.16	1.18	0.98	7	80	4.11 (26.10)
Subset (k = 2)	3.21	11.76	0.27	0.24	2.00	1.62	12	145	4.50 (17.29)
Subset (k = 3)	4.31	13.49	0.32	0.28	2.45	1.92	16	197	4.81 (13.55)
Subset (k = 4)	5.19	14.90	0.35	0.30	2.74	2.07	20	240	5.05 (11.42)
Subset (k = 5)	6.01	16.33	0.37	0.32	2.90	2.10	24	279	5.24 (10.05)
PC (IC = AIC)	8.31	25.16	0.33	0.26	−0.30	−1.95	48	—	7.25 (8.41)
PC (IC = BIC)	8.94	22.03	0.41	0.36	2.56	1.63	28	418	7.35 (9.39)
PC (IC = R^2)	7.43	22.17	0.34	0.26	0.94	−0.62	45	346	7.05 (8.27)

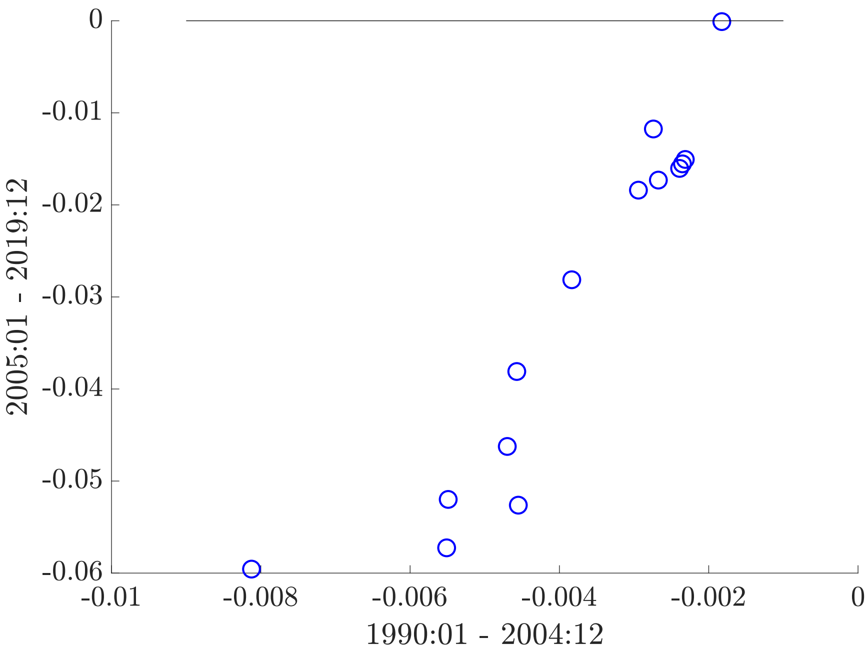
Notes. This table reports portfolio performance results for a mean-variance investor with relative risk aversion of three who monthly allocates her wealth between commodities and risk-free T-bills using either the HA benchmark forecast (static portfolio strategy) or the combination forecasts (dynamic portfolio strategy). For each portfolio strategy, we report the annualized percent mean realized return (μ_p), annualized percent realized volatility (σ_p), annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized percent utility gain (net of cost), $\triangle$ ($\triangle_\tau$), the turnover ratio (TO) ratio, the ratio of the average turnover of the dynamic strategy relative to that of the static strategy, the break-even transaction costs, τ^{BE}, and the alpha, α, estimates and Newey and West (1987) adjusted t-statistics (in parentheses) from regressions of the portfolio excess returns generated by each of the combination forecasts onto the excess returns on the S&P GSCI. 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 2019.

