Research Center
Finance & Information Management

Project Group
Business & Information
Systems Engineering

Discussion Paper

The Predictability of Aggregate Returns on Commodity Futures

by

Fabian Lutzenberger

in: Review of Financial Economics 23 (2014) 3, p. 120-130

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: F.-v.-Schiller-Str. 2a, 95444 Bayreuth
Phone: +49 921 55-4710 (Fax: -844710)

WI-440

www.fim-rc.de

This version: 16 January, 2014

The predictability of aggregate returns on commodity futures

Fabian Lutzenberger
University of Augsburg
FIM Research Center Finance & Information Management
Universitätsstraße 12, 86159 Augsburg, Germany
fabian.lutzenberger@wiwi.uni-augsburg.de
Phone: +49 821 598-4884 (Fax: -4899)
+49 821 598-4801 (Secretariat)

Abstract:

This paper provides evidence that aggregate returns on commodity futures (without the returns on collateral) are predictable, both in-sample and out-of-sample, by various lagged variables from the stock market, bond market, macroeconomics, and the commodity market. Out of the 32 candidate predictors we consider, we find that investor sentiment is the best in-sample predictor of short-horizon returns, whereas the level and slope of the yield curve has much in-sample predictive power for long-horizon returns. We find that it is possible to forecast aggregate returns on commodity futures out-of-sample through several combination forecasts (the out-of-sample return forecasting R^2 is up to 1.65% at the monthly frequency).

JEL classification: G12, G13, G17

Keywords: asset pricing, commodities, predictability of returns, predictive regressions, forecasting

1. Introduction

Are asset returns predictable? The 2013 Nobel Prize recipients Eugene Fama, Lars Peter Hansen and Robert Shiller find that this question "is as central as it is old" (The Royal Swedish Academy of Sciences, 2013). While several studies examine whether excess returns on asset classes such as stocks, treasuries, bonds, foreign exchange, sovereign debt, and houses are predictable (Cochrane, 2011 and the articles cited therein), there is less (up-to-date) evidence for the predictability of commodity futures returns, despite the fact that "commodity futures have [by now] moved into the investment mainstream" (Basu and Miffre, 2013).

This paper attempts to fill this research gap to some extent by studying the predictability of aggregate returns on commodity futures, that is, we test whether the null of unpredictable commodity futures returns can be rejected and seek to identify variables that show predictive power. For this purpose, we do not empirically test one specific theory of commodity futures returns and its implications, as the theory of storage of Kaldor (1939) and others. Instead, we follow an empirical asset pricing approach. That is, by working backwards, we come from the empirical side and test a large set of potential predictors. Most of these candidate predictors are standard choices in studies of return predictability of other asset classes, especially that of stocks and bonds. In addition, we propose some new factors, of which most are commodity-specific. The empirical facts we identify should be subject to further theoretical studies that seek to propose theoretical models that capture these given empirical patterns.[1]

To be somewhat more specific, we conduct predictive regressions that use the future return on an equal-weighted portfolio of 27 commodity futures (without the return on collateral) as the response variable. The right-hand sides of these regressions comprise the current values of subsets of 32 potential predictors from the stock market, bond market, macroeconomics, and the commodity market. The main sample period is from January 1972 to June 2010, with a monthly sample frequency. Predictive regressions are the most common approach to forecast aggregate returns (Kelly and Pruitt, 2013). If returns are unpredictable, regression coefficients beyond a constant should be insignificant in such models, and these models should not provide forecasts of future returns that are more accurate than the historical average of past returns. We evaluate both the in-sample (IS) and the out-of-sample (OOS) predictability. For the IS analysis, we employ single long-horizon predictive regressions with horizons of 1, 3, 12, 24, 36, 48, and 60 months ahead (Maio and Santa-Clara 2012; among others) as well as a

[1] Fama and French (2013) describe the research philosophy of empirical asset pricing.

procedure that selects the "best" multiple-variable regression out of the variety of candidate predictors we consider, as is proposed by Bossaerts and Hillion (1999) and Zakamulin (2013). The OOS evaluation comprises forecasts from single predictive regressions, from the model selection procedure that follows Bossaerts and Hillion (1999) and Zakamulin (2013), and from several combination forecasts that are proposed by Rapach et al. (2010).

Our main results include the following. First, aggregate returns on commodity futures appear to be predictable IS. Out of the set of candidate predictors we consider, a high level of Baker and Wurgler's (2006; 2007) stock market sentiment index seems to be the "best" single IS predictor for low subsequent short-horizon aggregate commodity futures returns (the R^2 for the one-month horizon is 1.96%), whereas much forecastability of long-horizon returns appears to come from the current level and slope of the yield curve (the R^2 for the 48-month horizon and the three-month U.S. Treasury bill rate as the independent variable is 33.96%, for instance). A model selection procedure even describes 7.85% (64.61%) of the variation in returns one month (48 months) ahead. Second, aggregate returns on commodity futures appear to be predictable OOS as well. In particular, we obtain a significantly positive OOS R^2 for the forecast combining methods proposed by Rapach et al. (2010). For instance, the OOS R^2 of the mean combination forecast is 1.65% over the OOS evaluation period January 1980 to June 2010. Hence, there are models that provide forecasts that significantly outperform the historical average of past returns.

The remainder of this paper is organized as follows. We describe our variables and data in section 2. Section 3 presents the econometric methodology, and in section 4, we present our empirical results, which we discuss in section 5. Section 6 presents our paper's conclusions.

2. Variables and Data

2.1 Response variable

We study the predictability of one variable – the return on a portfolio that consists of several commodity futures. Our sample period is from January 1972 to June 2010, with a monthly sample frequency. All prices and returns are denominated in U.S. dollars.

- *Return on commodity futures, CM:* We employ monthly returns on an equal-weighted portfolio of 27 commodity futures that is constructed by Asness et al. (2013). The portfolio covers aluminum, copper, nickel, zinc, lead, tin, brent crude oil, gas oil, live cattle, feeder cattle, lean hogs, corn, soybeans, soy meal, soy oil, wheat, WTI crude, RBOB gasoline, heating oil, natural gas, gold, silver, cotton, coffee, cocoa, sugar, and platinum. The futures returns are calculated by computing the daily excess return of the most liquid futures contract every day (typically, the nearest- or next nearest-to-delivery contract). The daily returns are then compounded to a total return index, and the monthly returns are computed from this index. The returns do not include the return on collateral associated with the futures contract. Thus, these returns are comparable to returns in excess of the risk-free interest rate. This is important to note because we are attempting to forecast the reward for risk, not the interest rate. The data are obtained from Tobias J. Moskowitz's website.

2.2 Potential predictor variables

The core of this paper is a test of the null hypothesis that returns on commodity futures are unpredictable against the alternative hypothesis that the expected returns depend on factors such as price levels and past price movements, economic conditions, and investor sentiment, and consequently vary trough time. Accordingly, our approach is not to test a specific theory of commodity futures returns that represents this alternative hypothesis and, at the same time, predetermines the set of potential predictor variables, but to choose the candidate predictors ourselves. While this approach examines variables that have not yet been suggested by any theory, the drawback of this approach is the selection of potential predictors that is, to some extent, arbitrary.

We employ a total of 32 variables that reflect price levels and past price movements, economic conditions and investor sentiment, which we present in **Table 1**.

– Please insert Table 1 (p. 24) about here. –

The variables are classified into four groups: stock (Panel A), bond (Panel B), macroeconomic (Panel C), and commodity characteristics (Panel D). A large number of these variables are relatively common choices in the literature that studies the predictability of stock and bond returns. We are interested in whether these variables also have forecasting power over commodity futures returns, and therefore, we include them in the set of candidate predictors. Because our space is limited, however, we refer to the studies mentioned in **Table 1** for a description of the motivation behind and construction of these variables. Some variables are less standard choices in the predictability-of-returns literature or have not yet been considered, and therefore, they are presented in more detail herein.