Table 4: Statistical performance of return forecasts in expansions and recessions

	Expansion			Recession		
Predictor	MSFE	R^2_{OS} (%)	*MSFE-adjusted*	MSFE	R^2_{OS} (%)	*MSFE-adjusted*
HA forecast	28.20			109.25		
Combination forecast						
Mean	28.16	0.13	0.66	105.96	3.01	2.51***
Median	28.21	−0.05	−0.19	108.85	0.36	1.58**
Trimmed mean	28.19	0.05	0.39	106.49	2.52	2.60***
Weighted mean	28.16	0.14	0.69	105.87	3.09	2.51***
DMSFE ($\theta = 0.9$)	28.16	0.13	0.67	105.55	3.39	2.50***
DMSFE ($\theta = 0.7$)	28.16	0.13	0.67	105.29	3.62	2.51***
ABMA	28.17	0.12	0.64	106.05	2.93	2.50***
Subset (k = 2)	28.16	0.15	0.71	103.18	5.55	2.51***
Subset (k = 3)	28.18	0.08	0.73	100.85	7.68	2.50***
Subset (k = 4)	28.21	−0.05	0.74	98.86	9.51	2.50***
Subset (k = 5)	28.27	−0.23	0.74	97.24	10.99	2.50***
PC (IC = AIC)	28.85	−2.29	0.78	91.04	16.67	2.47***
PC (IC = BIC)	28.72	−1.83	0.62	91.33	16.40	2.47***
PC (IC = R^2)	28.76	−1.99	0.87	92.84	15.02	2.47***

Notes. This table reports out-of-sample forecasting results for the combination forecasts of log excess commodity returns using the NBER-dated recession indicator. HA is the historical average benchmark forecast. MSFE is the mean squared forecast error. The R^2_{OS} statistic measures the proportional reduction in MSFE for the combination forecasts given in the first column relative to the HA forecast. Statistical significance for the R^2_{OS} 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 combination forecast MSFE against the one-sided (upper-tailed) alternative hypothesis that the HA forecast MSFE is greater than the combination forecast MSFE. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The out-of-sample forecast evaluation period is January 1990 to December 2019.

Figure 1: Log relative mean squared forecast error of combination forecasts



Notes. This figure plots the log relative mean squared forecast error (MSFE) of the combination forecasts 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 2019.

Table 5: 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}	α
HA forecast	0.01	0.01					0.05	0.05					
Combination forecast													
Mean	0.07	0.04	0.13	−0.07	6	24	0.70	0.68	10.83	10.65	4	1 021	4.17 (27.97)
Median	0.03	0.01	−0.08	−0.18	4	—	0.19	0.18	1.16	1.11	2	223	3.82 (36.41)
Trimmed mean	0.06	0.03	0.04	−0.13	6	19	0.60	0.59	9.02	8.87	4	889	4.10 (29.94)
Weighted mean	0.07	0.05	0.14	−0.06	6	24	0.71	0.70	11.07	10.89	4	1 046	4.17 (27.65)
DMSFE ($\theta = 0.9$)	0.07	0.05	0.16	−0.04	6	25	0.75	0.74	11.95	11.75	4	1 131	4.17 (27.00)
DMSFE ($\theta = 0.7$)	0.08	0.05	0.17	−0.03	6	26	0.80	0.79	12.98	12.77	5	1 224	4.14 (27.38)
ABMA	0.07	0.04	0.12	−0.08	6	23	0.68	0.67	10.59	10.42	4	997	4.16 (28.28)
Subset (k = 2)	0.09	0.05	0.12	−0.25	11	36	1.04	1.02	18.93	18.57	7	1 911	4.68 (20.42)
Subset (k = 3)	0.10	0.05	0.01	−0.52	15	45	1.20	1.18	24.74	24.21	10	2 669	5.10 (16.68)
Subset (k = 4)	0.11	0.05	−0.10	−0.78	19	—	1.31	1.28	29.01	28.35	12	3 261	5.44 (14.44)
Subset (k = 5)	0.12	0.05	−0.20	−1.01	22	—	1.35	1.33	31.93	31.13	15	3 790	5.72 (12.93)
PC (IC = AIC)	0.06	−0.03	−3.40	−5.10	46	—	1.33	1.31	31.51	30.45	21	5 943	7.87 (10.76)
PC (IC = BIC)	0.13	0.07	−1.19	−2.15	26	—	1.48	1.46	39.80	39.04	17	5 763	8.56 (14.29)
PC (IC = R^2)	0.11	0.02	−2.23	−3.82	43	—	1.30	1.27	31.28	30.15	19	4 528	7.50 (9.81)

Notes. This table reports portfolio performance results for a mean-variance investor with relative risk aversion of three who monthly allocates her wealth between commodities and risk-free T-bills using either the HA benchmark forecast (static portfolio strategy) or the combination forecasts (dynamic portfolio strategy). For each portfolio strategy, we report the annualized realized Sharpe ratio (net of cost), SR (SR_τ), annualized percent utility gain (net of cost), $\triangle$ ($\triangle_\tau$), 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}. 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. We also report the alpha, α, estimates and Newey and West (1987) adjusted t-statistics (in parentheses) from regressions of the portfolio excess returns generated by each of the combination forecasts onto the excess returns on the S&P GSCI and a recession dummy proxied by the NBER-dated recession indicator. The out-of-sample evaluation period is January 1990 to December 2019.

Table 6: Predictive regressions for stock market return and volatility

	$h = 1$ month			$h = 3$ months			$h = 12$ months		
Predictor	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
Panel A: Stock market return									
Mean	1.23	0.89	0.38	1.85	0.59	0.27	−6.57	−1.22	0.74
Median	−1.60	−0.94	0.25	−4.75	−1.40	0.68	−16.73	−1.23	1.82
Trimmed mean	0.99	0.67	0.20	1.14	0.35	0.08	−8.47	−1.30	1.01
Weighted mean	1.26	0.93	0.41	1.95	0.63	0.31	−6.28	−1.20	0.70
DMSFE ($\theta = 0.9$)	1.31	1.00	0.49	2.09	0.69	0.39	−6.33	−1.27	0.79
DMSFE ($\theta = 0.7$)	1.36	1.02	0.54	2.15	0.70	0.42	−6.87	−1.43	0.96
ABMA	1.19	0.85	0.34	1.75	0.55	0.23	−6.86	−1.25	0.78
Subset (k = 2)	1.12	1.26	0.86	1.97	0.97	0.83	−2.47	−0.83	0.29
Subset (k = 3)	0.92	1.38	1.08	1.67	1.11	1.12	−1.46	−0.67	0.19
Subset (k = 4)	0.79	1.46	1.23	1.44	1.19	1.30	−1.07	−0.61	0.16
Subset (k = 5)	0.69	1.50	1.32	1.28	1.23	1.41	−0.90	−0.61	0.16
PC (IC = AIC)	0.42	1.73*	1.47	0.72	1.33	1.35	0.19	0.18	0.02
PC (IC = BIC)	0.57	1.94*	2.12	1.10	1.58	2.43	0.27	0.26	0.03
PC (IC = R^2)	0.35	1.48	1.00	0.61	1.16	0.98	0.06	0.06	0.00
Panel B: Stock market volatility									
Mean	−1.30	−4.78***	8.44	−3.33	−3.13***	7.99	−7.78	−1.93*	3.60
Median	−0.72	−1.94*	1.01	−1.75	−1.20	0.85	−2.99	−0.33	0.20
Trimmed mean	−1.32	−4.56***	7.18	−3.41	−2.98***	6.87	−7.97	−1.77*	3.10
Weighted mean	−1.28	−4.76***	8.48	−3.28	−3.12***	8.04	−7.63	−1.94*	3.59
DMSFE ($\theta = 0.9$)	−1.19	−4.47***	8.02	−3.05	−2.91***	7.62	−6.76	−1.85*	3.13
DMSFE ($\theta = 0.7$)	−1.21	−4.58***	8.48	−3.12	−3.04***	8.16	−6.88	−2.02**	3.32
ABMA	−1.32	−4.80***	8.39	−3.38	−3.14***	7.94	−7.92	−1.92*	3.60
Subset (k = 2)	−0.90	−5.23***	10.86	−2.29	−3.46***	10.34	−5.30	−2.31**	4.67
Subset (k = 3)	−0.68	−5.33***	11.72	−1.75	−3.55***	11.23	−4.00	−2.38**	5.03
Subset (k = 4)	−0.56	−5.40***	12.19	−1.43	−3.61***	11.78	−3.26	−2.42**	5.22
Subset (k = 5)	−0.48	−5.48***	12.59	−1.24	−3.68***	12.24	−2.81	−2.47**	5.37
PC (IC = AIC)	−0.20	−3.71***	6.79	−0.50	−2.38**	5.90	−1.13	−1.54	2.57
PC (IC = BIC)	−0.27	−4.21***	9.52	−0.68	−2.65***	8.51	−1.54	−1.60	3.74
PC (IC = R^2)	−0.20	−3.70***	6.43	−0.48	−2.38**	5.61	−1.12	−1.58	2.53