The first three potential predictive variables, which we present at length, are stock market characteristics:

- *Equity premium (five-year cumulative sum), CRMRF:* Cumulative sum of the equity premium, which is the total return on the stock market in excess of the risk-free rate, over the last 60 months. We obtain monthly data for the U.S. equity premium as it is employed by Fama and French (1993) from Kenneth R. French's website, which includes all NYSE, AMEX, and NASDAQ firms. The intention behind the cumulative sum is to obtain a slow-moving predictive variable that corresponds to the equity premium. We choose the last 60 months rather than the total cumulative sum (or index level) because the variable constructed in this way is stationary, whereas the total cumulative sum is close to being non-stationary (the autocorrelation coefficient is around one). This approach is proposed by Maio and Santa-Clara (2012), who construct a predictive (state) variable that is associated with the momentum factor of Carhart (1997) as well as the following liquidity factor. We also apply this approach to several other return or growth rate series for which we want to obtain associated, slow-moving predictive variables.
- *Liquidity factor (five-year cumulative sum), CL:* Cumulative sum of Pástor and Stambaugh's (2003) non-traded liquidity factor over the last 60 months, which represents innovations in aggregate stock market liquidity (Pástor and Stambaugh 2003, equation (8)), from Lubos Pástor's website. This variable is employed by Maio and Santa-Clara (2012).
- *Investor sentiment, SENT:* The stock market sentiment index of Baker and Wurgler (2006; 2007) is based on the first principal component of six U.S. sentiment proxies – the NYSE turnover, dividend premium, closed-end fund discount, number of and first-day returns on IPOs and the equity share in new issues. The monthly data are obtained from Jeffrey Wurgler's website, and they are described in Baker and Wurgler (2007). We choose the series where each of the proxies has first been orthogonalized with respect to a set of macroeconomic conditions.

We then highlight a potential predictor that is a bond market characteristic:

- *Cochrane-Piazzesi factor, CP:* The factor proposed by Cochrane and Piazzesi (2005) is the fitted value from a regression of an average of excess bond returns on forward rates and is related to bond risk premia. We obtain the necessary data, which cover the period January 1972 to December 2003, to construct the *CP* from John H. Cochrane's website.

Our next set of independent variables, which we explicitly outline, can be categorized as primarily macroeconomic factors:

- *Industrial production growth (five-year), IP*: Five-year log growth of U.S. industrial production for which the data are obtained from the Federal Reserve Bank of St. Louis, as are the data for the following three variables.

- *M2 money stock growth (three-year), M2*: Three-year log growth of the U.S. M2 money stock.
- *GDP growth (three-year), GDP*: Three-year log growth of U.S. GDP. We linearly interpolate the quarterly GDP data to obtain a monthly series.
- *Return on U.S. dollar (five-year), USD*: Five-year log return on a trade weighted U.S. dollar index against major currencies. The series covers January 1978 to June 2010.
- *Return on Value Everywhere (five-year), CVAL*: Cumulative sum of the log excess return on Asness et al.'s (2013) value "everywhere" factor over the last 60 months. The factor comprises eight asset classes (U.S. equities, U.K. equities, continental Europe equities, Japanese equities, global equity indices, currencies, fixed income, and commodities). The variable seeks to represent the cross-sectional value return premium across these eight asset classes. We obtain the factor data from Tobias J. Moskowitz's website. The series constructed this way covers December 1976 to June 2010.
- *Return on Momentum Everywhere (five-year), CMOM*: Cumulative sum of the log excess return on Asness et al.'s (2013) momentum "everywhere" factor over the last 60 months. The factor comprises the same eight asset classes as the value "everywhere" factor. It represents the cross-sectional momentum return premium across these eight asset classes. The data are from Tobias J. Moskowitz's website. The resulting series covers December 1976 to June 2010.

Finally, we construct various factors from commodity market data that potentially show predictive power over commodity futures returns:

- *Commodity variance, CVAR*: In a manner analogous to *SVAR*, we compute the volatility of the aggregate commodity spot market as the sum of squared daily returns on the CRB BLS spot index. The price index data are obtained from Datastream.
- *Return on CRB BLS spot index (five-year), CCM_spot*: Cumulative sum of the monthly log return on the CRB BLS spot index over the last 60 months. Our intuition behind this variable is to capture potential "time series value" in commodities (see Asness et al., 2013 for the cross-sectional value effect in commodities).
- *Return on commodity futures (five-year), CCM*: Cumulative sum of *CM* over the last 60 months. This variable also seeks to capture potential "time series value". The series computed this way covers December 1976 to June 2010.
- *Return on CRB BLS spot index (12-month), C12CM_spot*: Cumulative sum of the monthly log return on the CRB BLS spot index over the last 12 months, where the most recent month's return is skipped. This measure, MOM2-12, is the common measure to capture

"momentum" (Jegadeesh and Titman, 1993; Asness et al., 2013). Significant "time series momentum" in commodity futures is found by Moskowitz et al. (2012).

- *Return on commodity futures (12-month), C12CM*: Cumulative sum of *CM* over the last 12 months, where the most recent month's return is skipped. This variable is also employed to capture "time series momentum". The resulting series covers December 1972 to June 2010.
- *Return on Commodities Value (five-year), CVAL_CM*: Cumulative sum of the log excess return on Asness et al.'s (2013) commodities value factor over the last 60 months. The variable represents the cross-sectional value return premium in commodities. The data are from Tobias J. Moskowitz's website, as they are for the following variable. The series constructed this way covers December 1976 to June 2010.
- *Return on Commodities Momentum (five-year), CMOM_CM*: Cumulative sum of the log excess return on Asness et al.'s (2013) commodities momentum factor over the last 60 months. The factor represents the cross-sectional momentum return premium in commodities. The series covers December 1976 to June 2010.

2.3 Summary statistics

– Please insert Table 2 (p. 26) about here. –

Table 2 shows summary statistics for the response and predictor variables. Observe that the first-order autoregressive coefficients of most predictors are above 0.9, indicating that most predictors are highly persistent. Some variables are correlated with others to some extent (not displayed in the table). Correlation coefficients above 0.85 are shown by *D/P* and *E/P* (0.91), *D/P* and *B/M* (0.91), *E/P* and *B/M* (0.91), *B/M* and *GDP* (0.87), and *TBL* and *LTY* (0.86).

3. Econometric Methodology

3.1 In-sample predictive regressions

We begin with IS single long-horizon predictive regressions, which are the common approach to assess the ability of a single potential predictor variable to forecast future returns (Cochrane 2011; Maio and Santa-Clara 2012; among many others):

$$r_{t,t+q} = a_q + b_q x_t + \varepsilon_{t,t+q}, \tag{1}$$

where $r_{t,t+q} \equiv r_{t+1} + \cdots + r_{t+q}$ represents the continuously compounded return over q periods, i.e., from $t+1$ to $t+q$, x_t is the value of the variable at time t whose predictive ability we want to assess, and $\varepsilon_{t,t+q}$ is a disturbance term (the forecasting error) with zero conditional mean, $E_t(\varepsilon_{t,t+q}) = 0$. The conditional expected return at time t can then be expressed as $E_t(r_{t,t+q}) = a_q + b_q x_t$. The forecasting power of x is assessed by regarding the degree of statistical significance of the slope coefficient, b_q, as well as by measuring the adj. R^2 of the regression. If the returns are unpredictable beyond a constant, i.e., i.i.d., b_q is statistically indistinguishable from zero. Following Maio and Santa-Clara (2012), we choose forecasting horizons of 1, 3, 12, 24, 36, 48, and 60 months ahead. The regressions are performed over the original sample period, January 1972 to June 2010, as well as over the two subsample periods January 1972 to December 1999 and January 2000 to June 2010, where q observations are lost in each respective q-horizon regression. By splitting the original sample into two subsamples, we seek to identify any structural changes over time, while the breakpoint is chosen to highlight the 2000s commodity boom. The regressions are conducted for each predictor proposed in section 2 whose data series covers the respective period. Following Maio and Santa-Clara (2012) and others, we compute both Newey and West (1987) and Hansen and Hodrick (1980) t-ratios with q lags to assess the statistical significance of the regression coefficients. The q lags are selected to correct for the serial correlation in the regression residuals that are induced by the overlapping observations.

In a next step, we extend this single predictive regression model to a multiple-variable predictive regression model. As our goal is not to test any existing theory that predetermines the right-hand side, it is unclear which of the variety of predictors we proposed should at once enter a multiple predictive regression. We rather seek to assess the marginal forecasting power of *each* candidate variable, conditional on *each* possible combination of all other variables. For this purpose, we employ a model selection procedure as used by Bossaerts and Hillion (1999), Zakamulin (2013), and others for each horizon q. This procedure seeks to select the "best" regression model out of 2^N competing specifications of the following form:

$$r_{t,t+q} = \begin{cases} a_q + \boldsymbol{b}_{\boldsymbol{q}}' \boldsymbol{x}_{\boldsymbol{t}} + \varepsilon_{t,t+q}, & if\ n > 0, \\ a_q + \varepsilon_{t,t+q}, & if\ n = 0, \end{cases} \tag{2}$$

where $0 \le n \le N$ and $\boldsymbol{x}_{\boldsymbol{t}}$ is a model-unique n-by-1 subvector of the vector of values at time t of all N candidate predictors. With regard to the large number of candidate predictors, we limit the set of competing regressions to those that comprise less than or equal to seven independent variables to reduce the risk of over-specification and to keep computation times in an acceptable range. Hence, we estimate each possible regression specification that includes not more than seven predictor variables, including the model with no predictors other than the constant, $n = 0$. The “best” model is then chosen according to a predefined model selection criterion. We choose the adj. R^2 for this purpose. If the returns are unpredictable beyond a constant, i.e., i.i.d., the procedure should select the specification $n = 0$. We perform the model selection procedure for the same horizons and sample periods as the single predictive regressions. However, to make the results across the three sample periods comparable, we only consider those predictors whose data series cover the original sample period December 1972 to June 2010. Accordingly, we regard $N = 23$ potential predictive variables.