Notes. This table reports the estimation results of predictive regressions for excess stock market return and volatility in (17) and (18), respectively at horizons of 1, 3, and 12 months ahead. The predictors are one of the 14 combination forecast of commodity returns (state variables) at a time. To the immediate right of slope estimates, β, are Newey and West (1987) heteroskedasticity and autocorrelation consistent t-statistics computed with $h - 1$ lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. R^2 denotes the coefficient of determination. The estimation period is January 1990 to December 2019.

Table 7: Predictive regressions for economic activity

	$h = 1$ month			$h = 3$ months			$h = 12$ months		
Predictor	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
Panel A: ADSI									
Mean	1.04	3.80***	12.37	3.02	2.70***	13.21	6.41	2.82***	5.03
Median	0.38	1.43	0.65	0.84	0.74	0.39	−0.36	−0.06	0.01
Trimmed mean	1.08	3.59***	10.98	3.08	2.52**	11.30	6.14	2.29**	3.78
Weighted mean	1.03	3.81***	12.55	2.99	2.71***	13.41	6.36	2.88***	5.13
DMSFE ($\theta = 0.9$)	1.01	3.81***	13.21	2.91	2.71***	13.99	5.98	2.99***	5.01
DMSFE ($\theta = 0.7$)	1.03	3.86***	13.98	2.95	2.74***	14.70	5.82	2.90***	4.87
ABMA	1.05	3.79***	12.19	3.05	2.69***	13.01	6.46	2.75***	4.92
Subset (k = 2)	0.74	4.38***	16.94	2.17	3.21***	18.60	4.89	3.88***	8.14
Subset (k = 3)	0.56	4.55***	18.46	1.66	3.35***	20.34	3.76	4.01***	9.10
Subset (k = 4)	0.46	4.64***	19.16	1.35	3.41***	21.10	3.06	4.03***	9.43
Subset (k = 5)	0.40	4.70***	19.58	1.16	3.45***	21.47	2.60	4.02***	9.46
PC (IC = AIC)	0.18	3.68***	12.70	0.58	2.87***	16.11	1.50	3.42***	9.42
PC (IC = BIC)	0.27	4.76***	21.84	0.84	3.67***	26.35	2.00	3.59***	13.06
PC (IC = R^2)	0.19	3.79***	13.52	0.58	2.96***	16.24	1.52	3.66***	9.54
Panel B: CFNAI									
Mean	0.71	3.23***	9.43	2.16	2.50**	13.42	4.69	2.45**	5.25