3.2 Out-of-sample forecasting

The informative value of IS predictive regressions is not without controversy. Rather, some authors argue that the results might be spurious and might not hold OOS (e.g., Bossaerts and Hillion, 1999; Welch and Goyal, 2008; Zakamulin, 2013). For instance, Rapach et al. (2010) show that single predictive regressions with *E/P*, *D/E*, *SVAR*, *B/M*, *TBL*, *LTY*, *LTR*, *TMS*, *DFY*, *DFR*, and *INFL* have no OOS predictive power over the U.S. equity premium, measured by quarterly S&P 500 excess returns over the period 1965 to 2005. For this reason, we also assess the OOS predictability of aggregate returns on commodity futures and employ an OOS predictive regression model that is based mainly on the OOS approaches used in Welch and Goyal (2008), Rapach et al. (2010), and Zakamulin (2013).

We firstly conduct individual OOS forecasts from single predictive regressions, as described in Rapach et al. (2010). Therefore, we start with a single predictive regression model for each candidate predictor, as formulated in equation (1), but we refrain from incorporating multiple horizons:

$$r_{t+1} = a + bx_t + \varepsilon_{t+1}, \qquad (3)$$

where r_{t+1} is the return, and ε_{t+1} is the disturbance term at time $t + 1$. OOS forecasts are then generated with a recursive (expanding) estimation window. We then split the total sample of observations for r_t and x_t into an initial IS period that consists of the first m observations and an OOS period that includes the last s observations. The initial OOS forecast of the return at time $m + 1$, based on a single predictor, is then computed as:

$$\hat{r}_{m+1} = \hat{a}_m + \hat{b}_m x_m, \tag{4}$$

where $\hat{a}_m$ and $\hat{b}_m$ are the estimates of a and b from equation (3), which are computed using observations of r_t from $t = 2$ to $t = m$ and of x_t from $t = 1$ to $t = m - 1$. In the next step, the estimation window is expanded by one period such that we obtain an estimate of the return at $m + 2$, $\hat{r}_{m+2}$, via observations of r_t from $t = 2$ to $t = m + 1$ and of x_t from $t = 1$ to $t = m$. We continue this procedure through the end of the OOS period and obtain a series of s OOS return forecasts based on a single predictor. This approach is conducted for each candidate predictor variable that we proposed in section 2 and whose data series covers the original sample period January 1972 to June 2010.

We consider three different OOS periods: January 1980 to June 2010, January 1980 to December 1999, and January 2000 to June 2010. The IS period starts eight years prior to the beginning of the respective OOS period.[2]

In addition to these individual forecasts, we employ three combination forecasts that are proposed by Rapach et al. (2010): mean, median, and trimmed mean. The mean combining method computes the arithmetic average of all individual forecasts of r_{t+1} made at time t to obtain another forecast of r_{t+1}. In an analogous manner, the median combination forecast of r_{t+1} is the median of all individual forecasts of r_{t+1} made at t. Finally, the trimmed mean combination forecast computes the arithmetic average of all but the smallest and largest individual forecasts of r_{t+1} obtained at t.[3]

Finally, we use the OOS recursive forecasting procedure proposed by Zakamulin (2013). This procedure follows the individual forecasting method in equations (3) and (4), but instead of single predictive regressions, it conducts the model selection procedure described in section 3.1. Hence, the first m observations are used to find the optimal (multiple-variable) predictive model to make the forecast $\hat{r}_{m+1}$. Following that, the IS period is expanded by one month, and we repeat the procedure to find the best model using data from $t = 1$ to $t = m + 1$ and compute the forecast $\hat{r}_{m+2}$. This procedure is continued through the end of the OOS period. As a result, we obtain a series of s OOS return forecasts, where each forecast is based on the best (multiple-variable) model using only data prior to the month for which the forecast is made. To keep computation times manageable, we limit the set of potential predictors and,

---

[2] We also considered IS periods that start at the beginning of our original sample, i.e., in January 1972, what follows Rapach et al. (2010). However, the results were less convincing.

[3] We also implemented more complex combining methods proposed by Rapach et al. (2010), which require a holdout OOS period to estimate the combining weights. However, the performance of these forecasts was not better than the simple schemes described above, thus confirming Rapach et al.'s (2010) results.

similar to Zakamulin (2013), only consider nine variables. Although this induces some look-ahead-bias, we use predictors that have performed relatively well IS: *D/E*, *SENT*, *LTY*, *DFR*, *CAY*, *M2*, *CCM_spot*, *C12CM_spot*, and *CVAR*.

We employ two measures to evaluate the individual, combination, and model selection OOS forecasts: the OOS R^2 statistic, R^2_{OS}, proposed by Campbell and Thompson (2008) and used by Rapach et al. (2010), among others, and the Henriksson-Merton test statistic, which is based on the Henriksson and Merton test of directional accuracy (Henriksson and Merton, 1981; Pesaran and Timmermann, 1992) and employed by Zakamulin (2013).

The R^2_{OS} is based on a series of moving historical averages of returns, $\bar{r}_{t+1} = \frac{1}{t}\sum_{j=1}^{t} r_j$, which are used as the benchmark for the respective return forecasting method under evaluation. In particular, the statistic is computed as:

$$R^2_{OS} = 1 - \sum_{k=1}^{s}(r_{m+k} - \hat{r}_{m+k})^2 / \sum_{k=1}^{s}(r_{m+k} - \bar{r}_{m+k})^2. \tag{5}$$

The forecasting model under investigation, which generates the forecasts $\hat{r}_{t+1}$, outperforms the historical average forecast in terms of mean squared prediction errors if $R^2_{OS} > 0$.

In a second step, we test whether the R^2_{OS} is significantly greater than zero in two ways. First, we follow Rapach et al. (2010) and compute the *MSPE-adjusted* statistic proposed by Clark and West (2007). For this purpose, we first calculate:

$$f_{t+1} = (r_{t+1} - \bar{r}_{t+1})^2 - [(r_{t+1} - \hat{r}_{t+1})^2 - (\bar{r}_{t+1} - \hat{r}_{t+1})^2], \tag{6}$$

and then regress the series of f_{t+1} on a constant and compute its t-statistic. The p-value for rejecting the null hypothesis $R^2_{OS} \leq 0$ is then obtained using the standard normal distribution.

In addition to calculating the *MSPE-adjusted* statistic, we employ the non-parametric bootstrap method proposed by Zakamulin (2013). The null hypothesis, $R^2_{OS} \leq 0$, corresponds to the null that returns are unpredictable and therefore i.i.d. The bootstrap method is used to estimate the sampling distribution of R^2_{OS} under the conditions given by this null hypothesis. To be specific, after having computed the R^2_{OS} for each forecasting model under investigation using the original time series of returns and predictors, we bootstrap the original time series to obtain random resamples of the returns and predictive variables. Accordingly, we resample the entire vector of returns and predictors at each time t to maintain the historical intratemporal correlations between these variables. We then compute the R^2_{OS} of each forecasting model using the resampled time series of returns and predictive variables. This procedure is repeated numerous times, and we count how many times the R^2_{OS} is above the R^2_{OS} obtained using the original time-series to obtain empirical p-values for the null hypothesis, $R^2_{OS} \leq 0$, against the alternative $R^2_{OS} > 0$.

Finally, the Henriksson-Merton test statistic is computed as in Zakamulin (2013):

$$HM - S = Prob(\hat{r}_{t+1} > 0 | r_{t+1} > 0) + Prob(\hat{r}_{t+1} \leq 0 | r_{t+1} \leq 0), \tag{7}$$

where $Prob(\hat{r}_{t+1} > 0 | r_{t+1} > 0)$ is the conditional probability of obtaining a correct forecast, $\hat{r}_{t+1}$, of a positive return at time $t+1$ using the respective forecasting model under evaluation, given that the realized return at $t+1$, r_{t+1}, is positive. A forecasting model that is able to forecast the sign of the return generates a Henriksson-Merton test statistic that is greater than one, $HM - S > 1$. If the null that the return is unpredictable is true, the statistic should be unity, $HM - S = 1$. Following Zakamulin (2013), we test the null $HM - S \leq 1$ against the alternative hypothesis, $HM - S > 1$, by obtaining empirical p-values through the same non-parametric bootstrap method as the one described above.