Median	0.22	1.14	0.37	0.52	0.64	0.30	−0.99	−0.24	0.09
trimmed mean	0.74	3.07***	8.56	2.21	2.35**	11.57	4.45	2.03**	3.88
Weighted mean	0.70	3.23***	9.55	2.13	2.51**	13.59	4.65	2.50**	5.35
DMSFE ($\theta = 0.9$)	0.68	3.20***	10.06	2.08	2.49**	14.21	4.39	2.58***	5.28
DMSFE ($\theta = 0.7$)	0.70	3.26***	10.87	2.13	2.56**	15.27	4.36	2.58***	5.35
ABMA	0.72	3.23***	9.31	2.18	2.50**	13.24	4.73	2.41**	5.15
Subset (k = 2)	0.50	3.64***	12.94	1.55	2.95***	18.95	3.64	3.37***	8.85
Subset (k = 3)	0.38	3.77***	14.18	1.19	3.09***	20.81	2.83	3.53***	10.07
Subset (k = 4)	0.32	3.84***	14.81	0.97	3.15***	21.68	2.32	3.59***	10.57
Subset (k = 5)	0.27	3.91***	15.26	0.84	3.20***	22.21	1.98	3.63***	10.76
PC (IC = AIC)	0.11	2.81***	7.94	0.39	2.42**	14.52	1.04	2.73***	8.74
PC (IC = BIC)	0.18	3.81***	15.98	0.60	3.28***	26.67	1.52	3.10***	14.67
PC (IC = R^2)	0.12	2.93***	8.76	0.39	2.51**	14.86	1.05	2.92***	8.84
Panel C: IP									
Mean	0.82	3.41***	7.86	2.70	3.75***	18.76	6.23	4.44***	10.46
Median	0.63	2.48**	1.79	1.72	2.03**	2.95	2.92	0.71	0.87
Trimmed mean	0.88	3.46***	7.56	2.84	3.50***	17.03	6.27	3.72***	8.67
Weighted mean	0.81	3.40***	7.90	2.67	3.78***	19.01	6.17	4.55***	10.64
DMSFE ($\theta = 0.9$)	0.78	3.31***	8.15	2.60	3.82***	19.76	5.81	4.73***	10.46
DMSFE ($\theta = 0.7$)	0.78	3.19***	8.34	2.59	3.78***	20.10	5.60	4.56***	9.96
ABMA	0.83	3.42***	7.81	2.73	3.72***	18.50	6.29	4.33***	10.28
Subset (k = 2)	0.55	3.42***	9.59	1.82	4.26***	23.40	4.35	5.68***	14.26
Subset (k = 3)	0.41	3.43***	10.19	1.37	4.39***	24.81	3.28	5.68***	15.27
Subset (k = 4)	0.34	3.44***	10.54	1.12	4.45***	25.44	2.65	5.62***	15.57
Subset (k = 5)	0.29	3.46***	10.79	0.95	4.49***	25.82	2.24	5.59***	15.54
PC (IC = AIC)	0.12	2.47***	5.90	0.44	3.54***	16.58	1.25	5.23***	14.25
PC (IC = BIC)	0.17	2.60***	8.31	0.61	4.33***	24.51	1.58	5.02***	17.88
PC (IC = R^2)	0.12	2.72***	6.01	0.45	3.75***	17.74	1.26	5.55***	14.43