4. Results

4.1 In-sample

Table 3 shows the results for the IS single predictive regressions. With three sample periods, 23 or more candidate predictors for each sample period and seven forecasting horizons, there is reason to limit the results shown to the original sample period (January 1972 to June 2010) as well as to variables that significantly predict future returns at least one horizon at the 5% level (as indicated by either Newey-West or Hansen-Hodrick t-ratios). To save even more space, the Hansen-Hodrick t-ratios are only shown for those variables that would have not been included in the table by only regarding the Newey-West values.

– Please insert Table 3 (p. 27) about here. –

Note that *SVAR*, *B/M*, *LTR*, *DFY*, *INFL*, *I/K*, *IP*, *GDP*, *CCM_spot*, and *C12CM_spot* are unable to forecast either short-term or long-term returns within the original sample period. Hence, the results for these variables are only available upon request. Nonetheless, the regression coefficients on the other variables we consider are significantly different from zero. On the one hand, *D/P*, *E/P*, *CRMRF*, *CL*, *SENT*, *TBL*, *LTY*, and *CAY* consistently forecast negative returns. On the other hand, *D/E*, *TMS*, *DFR*, and *M2* predict positive returns.

Additionally, the predictive sign on *CVAR* is negative for horizons of one and three months (but without significance), and becomes positive for longer horizons. As a consequence, aggregate returns on commodity futures seem to be predictable IS beyond a constant, and therefore are not i.i.d. Observe that the predictive power of most variables increases with the horizon according to the adj. R^2 values. This is a result of the high persistence of most predictors, i.e., their slow movement, as it is indicated by their autocorrelation coefficients above 0.9 (Cochrane, 2005, chapter 20).

For horizons of one to 12 months, *SENT* shows the highest forecasting power according to the t-statistics and R^2. Thus – out of the set of candidate predictors we consider – a high sentiment level of stock market investors seems to be the "best" single predictor for low subsequent short-horizon aggregate returns on commodity futures. Regarding horizons of 24 months and longer, we see that a high level of *TBL* – followed by *LTY* and *TMS* – does a good job of indicating low future long-horizon returns. Consequently, much forecastability of long-horizon aggregate returns on commodity futures appears to come from the current level and slope of the yield curve.

Regarding the first subsample, January 1972 to December 1999 (not reported), the regression coefficients on *D/E*, *SVAR*, *B/M*, *CSP*, *DFR*, *I/K*, *IP*, *GDP*, and *CCM_spot* are insignificant at the 5% level for each horizon. Hence, the significance of *D/E* and *DFR* vanishes when we only consider the first subsample. On the other hand, *LTR* now significantly forecasts positive returns at the 60-month horizon (according to the Hansen-Hodrick t-ratio), while *DFY* and *INFL* become significant predictors of negative returns at several horizons. Moreover, the results for *C12CM_spot* at the one- to 12-month horizons indicate time series momentum within this subsample: a high cumulative spot return over the last 12 months forecasts high future short-term returns, which coincides with the findings of Moskowitz et al. (2012). The predictors introduced in the first subsample are *CSP* and *CP*. While the slope of *CSP* is insignificant, *CP* significantly predicts negative returns at horizons of one and three months (at the 10% level) and positive returns at the 48-month horizon (according to the Hansen-Hodrick t-ratio). Moreover, *D/P*, *E/P*, *CRMRF*, *CL*, *SENT*, *TBL*, *LTY*, and *CAY* still consistently forecast negative returns, while *TMS* and *M2* still predict positive returns.

Examining the second subsample, January 2000 to June 2010 (not reported), we see that, in contrast to the first subsample, all candidate predictors, except *D/P*, are now significant for at least one horizon. Consequently, the predictability appears to have increased at the millennium. Moreover, there are several variables whose predictive signs have changed compared to the first subsample (at least for some horizons). For instance, *DFY* and *CAY* now predict long-horizon returns significantly positively. Thus, the results indicate some structural change in commodity markets over time.

– Please insert Table 4 (p. 28) about here. –

Table 4 displays the results of the IS model selection procedure for the original sample period (January 1972 to June 2010). The results for the two subsamples (January 1972 to December 1999 and January 2000 to June 2010) are not reported, but available upon request. First, observe that the procedure selects the specification $n = 0$ for neither forecasting horizon. Instead, seven predictors are chosen for each horizon. Consequently, these results confirm our suggestion from the single regressions that we can reject the null hypothesis of returns that are unpredictable beyond a constant, i.e., returns that are i.i.d. through time. Second, observe that the adj. R^2 again increases with horizons and that they are considerably higher than for the single regressions. Third, note that the variables that are chosen by the model selection procedure, and which thus build the best predictive model, depend on both the horizon and the sample period. Nevertheless, there are some predictors that seem to be particularly important and robust, being represented in most models: *CCM_spot* (19 out of 21 models), *M2* (15 models), *CAY* (14 models), and *SENT* (13 models). (It is remarkable that *CCM_spot* has been insignificant within all single predictive regressions.) Hence, a combination of the current spot "time-series value", monetary policy, consumption-wealth-income-ratio, and investor sentiment seems to represent a good portion of return predictability. *INFL* is the sole variable that is represented in none of the best models.

4.2 Out-of-sample

Panel A of Table 5 displays the OOS forecasting results over the OOS period January 1980 to June 2010. We see that the combining methods proposed by Rapach et al. (2010) perform quite well, with R^2_{OS} between 0.96% (median combination) and 1.65% (mean combination). Moreover, both the p-values from the bootstrap procedure as well as the p-values obtained using Clark and West's (2007) *MSPE-adjusted* statistic indicate statistical significance at the 1% level. According to Campbell and Thompson (2008) and Rapach et al. (2010), even small positive R^2_{OS}, such as the 0.5% for a monthly sample frequency, can indicate a degree of return predictability that is economically meaningful. Thus, the R^2_{OS} we obtain from the combining methods are economically significant and indicate that aggregate returns on commodity futures are predictable OOS.

– Please insert Table 5 (p. 28) about here. –

The R^2_{OS} of the model selection procedure as well as of the individual forecasts paint a picture that is less favorable for the alternative hypothesis of return predictability. First, though the Clark and West (2007) p-values that we obtain for the model selection procedure indicates that the null hypothesis of having a higher mean square prediction error than the historical average forecast can be rejected at the 5% level, both the point estimate of the R^2_{OS} and its p-value obtained from the bootstrap procedure do not support this indication. Second, all but two of the individual forecasts' R^2_{OS} are negative. Some individual forecasts show Clark and West (2007) p-values that indicate that the null hypothesis of having a higher mean square prediction error than the historical average forecast can be rejected at the 10% level and below (*D/P*, *TBL*, *LTY*, *DFY*, *CAY*, *M2*, and *C12CM_spot*). However, both the R^2_{OS} point estimates and the bootstrap p-values speak against this supposition.

The R^2_{OS} statistic measures the closeness of forecasted returns to actually realized returns. Thus, a forecaster who is concerned about the ability of the forecasting model to correctly predict the magnitude of future returns should regard the R^2_{OS} and choose a combining method in our case. However, forecasters who are more interested in a forecasting model's ability to provide the correct sign of future returns (for instance, an investor who is only interested in the correct *direction* for trading activity) should regard the Henriksson-Merton test statistic (Zakamulin 2013). According to the values that we obtain for this measure, the model selection procedure as well as individual forecasts from single predictive regressions, including *SENT*, *TBL*, *LTY*, *DFY*, and *C12CM_spot*, seem to outperform the historical mean model with $HM - S$ values between 1.06 and 1.09 that are statistically significant at the 5% level and below. In contrast, the combining methods do not seem to outperform the historical average with regard to $HM - S$.

Panel B of Table 5 shows the results for the OOS period January 1980 to December 1999. Basically, the combing methods perform, to some extent, better than during January 1980 to June 2010, with R^2_{OS} values between 1.83% (median combination) and 2.88% (mean combination). Moreover, the model selection procedure still performs poorly in terms of R^2_{OS}. Furthermore, there are two individual forecasts that show a performance that is comparable to the combined forecasts: TBL and LTY with R^2_{OS} values of 1.50% and 2.72%, which are statistically significant at the 5% and 1% level, respectively, according to both the p-values from bootstrap and from Clark and West (2007). However, the best performance is shown by the forecast from a predictive regression with *C12CM_spot* as the single predictor variable (the R_{OS}^2 is 4.27% and is statistically significant at the 1% level according to both bootstrap and Clark and West (2007) p-values). Significantly positive $HM - S$ are shown by the individual forecasts from these three variables as well as from *B/M* and *DFY*. Overall, the predictability of the magnitude of future returns as measured by R_{OS}^2 appears to be somewhat higher within the January 1980 to December 1999 period than in the January 1980 to June 2010 period.