Table 7 *continued*

Predictor	$h = 1$ month			$h = 3$ months			$h = 12$ months		
	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)	β	t-stats	R^2 (%)
Panel D: SRP									
Mean	−0.17	−1.63	3.00	−0.60	−1.31	4.13	−1.02	−0.85	0.93
Median	0.20	2.32**	1.62	0.62	1.48	1.71	3.18	1.59	3.42
Trimmed mean	−0.15	−1.31	1.88	−0.53	−1.03	2.65	−0.63	−0.47	0.30
Weighted mean	−0.17	−1.65*	3.12	−0.60	−1.33	4.30	−1.04	−0.89	1.01
DMSFE ($\theta = 0.9$)	−0.17	−1.67*	3.47	−0.60	−1.36	4.82	−1.05	−0.95	1.13
DMSFE ($\theta = 0.7$)	−0.18	−1.71*	3.87	−0.63	−1.40	5.33	−1.05	−0.94	1.17
ABMA	−0.17	−1.61	2.88	−0.59	−1.28	3.96	−0.99	−0.81	0.85
Subset (k = 2)	−0.15	−2.27**	6.30	−0.52	−1.89*	8.43	−1.16	−1.69*	3.34
Subset (k = 3)	−0.12	−2.49**	7.69	−0.41	−2.10**	10.18	−0.96	−1.95*	4.38
Subset (k = 4)	−0.10	−2.59***	8.41	−0.35	−2.20**	11.04	−0.81	−2.06**	4.82
Subset (k = 5)	−0.09	−2.67***	8.86	−0.30	−2.26**	11.53	−0.70	−2.10**	4.97
PC (IC = AIC)	−0.04	−2.12**	5.64	−0.15	−1.91*	8.32	−0.45	−2.09**	6.18
PC (IC = BIC)	−0.08	−3.30***	15.25	−0.26	−3.03***	20.09	−0.70	−2.82***	11.87
PC (IC = R^2)	−0.04	−2.06**	5.36	−0.14	−1.88*	7.89	−0.43	−2.07**	5.68
Panel E: TCU									
Mean	0.72	3.00***	6.09	2.42	3.28***	15.09	4.69	2.80***	6.30
Median	0.33	1.31	0.50	0.80	0.98	0.64	−1.70	−0.47	0.31
Trimmed mean	0.75	2.90***	5.38	2.42	2.89***	12.49	4.06	2.18**	3.87
Weighted mean	0.71	3.01***	6.21	2.41	3.34***	15.53	4.74	2.89***	6.66
DMSFE ($\theta = 0.9$)	0.71	3.01***	6.70	2.39	3.49***	16.83	4.65	3.00***	7.11
DMSFE ($\theta = 0.7$)	0.70	2.88***	6.72	2.36	3.42***	16.78	4.32	2.99***	6.30
ABMA	0.73	2.98***	5.96	2.43	3.23***	14.66	4.64	2.71***	5.94
Subset (k = 2)	0.51	3.22***	8.33	1.73	4.07***	21.00	3.77	4.47***	11.33
Subset (k = 3)	0.39	3.28***	9.15	1.32	4.33***	22.97	2.94	4.93***	13.03
Subset (k = 4)	0.32	3.32***	9.60	1.08	4.44***	23.86	2.41	5.07***	13.67
Subset (k = 5)	0.28	3.36***	9.89	0.92	4.50***	24.34	2.05	5.12***	13.78
PC (IC = AIC)	0.12	2.50***	5.97	0.45	3.75***	17.07	1.23	4.79***	14.85
PC (IC = BIC)	0.19	2.87***	10.35	0.67	5.37***	30.09	1.81	5.67***	24.88
PC (IC = R^2)	0.12	2.64***	5.57	0.44	3.76***	17.04	1.19	4.55***	13.67
Panel F: PAYEMS									
Mean	0.12	1.62	2.81	0.37	1.10	3.41	1.12	1.01	2.17
Median	−0.03	−0.44	0.05	−0.12	−0.48	0.14	−1.16	−0.84	0.89
Trimmed mean	0.13	1.62	2.62	0.39	1.07	3.06	1.03	0.86	1.53
weighted mean	0.12	1.60	2.79	0.37	1.08	3.41	1.10	1.02	2.20
DMSFE ($\theta = 0.9$)	0.12	1.53	2.84	0.36	1.04	3.53	1.05	1.02	2.20
DMSFE ($\theta = 0.7$)	0.13	1.67*	3.52	0.39	1.14	4.34	1.10	1.08	2.50
ABMA	0.13	1.65*	2.83	0.38	1.11	3.41	1.13	1.01	2.13
Subset (k = 2)	0.10	1.94*	4.62	0.30	1.34	5.81	0.97	1.46	4.58
Subset (k = 3)	0.08	2.05**	5.37	0.23	1.43	6.79	0.77	1.59	5.51
Subset (k = 4)	0.06	2.12**	5.81	0.19	1.48	7.33	0.64	1.65*	5.93
Subset (k = 5)	0.06	2.18**	6.17	0.17	1.51	7.73	0.55	1.68*	6.14
PC (IC = AIC)	0.02	1.45	2.70	0.07	1.12	4.14	0.28	1.44	4.80
PC (IC = BIC)	0.04	2.04**	6.09	0.12	1.54	8.94	0.46	1.87*	9.89
PC (IC = R^2)	0.02	1.50	2.91	0.07	1.15	4.29	0.28	1.49	4.79