Finally, the results for the OOS period January 2000 to June 2010 are displayed in Panel C of Table 5. On the one hand, we see that the R_{OS}^2 values of the three combining methods are much lower than for the two other OOS periods. Only the R_{OS}^2 of the mean combining method is significantly positive, and only according to the p-value obtained from bootstrap. Furthermore, the model selection procedure performs very poorly according to the R_{OS}^2. Moreover, only *SENT* generates an individual forecast whose R_{OS}^2 of 3.80% is significantly positive according to both methods for p-value estimation. On the other hand, the majority of forecasting models (including the three combining methods, but not the model selection procedure) generates significantly positive $HM - S$ values. Thus, the direction of future returns appears to be more predictable within January 2000 to June 2010 than within the two other OOS periods we considered.

Overall, the results indicate that both the magnitude and the sign of future aggregate commodity returns are predictable OOS. However, the predictability with regard to the correct forecast magnitude relies mainly on the application of forecast combing methods, whereas the predictability with regard to the correct forecast direction is mainly based on individual predictive regressions.

5. Discussion

The results of this paper provide new insights into mainly two bodies of literature. First, they contribute to the literature on the time series drivers and predictability of returns on commodity futures. Some key findings of this body of literature thus far include Jensen et al. (2000; 2002), who show that a measure of the U.S. monetary policy significantly predicts the performance and role of commodity futures in mean-variance efficient portfolios. Several studies indicate that returns on commodity futures react to changes in various economic variables, such as bond yields, inflation rates, term spread, and default spread (Erb and Harvey, 2006, and the articles cited therein). Gorton et al. (2013) find that excess returns on commodity futures are predictable by the level of physical inventories and that various price measures, such as the futures basis, prior futures returns, prior spot returns, and spot price volatilities, indicate the state of inventories and accordingly have predictive power over excess returns on commodity futures as well. Moskowitz et al. (2012) find significant time series momentum in commodity futures, that is, that the past 12-month excess return is a positive predictor of the future return.

Our results first contribute to this body of literature by confirming the findings above within a different data set. Accordingly, we demonstrate that the U.S. monetary policy (measured by *M2* in this paper), bond yields (*TBL* and *LTY*), the inflation rate (*INFL*), the term spread (*TMS*), and the default spread (*DFY* and *DFR*), as well as other macroeconomic variables, have predictive power over returns on commodity futures. Moreover, the results confirm that several commodity price measures show some forecasting power (especially the spot price volatility, *CVAR*, and prior spot returns, *C12CM_spot*). Furthermore, the results verify the findings of Moskowitz et al. (2012) and show that there is significant time series momentum (particularly within the IS period January 1972 to December 1999 and the OOS period January 1980 to December 1999). Second, our results contribute to this body of literature by identifying several significant IS predictors that have not yet been considered. For instance, *SENT*, *CL*, *CAY*, and *CRMRF* seem to perform as well as, or even better than, the more common variables mentioned above in predicting aggregate commodity futures returns IS. Third, we contribute to this body of literature by being the first to show that aggregate returns on commodity futures are predictable OOS, especially by the combination forecasts proposed by Rapach et al. (2010).

A task for future research is to connect these predictability patterns to established theories, for instance, to the theory of storage of Kaldor (1939) and others and to refine these theories to account for the observed empirical patterns where required.

Additionally, this paper's results contribute to studies on the predictability of asset returns in general. Much evidence, which is surveyed by Cochrane (1999; 2005, chapter 20; 2011), Spiegel (2008), Koijen and Van Nieuwerburgh (2011) and others, suggests that aggregate excess stock returns are predictable. Some important facts include the following. There are various predictors that perform quite well IS such as low aggregate stock prices relative to fundamentals, including aggregate dividends (Fama and French, 1988), earnings, book values, and moving averages of past prices and that predict higher subsequent stock returns. Moreover, variables from the bond market, such as the term spread, the default spread, and the U.S. Treasury bill rate (Fama and French, 1989), as well as macroeconomic variables, such as the investment-capital ratio (Cochrane, 1991) and the consumption, wealth, income ratio (Lettau and Ludvigson, 2001), show predictive power over future stock returns. However, the OOS performance of many predictive regression models is rather poor, which brings some authors to doubt that aggregate excess stock returns are predictable (e.g., Goyal and Welch, 2008). Nonetheless, more recently, the finance literature developed more sophisticated predictive regression models that show, for stock returns, a convincing OOS performance as well (e.g., Rapach et al., 2010; Ferreira and Santa-Clara, 2011; Kelly and Pruitt, 2013). In an analogous manner to excess stock returns, excess returns on treasuries, bonds, foreign exchange, sovereign debt, and houses appear to be predictable, especially based on various yields or valuation ratios (Cochrane, 2011 and the articles cited therein).

The findings of this paper contribute to this second body of literature by indicating that one should add commodity futures to this long list of predictable assets. Hence, they support the alternative hypothesis that asset returns in general are predictable, that asset returns are not i.i.d. and that asset prices do not follow random walks. In particular, they provide further support to the proposition that asset returns are predictable IS by various variables. The findings also validate that asset returns are predictable OOS when applying more sophisticated forecasting approaches (as shown by the combining methods of Rapach et al., 2010 in this paper), although the OOS performances of many predictive regression models are rather poor in terms of OOS R^2 (in our case, this applies to most of the individual forecasts and the model selection procedure).

6. Conclusions

In this paper, we test whether the null hypothesis of unpredictable aggregate returns on commodity futures can be rejected, and we attempt to identify variables that show predictive power over these returns. For this purpose, we propose a set of 32 candidate predictors that include stock, bond, macroeconomic, and commodity characteristics and test both their IS and OOS forecasting abilities. Our results suggest that many of the candidate predictors have IS predictive power over short- and long-horizon commodity futures returns. Moreover, they indicate that it is possible to forecast returns on commodity futures OOS, especially through the forecast combining methods proposed by Rapach et al. (2010), although the majority of individual forecasts as well as a model selection procedure performs rather poorly in terms of OOS R^2.

Hence, the results of this paper indicate that the null hypothesis, that is, commodity futures returns are unpredictable, can be rejected. Rather, the results support the alternative hypothesis that expected returns on commodity futures depend on factors such as price levels and past price movements, economic conditions and investor sentiment and thus vary over time. Overall, the evidence presented in this paper can be interpreted as one more data point that supports a rejection of the null hypothesis that asset returns *in general* are unpredictable.

The major limitation of this paper is that it is strictly empirical. This paper intentionally does not test – or discuss its results against the background of – existing theories of commodity or, more generally, asset returns. However, our finding that investor sentiment seems to be relatively successful in predicting commodity futures returns makes this discussion quite interesting. This result brings the field of behavioral finance into play against theories of efficient markets and rational investors. Furthermore, we present this factor with just one out of many different possible proxies (Baker and Wurgler, 2007 provide an overview), which constitutes an additional limitation. Hence, predictability tests of commodity futures returns with further sentiment proxies and other potential predictive variables from behavioral finance would be interesting. Overall, we reveal that investigating the underlying economic foundations of the empirical patterns is an important avenue for future research.

Acknowledgments

I extend my sincere appreciation to an anonymous referee, Andreas W. Rathgeber and Stefan Stöckl for their helpful comments.

References

Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, *68*(3), 929-985.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645-1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129-151.

Basu, D., & Miffre, J. (2013). Capturing the risk premium of commodity futures: The role of hedging pressure. *Journal of Banking & Finance*, *37*(7), 2652-2664.

Bossaerts, P., & Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies*, *12*(2), 405-428.

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

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*, 57-82.

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

Cochrane, J. H. (1991). Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance*, *46*(1), 209-237.

Cochrane, J. H. (1999). New facts in finance. *Economic perspectives*, *23*, 36-58.

Cochrane, J. H. (2005). *Asset pricing* (Rev. ed.). Princeton, N.J: Princeton University Press.

Cochrane, J. H. (2011). Presidential address: discount rates. *Journal of Finance*, *66*, 1047-1108.

Cochrane, J. H., & Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, *95*, 138-160.

Erb, C. B., & Harvey, C. R. (2006). The strategic and tactical value of commodity futures. *Financial Analysts Journal*, *62*(2), 69-97.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3-25.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23-49.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3-56.

Fama, E. F., & French, K. R. (2013). A Five-Factor Asset Pricing Model. *Fama-Miller Working Paper*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2287202

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, *100*(3), 514-537.

Gorton, G. B., Hayashi, F., & Rouwenhorst, K. G. (2013). The fundamentals of commodity futures returns. *Review of Finance*, *17*(1), 35-105.

Guo, H. (2006). On the out-of-sample predictability of stock market returns. *Journal of Business*, *79*(2), 645-670.

Hansen, L. P., & Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. *Journal of Political Economy*, *88*(5), 829-853.

Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*(4), 513-533.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, *48*, 65-91.

Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2000). Efficient use of commodity futures in diversified portfolios. *Journal of Futures Markets*, *20*(5), 489-506.

Jensen, G. R., Johnson, R. R., & Mercer, J. M. (2002). Tactical asset allocation and commodity futures. *Journal of Portfolio Management*, *28*(4), 100-111.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*, *7*(1), 1-27.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *Journal of Finance*, *68*(5), 1721-1756.

Koijen, R. S. J., & van Nieuwerburgh, S. (2011). Predictability of returns and cash flows. *Annual Review of Financial Economics*, *3*(1), 467-491.

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, *56*(3), 815-849.

Maio, P., & Santa-Clara, P. (2012). Multifactor models and their consistency with the ICAPM. *Journal of Financial Economics*, *106*, 586-613.

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

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703-708.

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*, 642-685.

Pesaran, M. H., & Timmermann, A. (1992). Is value riskier than growth? *Journal of Business & Economic Statistics*, *10*(4), 461-465.
Polk, C., Thompson, S., & Vuolteenaho, T. (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics, 81*(1), 101-141.
Prokopczuk, M., & Symeonidis, L. (2013). The economic drivers of time-varying commodity market volatility. *Working Paper, Zeppelin University, University of Reading*. Retrieved from http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2013-Reading/papers/EFMA2013_0441_fullpaper.pdf
Rangvid, J. (2006). Output and expected returns. *Journal of Financial Economics*, *81*(3), 595-624.
Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: combination forecasts and links to the real economy. *Review of Financial Studies*, *23*(2), 821-862.
Spiegel, M. (2008). Forecasting the equity premium: where we stand today. *Review of Financial Studies*, *21*(4), 1453-1454.
The Royal Swedish Academy of Sciences (2013). Understanding asset prices. Retrieved from http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2013/advanced-economicsciences2013.pdf
Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies, 21*(4), 1455-1508.
Zakamulin, V. (2013). Forecasting the size premium over different time horizons. *Journal of Banking & Finance*, *37*(3), 1061-1072.

Table 1

Potential predictors of commodity futures returns.

Variable	Description	Studies (among others)	Proxy	Data source	Sample period
Panel A: Stock market					
D/P	Dividend-price ratio	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010)	Log of S&P 500 dividend-price ratio (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
E/P	Earnings-price ratio	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010)	Log of S&P 500 earnings-price ratio (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
D/E	Dividend-payout ratio	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010)	Log of S&P 500 dividend-earnings ratio (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
SVAR	Stock variance	Stocks: Guo (2006); Welch and Goyal (2008); Rapach et al. (2010)	Sum of squared daily S&P 500 returns (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
B/M	Book-to-market ratio	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010)	Dow Jones Industrial Average book-market ratio (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
CSP	Polk et al.'s (2006) cross-sectional premium	Stocks: Polk et al. (2006); Welch and Goyal (2008)	-	Amit Goyal	01/1972 - 12/2002
CRMRF	Cumulative equity premium	-	Five-year cumulative sum of Fama and French's (1993) U.S. market excess return	Kenneth R. French	01/1972 - 06/2010
CL	Stock liquidity	Stocks: Maio and Santa-Clara (2012)	Five-year cumulative sum of Pástor and Stambaugh’s (2003) non-traded liquidity factor	Lubos Pástor	01/1972 - 06/2010
SENT	Investor sentiment	Baker and Wurgler (2006; 2007)	Stock market sentiment index of Baker and Wurgler (2006; 2007)	Jeffrey Wurgler	01/1972 - 06/2010
Panel B: Bond market					
TBL	Treasury bills	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010); Zakamulin (2013)	Three-month U.S. Treasury bill rate (secondary market) (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
LTY	Long-term yield	Stocks: Welch and Goyal (2008); Rapach et al. (2010)	Yield on long-term U.S. government bonds (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
LTR	Long-term return	Stocks: Welch and Goyal (2008); Rapach et al. (2010)	Return on long-term U.S. government bonds (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
TMS	Term spread	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010); Zakamulin (2013)	Long-term yield minus U.S. Treasury bill rate (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
DFY	Default yield spread	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010); Zakamulin (2013)	Yield on BAA- minus yield on AAA-rated corporate bonds (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
DFR	Default return spread	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010)	Return on long-term corporate bonds minus return on long-term U.S. government bonds (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
CP	Cochrane and Piazzesi's (2005) factor	Bonds: Cochrane and Piazzesi (2005); Stocks: Maio and Santa-Clara (2012)	-	John H. Cochrane	01/1972 - 12/2003

Table 1 (continued)

Potential predictors of commodity futures returns.

Variable	Description	Studies (among others)	Proxy	Data source	Sample period
Panel C: Macroeconomics					
INFL	Inflation	Stocks: Welch and Goyal (2008) and the articles cited therein; Rapach et al. (2010); Zakamulin (2013)	U.S. CPI inflation (all urban consumers) lagged by one month (Welch and Goyal 2008)	Amit Goyal	01/1972 - 06/2010
I/K	Cochrane's (1991) investment-to-capital ratio	Cochrane (1991); Welch and Goyal (2008); Rapach et al. (2010)	(The quarterly data are linearly interpolated to obtain monthly data.)	Amit Goyal	01/1972 - 06/2010
CAY	Lettau and Ludvigson's (2001) consumption, wealth, income ratio	Stocks: Lettau and Ludvigson (2001); Welch and Goyal (2008)	(The quarterly data are linearly interpolated to obtain monthly data.)	Amit Goyal	01/1972 - 06/2010
IP	Industrial production	Commodities (volatility): Prokopczuk and Symeonidis (2013)	Five-year log growth of U.S. industrial production	Federal Reserve Bank of St. Louis	01/1972 - 06/2010
M2	M2 money stock	Commodities (volatility): Prokopczuk and Symeonidis (2013)	Three-year log growth of the U.S. M2 money stock	Federal Reserve Bank of St. Louis	01/1972 - 06/2010
GDP	GDP	Stocks: Rangvid (2006)	Three-year log growth of U.S. GDP. (The quarterly data are linearly interpolated to obtain monthly data.)	Federal Reserve Bank of St. Louis	01/1972 - 06/2010
USD	Return on U.S. dollar	Commodities (volatility): Prokopczuk and Symeonidis (2013)	Five-year log return on a trade weighted U.S. dollar index against major currencies	Federal Reserve Bank of St. Louis	01/1978 - 06/2010
CVAL	Asness et al.'s (2013) value "everywhere" factor	-	Five-year log excess return on Asness et al.'s (2013) value "everywhere" factor	Tobias J. Moskowitz	12/1976 - 06/2010
CMOM	Asness et al.'s (2013) momentum "everywhere" factor	-	Five-year log excess return on Asness et al.'s (2013) momentum "everywhere" factor	Tobias J. Moskowitz	12/1976 - 06/2010
Panel D: Commodity market					
CVAR	Commodity variance	-	Sum of squared daily CRB BLS spot index returns	Datastream	01/1972 - 06/2010
CCM_spot	Commodity spot return	-	Five-year log return on CRB BLS spot index	Datastream	01/1972 - 06/2010
CCM	Commodity futures return	-	Five-year log return on Asness et al.'s (2013) commodity futures portfolio	Tobias J. Moskowitz	12/1976 - 06/2010
C12CM_spot	Commodity spot momentum	-	12-month log return on CRB BLS spot index (most recent month's return is skipped)	Datastream	01/1972 - 06/2010
C12CM	Commodity futures momentum	Moskowitz et al. (2012)	12-month log return on Asness et al.'s (2013) commodity futures portfolio (most recent month's return is skipped)	Tobias J. Moskowitz	12/1972 - 06/2010
CVAL_CM	Asness et al.'s (2013) commodity value factor	-	Five-year log excess return on Asness et al.'s (2013) commodity value factor	Tobias J. Moskowitz	12/1976 - 06/2010
CMOM_CM	Asness et al.'s (2013) commodity momentum factor	-	Five-year log excess return on Asness et al.'s (2013) commodity momentum factor	Tobias J. Moskowitz	12/1976 - 06/2010

Table 2

Summary statistics for response and predictor variables.