Notes. This table reports the estimation results of predictive regressions for economic activity in (19) at horizons of 1, 3, and 12 months ahead. The economic activity variables are the Aruoba et al. (2009) business condition index (ADSI), the Chicago Fed National Activity Index (CFNAI), log growth in U.S. industrial production index (IP), the smoothed recession probability (SRP) of Chauvet (1998), change in total capacity utilization (TCU), or log growth in total nonfarm payroll employment (PAYEMS). To the immediate right of slope estimates, β, are Newey and West (1987) heteroskedasticity and autocorrelation consistent t-statistics computed with $h - 1$ lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. R^2 denotes the coefficient of determination. The estimation period is January 1990 to December 2019.

Table 8: Predictive regressions for the business cycle

Business cycle variable	β	t-stats	R^2 (%)
S&P 500 return	−0.0238	−0.50	0.10
Dividend yield	0.1026	0.30	0.02
Term spread	0.0012	0.10	0.00
Default spread	−0.0139	−2.08**	2.99
Unemployment rate	−0.0067	−0.56	0.05
SRP	−0.5691	−1.57	1.84
ADSI	2.5514	2.30**	3.84
IP	0.0071	0.98	0.35
TCU	0.0080	1.11	0.44
PAYEMS	0.0052	2.12**	2.36

Notes. This table reports the estimation results for the predictive regression $y_{t+1} = \alpha + \beta z_t + \varepsilon_{t+1}$ at 1-month horizon, where y_{t+1} is a business cycle variable and z_t is the excess return on the S&P GSCI. y is either the return on the S&P 500 index (S&P return), dividend yield on the S&P 500 (Dividend yield), Term spread, Default spread, U.S. Unemployment rate, the Smoothed recession probability of Chauvet (1998) (SRP), the Aruoba et al. (2009) business condition index (ADSI), log growth in U.S. industrial production index (IP), change in total capacity utilization (TCU), or log growth in total nonfarm payroll employment (PAYEM). To the immediate right of slope estimates, β, are heteroskedasticity consistent t-statistics. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. R^2 denotes the coefficient of determination. The estimation period is January 1990 to December 2019.