Variable	Mean	Std.	AC_1
CM	0.0044	0.0433	0.133
D/P	-3.5801	0.4508	0.995
E/P	-2.8186	0.5118	0.990
D/E	-0.7615	0.3426	0.983
SVAR	0.0025	0.0051	0.464
B/M	0.5117	0.3008	0.995
CSP	-0.0013	0.0010	0.947
CRMRF	0.2043	0.3220	0.977
CL	-0.0672	0.5828	0.991
SENT	-0.0260	0.9205	0.986
TBL	0.0561	0.0312	0.987
LTY	0.0761	0.0245	0.990
LTR	0.0074	0.0311	0.039
TMS	0.0200	0.0153	0.947
DFY	0.0111	0.0048	0.963
DFR	-0.0001	0.0141	-0.012
CP	0.0113	0.0242	0.741
INFL	0.0036	0.0038	0.618
I/K	0.0362	0.0035	0.997
CAY	0.0031	0.0224	0.995
IP	0.1209	0.0800	0.992
M2	0.1999	0.0751	0.997
GDP	0.2029	0.0702	0.998
USD	-0.0424	0.1859	0.988
CVAL	0.2003	0.1487	0.972
CMOM	0.3530	0.2171	0.986
CCMspot	0.1636	0.2694	0.991
C12CMspot	0.0316	0.1297	0.957
CVAR	0.0004	0.0006	0.504
CCM	0.1774	0.3906	0.989
C12CM	0.0476	0.1926	0.957
CVAL_CM	0.1558	0.5979	0.988
CMOM_CM	0.5514	0.4122	0.978

Note: This table reports the mean, standard deviation (Std.) and first-order autocorrelation coefficient (AC_1) of the response and predictor variables employed in this study. The sample period for the majority of variables is January 1972 to June 2010. Some variables are only available for a shorter sample period. The data series as well as the sources are described in sections 2.1 and 2.2.

Table 3

Single predictive regressions.

Predictor	*q* = 1	*q* = 3	*q* = 12	*q* = 24	*q* = 36	*q* = 48	*q* = 60
D/P	0.00	-0.01	-0.03	-0.07	-0.16	-0.25	-0.32
	(-0.55)	(-0.66)	(-0.46)	(-0.74)	(-1.23)	(-1.64)	(<u>-2.09</u>)
R^2 (%)	-0.16	-0.03	0.10	1.10	5.36	10.43	14.06
E/P	0.00	-0.01	-0.05	-0.13	-0.25	-0.34	-0.41
	(-0.56)	(-0.84)	(-1.05)	(-1.34)	(*-1.95*)	(<u>-2.46</u>)	(**-2.88**)
R^2 (%)	-0.13	0.17	1.10	3.86	12.00	19.01	23.18
D/E	0.00	0.01	0.06	0.25	0.35	0.43	0.51
	(0.23)	(0.57)	(1.53)	(1.20)	(1.51)	(*1.87*)	(*1.73*)
	[0.22]	[0.50]	[1.28]	[1.05]	[*1.70*]	[**2.65**]	[<u>2.27</u>]
R^2 (%)	-0.21	-0.09	0.75	2.97	5.01	5.64	6.06
CRMRF	-0.01	-0.03	-0.18	-0.34	-0.36	-0.27	-0.24
	(-1.53)	(*-1.70*)	(<u>-2.49</u>)	(**-3.27**)	(**-3.39**)	(<u>-2.34</u>)	(-1.59)
R^2 (%)	0.40	1.29	8.14	13.81	13.00	6.08	3.83
CL	-0.01	-0.02	-0.07	-0.12	-0.15	-0.14	-0.11
	(*-1.70*)	(*-1.83*)	(<u>-2.06</u>)	(<u>-2.33</u>)	(**-2.66**)	(<u>-1.99</u>)	(-1.23)
R^2 (%)	0.57	1.56	4.39	6.13	8.00	5.83	3.02
SENT	-0.01	-0.02	-0.07	-0.11	-0.10	-0.08	-0.05
	(**-2.93**)	(**-3.09**)	(**-3.05**)	(**-2.86**)	(<u>-2.23</u>)	(-1.47)	(-0.79)
R^2 (%)	1.96	4.59	10.89	11.57	8.25	4.63	1.20
TBL	-0.12	-0.39	-1.84	-4.08	-6.17	-6.93	-6.59
	(-1.54)	(*-1.75*)	(<u>-2.31</u>)	(**-4.07**)	(**-6.94**)	(**-5.86**)	(**-3.96**)
R^2 (%)	0.58	1.87	7.21	16.55	32.58	33.96	25.61
LTY	-0.20	-0.62	-2.49	-4.34	-5.97	-7.07	-7.72
	(<u>-2.24</u>)	(<u>-2.40</u>)	(<u>-2.27</u>)	(<u>-2.47</u>)	(**-2.82**)	(**-2.59**)	(<u>-2.52</u>)
R^2 (%)	1.04	3.02	8.45	12.12	19.48	21.58	20.65
TMS	0.01	0.03	1.02	4.78	8.71	10.06	7.98
	(0.03)	(0.08)	(0.60)	(<u>2.15</u>)	(**3.39**)	(**3.34**)	(*1.82*)
R^2 (%)	-0.22	-0.21	0.35	5.82	17.37	18.34	9.49
DFR	0.38	0.59	1.37	1.79	1.17	1.72	1.62
	(1.39)	(1.17)	(<u>2.25</u>)	(*1.80*)	(1.13)	(*1.67*)	(1.59)
R^2 (%)	1.30	0.73	0.61	0.23	-0.08	0.03	-0.04
CAY	-0.23	-0.65	-2.03	-2.45	-2.79	-3.57	-4.48
	(<u>-2.44</u>)	(<u>-2.30</u>)	(-1.40)	(-0.97)	(-1.02)	(-1.06)	(-1.02)
R^2 (%)	1.20	2.81	4.79	3.42	3.72	4.60	5.80
M2	0.05	0.16	0.58	0.87	0.98	1.04	1.09
	(<u>2.04</u>)	(<u>1.97</u>)	(1.52)	(1.35)	(1.37)	(1.28)	(1.13)
R^2 (%)	0.69	1.85	4.36	4.96	5.56	5.14	4.71
CVAR	-4.43	-6.33	21.41	9.94	20.30	37.93	76.01
	(-1.04)	(-0.67)	(0.96)	(0.31)	(0.70)	(1.26)	(*1.72*)
	[-0.94]	[-0.59]	[1.11]	[0.41]	[0.81]	[<u>1.96</u>]	[<u>2.06</u>]
R^2 (%)	0.18	0.00	0.19	-0.20	-0.13	0.07	0.80

Note: This table displays the results for single predictive regressions for the monthly continuously compounded return on an equal-weighted portfolio of 27 commodity futures at horizons q = 1, 3, 12, 24, 36, 48, and 60 months ahead. The returns do not include the returns on collateral from transacting in futures contracts. The original sample is January 1972 to June 2010, and q observations are lost in each of the respective q-horizon regressions. The first line corresponding to each model reports the slope estimates. Line 2 reports the Newey-West t-ratios (in parentheses), and, where it is required, line 3 shows the Hansen-Hodrick t-ratios (in brackets), both computed with q lags. Italic, underlined, and bold t-statistics denote statistical significance according to the standard normal distribution at the 10%, 5%, and 1% levels, respectively. The last line, R^2 (%), shows the values of the adjusted coefficient of determination (in %). The table reports only those regression models whose slope coefficients are significant for at least one horizon according to either the Newey-West or Hansen-Hodrick t-ratios.

Table 4

Multiple predictive regressions.