Table 9: Factor risk premia estimates from ICAPM

Model	λ_0	t_0	λ_M	t_M	λ_z	t_z	R^2_{OLS}
Panel A: Cross-sectional regressions without intercept							
Mkt (CAPM)			0.80	3.16***			−57.83
Mkt + S&P GSCI			0.78	3.04***	−0.57	−0.42	−65.56
Mkt + Mean			0.62	2.53**	0.04	2.14**	29.07
Mkt + Median			0.63	2.53**	0.02	1.96**	23.42
Mkt + Trimmed mean			0.62	2.54**	0.04	2.11**	27.95
Mkt + Weighted mean			0.62	2.53**	0.04	2.14**	28.94
Mkt + DMSFE ($\theta = 0.9$)			0.62	2.56**	0.04	2.14**	28.32
Mkt + DMSFE ($\theta = 0.7$)			0.63	2.59***	0.04	2.15**	27.90
Mkt + ABMA			0.62	2.46**	0.04	2.16**	29.20
Mkt + Subset (k = 2)			0.61	2.44**	0.08	2.17**	29.85
Mkt + Subset (k = 3)			0.61	2.44**	0.11	2.15**	29.47
Mkt + Subset (k = 4)			0.61	2.43**	0.14	2.13**	29.00
Mkt + Subset (k = 5)			0.61	2.43**	0.17	2.10**	28.32
Mkt + PC (IC = AIC)			0.60	2.35**	0.34	2.33**	31.64
Mkt + PC (IC = BIC)			0.65	2.54**	0.24	2.60***	32.12
Mkt + PC (IC = R^2)			0.60	2.37**	0.34	2.24**	29.14
Panel B: Cross-sectional regressions with intercept							
Mkt (CAPM)	0.48	2.17**	0.29	0.90			29.89
Mkt + S&P GSCI	0.47	2.01**	0.28	0.88	−0.28	−0.21	29.35
Mkt + Mean	0.28	1.26	0.41	1.37	0.02	1.05	43.22
Mkt + Median	0.30	1.38	0.39	1.29	0.01	0.94	43.98
Mkt + Trimmed mean	0.28	1.29	0.40	1.36	0.02	1.03	43.15
Mkt + Weighted mean	0.28	1.27	0.41	1.37	0.02	1.05	43.12
Mkt + DMSFE ($\theta = 0.9$)	0.28	1.27	0.41	1.38	0.02	1.03	42.80
Mkt + DMSFE ($\theta = 0.7$)	0.28	1.28	0.41	1.39	0.02	1.02	42.50
Mkt + ABMA	0.28	1.26	0.41	1.37	0.02	1.06	43.32
Mkt + Subset (k = 2)	0.27	1.26	0.41	1.37	0.04	1.08	43.30
Mkt + Subset (k = 3)	0.27	1.28	0.40	1.36	0.06	1.08	43.18
Mkt + Subset (k = 4)	0.28	1.30	0.40	1.35	0.08	1.07	43.05
Mkt + Subset (k = 5)	0.28	1.31	0.40	1.34	0.09	1.05	42.89
Mkt + PC (IC = AIC)	0.26	1.21	0.41	1.32	0.20	1.17	42.59
Mkt + PC (IC = BIC)	0.25	1.18	0.45	1.46	0.14	1.22	41.11
Mkt + PC (IC = R^2)	0.27	1.27	0.40	1.31	0.19	1.14	41.28

Notes. This table reports estimation and evaluation results for the two-factor ICAPM: $\bar{R}_i = \lambda_0 + \lambda_M \hat{\beta}_{i,M} + \lambda_z \hat{\beta}_{i,z} + \alpha_i, \ \forall i$. The estimation procedure is the Fama and MacBeth (1973) two-step cross-sectional regression approach. The testing assets (R_is) are the excess returns on the 24 individual commodity futures and the 25 equity portfolios formed on size and book-to-market. λ_0, λ_M and λ_z denote the zero beta rate (intercept), and risk premium estimates (in percent per month) for the market and state variable factors, respectively. The market factor is the excess return on the CRSP value-weighted stock market portfolio. The state variable factors represent the innovations in each of the 14 combination forecasts of commodity returns (as measured via the VAR(1) residuals). Panels A and B report results for the pricing equation without and with intercept, respectively. Next to the risk premia estimates are displayed t-statistics based on GMM standard errors corrected for errors-in-variable bias, heteroskedasticity and autocorrelation in the residuals. The number of lags to include in the computation of the standard errors is based on the Bartlett kernel with Newey and West (1994) optimal bandwidth selection. ** and *** indicate significance at the 5% and 1% levels, respectively. 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 estimation period is January 1990 to December 2019.