	D/P	*E/P*	*D/E*	*SVAR*	*B/M*	*CRMRF*	*CL*	*SENT*	*TBL*	*LTY*	*LTR*	*TMS*	*DFY*	*DFR*	*INFL*	*I/K*	*CAY*	*IP*	*M2*	*GDP*	*CCM_spot*	*C12CM_spot*	*CVAR*	R^2 (%)
q = 1	–	–	0.01	–	–	–	–	-0.01	–	–	–	–	–	0.33	–	–	-0.43	–	–	–	-0.04	0.05	-8.32	7.85
			(1.30)					(**-3.90**)						(1.24)			(**-4.27**)				(**-3.41**)	(2.36)	(-2.45)	
			[1.24]					[**-3.80**]						[1.24]			[**-4.13**]				[**-3.30**]	[2.32]	[-2.38]	
q = 3	0.10	–	–	–	-0.13	–	–	-0.04	–	–	–	–	–	–	–	–	-2.05	–	–	–	-0.16	0.11	-16.63	17.01
	(2.30)				(*-1.78*)			(**-4.39**)									(**-3.94**)				(**-3.59**)	(2.34)	(-2.48)	
	[2.00]				[-1.56]			[**-3.76**]									[**-3.33**]				[**-3.05**]	[2.12]	[-2.37]	
q = 12	–	–	–	–	–	-0.30	0.23	-0.17	–	–	–	–	–	–	–	–	-4.08	–	0.77	–	-0.54	0.32	–	47.40
						(**-4.21**)	(**3.16**)	(**-5.77**)									(**-3.90**)		(2.27)		(**-4.85**)	(2.37)		
						[**-3.76**]	[**2.72**]	[**-5.09**]									[**-3.59**]		[*1.88*]		[**-4.53**]	[2.02]		
q = 24	–	–	–	–	–	-0.64	0.37	-0.21	–	-2.95	–	–	–	–	–	–	-3.95	–	1.71	–	-0.73	–	–	61.59
						(**-4.45**)	(**2.86**)	(**-3.62**)		(*-1.66*)							(-1.64)		(**2.69**)		(**-4.56**)			
						[**-3.91**]	[**2.60**]	[**-3.21**]		[-1.64]							[-1.50]		[2.52]		[**-4.44**]			
q = 36	–	–	0.41	–	–	-0.37	–	–	-6.36	–	–	–	–	–	–	–	-2.87	0.99	1.23	–	-0.32	–	–	65.35
			(1.63)			(**-2.79**)			(**-7.37**)								(-1.12)	(2.44)	(2.41)		(-2.44)			
			[1.47]			[**-2.66**]			[**-6.50**]								[-1.04]	[**2.62**]	[**2.85**]		[-2.09]			
q = 48	–	–	0.45	–	–	–	–	-0.08	-4.98	–	–	–	–	–	–	–	-4.75	–	2.23	-1.63	-0.32	–	–	64.61
			(*1.80*)					(*-1.83*)	(**-2.60**)								(-1.48)		(**4.76**)	(-0.93)	(*-1.81*)			
			[*1.66*]					[**-3.32**]	[-2.43]								[-1.46]		[**10.14**]	[-0.83]	[*-1.73*]			
q = 60	–	-0.67	–	–	–	-0.48	0.16	-0.13	–	–	–	–	11.17	–	–	–	–	–	2.87	–	-0.33	–	–	62.40
		(**-3.42**)				(*-1.81*)	(0.83)	(-2.36)					(0.86)						(**3.11**)		(-1.18)			
		[**-2.85**]				[-1.58]	[0.79]	[**-2.71**]					[0.88]						[**3.84**]		[-1.00]			

Note: This table displays the results for multiple-variable predictive regressions for the monthly continuously compounded return on an equal-weighted portfolio of 27 commodity futures at horizons q = 1, 3, 12, 24, 36, 48, and 60 months ahead. The returns do not include the returns on collateral from transacting in futures contracts. The multiple-variable models are chosen by a model selection procedure that searches for the model with the maximum adjusted R-squared out of all possible combinations of the variables shown in the table, with a maximum of 7 variables at once. The original sample is January 1972 to June 2010, and q observations are lost in each of the respective q-horizon regressions. The first line corresponding to each model reports the slope estimates. A hyphen for the slope estimate means that the selection procedure does not include this variable in the model. Line 2 reports the Newey-West t-ratios (in parentheses), and line 3 shows the Hansen-Hodrick t-ratios (in brackets), both computed with q lags. Italic, underlined, and bold t-statistics denote statistical significance according to the standard normal distribution at the 10%, 5%, and 1% levels, respectively. The last line, R^2 (%), shows the values of the adjusted coefficient of determination (in %).

Table 5

Out-of-sample forecasting.

Comb. method or predictor	R^2_{OS} (%)	p(CW)	$HM-S$	Predictor	R^2_{OS} (%)	p(CW)	$HM-S$
Panel A: OOS period: January 1980 to June 2010							
Mean	**1.65**	0.002	1.00	*LTR*	-0.42	0.438	0.99
Median	**0.96**	0.000	0.99	*TMS*	-2.56	0.918	1.00
Trimmed mean	**1.37**	0.002	1.00	*DFY*	-1.98	0.021	<u>1.06</u>
Model selection	-7.36	0.018	<u>1.06</u>	*DFR*	*0.30*	0.210	0.97
D/P	-2.72	0.075	1.02	*INFL*	-1.13	0.510	0.99
E/P	-2.22	0.207	1.02	*I/K*	-2.93	0.922	1.00
D/E	-1.39	0.615	1.01	*CAY*	-3.63	0.005	0.98
SVAR	-4.66	0.691	1.01	*IP*	-5.19	0.738	0.99
B/M	-2.58	0.256	*1.03*	*M2*	-1.34	0.002	1.01
CRMRF	-0.71	0.279	0.98	*GDP*	-0.57	0.477	0.99
CL	-2.97	0.737	0.97	*CCM_spot*	-0.95	0.666	0.97
SENT	-2.78	0.450	**1.09**	*C12CM_spot*	-0.27	0.035	<u>1.08</u>
TBL	-0.33	0.042	<u>1.07</u>	*CVAR*	*0.28*	0.218	1.01
LTY	-0.21	0.001	<u>1.06</u>				
Panel B: OOS period: January 1980 to December 1999							
Mean	**2.88**	0.002	1.00	*LTR*	-0.37	0.415	1.00
Median	**1.83**	0.000	0.99	*TMS*	-4.56	0.908	1.00
Trimmed mean	**2.52**	0.002	1.00	*DFY*	-3.08	0.019	*1.04*
Model selection	-9.77	0.044	1.04	*DFR*	-0.36	0.489	0.98
D/P	-4.91	0.071	1.02	*INFL*	-1.61	0.441	0.98
E/P	-3.82	0.182	1.03	*I/K*	-5.13	0.905	0.98
D/E	-1.09	0.603	1.00	*CAY*	-6.31	0.007	0.89
SVAR	-8.37	0.757	0.99	*IP*	-8.89	0.701	0.98
B/M	-4.61	0.240	*1.04*	*M2*	-2.58	0.002	0.95
CRMRF	-1.34	0.357	0.96	*GDP*	-0.34	0.332	1.00
CL	-6.01	0.793	0.93	*CCM_spot*	-0.98	0.550	0.95
SENT	-8.18	0.729	1.03	*C12CM_spot*	**4.27**	0.000	<u>1.08</u>
TBL	<u>1.50</u>	0.025	<u>1.10</u>	*CVAR*	0.02	0.400	1.00
LTY	**2.72**	0.001	*1.06*				

Note: The table reports the results of out-of-sample (OOS) forecasts of the monthly return on an equal-weighted portfolio of 27 commodity futures. The returns do not include the returns on collateral from transacting in futures contracts. The forecasts are obtained from individual predictive regressions, from mean, median, and trimmed mean combining methods as well as from a model selection procedure. The OOS evaluation periods are January 1980 to June 2010 (Panel A), January 1980 to December 1999 (Panel B), and January 2000 to June 2010 (Panel C). The in-sample estimation periods start eight years prior to the OOS periods. R^2_{OS} (%) denotes the Campbell and Thompson (2008) OOS R-squared statistic (in %), and $HM-S$ denoted the Henriksson-Merton test statistic. Italic, underlined, and bold indicate statistical significance according to a bootstrap procedure at the 10%, 5%, and 1% levels, respectively. p(CW) is the p-value for rejecting the null hypothesis $R^2_{OS} \leq 0$ according to the MSPE-adjusted statistic of Clark and West (2007).

Table 5 (continued)

Out-of-sample forecasting.

Comb. method or predictor	R^2_{OS} (%)	p(CW)	$HM-S$	Predictor	R^2_{OS} (%)	p(CW)	$HM-S$
Panel C: OOS period: January 2000 to June 2010							
Mean	<u>1.05</u>	0.149	<u>1.08</u>	*LTR*	-1.25	0.686	<u>1.09</u>
Median	-0.24	0.690	<u>1.06</u>	*TMS*	-0.84	0.584	1.03
Trimmed mean	0.37	0.233	<u>1.08</u>	*DFY*	-4.61	0.603	1.04
Model selection	-19.85	0.037	1.07	*DFR*	-2.26	0.269	0.95
D/P	-0.89	0.884	**1.17**	*INFL*	-2.38	0.657	1.03
E/P	-2.35	0.675	<u>1.14</u>	*I/K*	*0.58*	0.174	<u>1.11</u>
D/E	-2.60	0.488	<u>1.10</u>	*CAY*	-1.23	0.824	<u>1.14</u>
SVAR	-7.15	0.175	<u>1.09</u>	*IP*	-0.59	0.399	*1.07*
B/M	-0.53	0.833	*1.07*	*M2*	-1.17	0.862	<u>1.11</u>
CRMRF	-0.10	0.281	*1.07*	*GDP*	-2.42	0.771	*1.08*
CL	0.08	0.266	1.01	*CCM_spot*	-1.85	0.807	0.93
SENT	**3.80**	0.001	*1.08*	*C12CM_spot*	-1.68	0.681	<u>1.10</u>
TBL	-1.47	0.654	1.06	*CVAR*	-23.63	0.094	1.04
LTY	-2.19	0.828	<u>1.10</u>				