Heliyon 5 (2019) e01843

Contents lists available at ScienceDirect

Heliyon

www.elsevier.com/locate/heliyon

Return predictability in metal futures markets: new evidence ☆

Björn Tharann *

Leibniz University Hannover, Koenigsworther Platz 1, D-30167 Hannover, Germany

ARTICLE INFO

Keywords:
Economics
Commodities
Metal futures
Return predictability

ABSTRACT

This paper studies the predictability of metal futures returns. Additionally, we identify years of high predictability. Generally, we find a substantial degree of predictability both in- and out-of-sample. Gold returns seem to be best predictable out-of-sample. A timing strategy leads to utility gains of 2.18% p.a. In particular, the Aruoba–Diebold–Scotti (ADS) business conditions index incorporates relevant information for metal returns, and strongly predicts gold returns.

1. Introduction

Are returns predictable? This question has been analyzed at least since one century. Initial attempts to predict stock returns were already performed by Dow (1920). Numerous studies have tackled the question of return predictability and provided evidence either in favor or against predictability. Goyal and Welch (2008) argue that the historical mean is a tough benchmark to beat and that the so far observed predictability is mainly driven by the period of the oil crises. In contrast, Campbell and Thompson (2008) provide evidence in favor of return predictability when including two economically motivated restrictions. Cochrane (2008) shows that return predictability results from the time-variation of expected returns rather than dividend growth, and thus contradicts the random walk hypothesis. Overall, return predictability is challenging and the answer of it is still inconclusive.

We accept this challenge and analyze the predictability of metal futures returns. The importance of commodities has increased dramatically since the beginning of the millennium. In December 2000, the release of the Commodity Futures Modernization Act (CFMA) took place, which is typically regarded as the beginning of the financialization of commodities. As a consequence, the trading of commodity futures has been facilitated and commodities have moved more into the focus of institutional investors for the purpose of diversification, asset allocation as well as risk management. From a practical standpoint, to have knowledge about future price developments of commodities goes along with an improved investment performance. Investors may better select assets for their asset allocations. Due to the performance of commodities, investors have detected commodities as a new investment class (e.g., Bessembinder, 1992; Gorton and Rouwenhorst, 2006). Erb and Harvey (2006) show that commodity portfolios have similar average returns than stock and bond portfolios.

Commodities exhibit advantageous properties that improve the portfolio performance. They are used for diversification due to a low correlation with stocks and bonds, and due to a positive correlation with inflation and unexpected inflation, commodities serve as hedge against inflation (e.g., Gorton and Rouwenhorst, 2006, Symeonidis et al., 2012). Gorton and Rouwenhorst (2006) show that commodities perform well in the early stage of recessions, when stocks typically underperform. Belousova and Dorfleitner (2012) analyze the diversification benefits from the perspective of a euro investor and provide evidence that industrial metals lead to a reduction of risk, whereas precious metals to both a risk reduction as well as an improvement of return. Bessler and Wolff (2015) analyze numerous asset-allocation strategies and provide evidence that in particular metal commodities significantly improve the performance of traditional stock-bond-portfolios. Hammoudeh et al. (2011) and Zhang and Zhang (2016) refer to the necessity to adjust risk management metrics for precious metals, since the prices of precious metals are more volatile than their historical trend. Hammoudeh et al. (2013) argue that these adjustments increase the regulatory compliance by affecting the daily capital charges under the Basel Accord rule.

The main goal of this study is to provide evidence on return predictability of metal commodity futures. In doing so, we make two contributions. First, we analyze not only the predictability in-sample, but also out-of-sample. Here, we use five distinct time series of metal futures and 12 variables that are supposed to predict stock returns. Moreover,

☆ We are grateful to Paige Shaklee (the editor), an anonymous associate editor, an anonymous referee, Fabian Bätje, Fabian Hollstein, and Marcel Prokopczuk for helpful comments and suggestions.

* Corresponding author.
E-mail address: tharann@fmt.uni-hannover.de.

https://doi.org/10.1016/j.heliyon.2019.e01843
Received 2 January 2019; Received in revised form 26 April 2019; Accepted 24 May 2019

Table 1
Summary statistics metal futures excess returns. This table summarizes (annualized) key statistics for metal futures excess returns. "Mean", "Std Dev", "Skew", and "Kurt" denote the mean, standard deviation, skewness, and kurtosis, respectively. The next three columns show the first-order autoregressive coefficient and the p-value of the Jarque-Bera and Augmented Dicky Fuller test, respectively. "First Obs." and "Nobs" denote the first observation of the time series and the number of observations. All data are sampled at the monthly frequency.

Commodity	Mean	Std Dev	Skew	Kurt	AR(1)	JB p-value	ADF p-value	First obs.	Nobs
Copper	0.03	0.28	0.21	3.76	0.93	<0.01	<0.05	31.08.1988	342
Gold	0.04	0.14	0.30	2.94	0.90	<0.1	<0.05	31.08.1988	342
Palladium	0.08	0.41	-0.38	2.71	0.93	<0.05	0.26	28.02.1995	264
Platinum	0.02	0.21	-0.20	3.46	0.91	<0.1	<0.01	31.08.1988	342
Silver	0.04	0.24	0.68	3.75	0.89	<0.01	<0.01	31.08.1988	342

we focus on several sample periods and use different techniques to identify years of high predictability. We also use forecast combinations to improve the out-of-sample predictability. In addition, we do not only analyze the return predictability, but also the economic value that arises if an investor can utilize knowledge about future price movements.

Second, we introduce and analyze the ADS index, developed by Aruoba et al. (2009), to examine the potential effects on metal futures returns and on the behavior over business cycles. The ADS index claims to accurately measure business conditions in real-time. Through the use of information capturing potential co-movements of business cycles at different time frequencies, the ADS index complements the information content of less frequently published variables such as the industrial production and the unemployment rate. Moreover, since the ADS index is reflecting information on business cycle states, it might affect the return predictability in both equity and commodity markets as a result of a sufficient market integration. As a consequence, the index should incorporate information that are relevant for metal commodities.

We find excessive evidence for predictability for the next year's excess return across all metal commodities, both in- and out-of-sample. The best performing variable is the long-term government bond yield, indicated by an out-of-sample R^2 of 37.90% in the case of gold. The mean forecast combination approach provides evidence for an improved and especially stable out-of-sample predictability, in particular for gold and platinum, indicated by out-of-sample R^2s up to 18.57%. Gold returns seem to be best predictable out-of-sample. A timing strategy leads to utility gains of 2.18% p.a., when using aggregated information in predicting gold returns.

The ADS index shows remarkable correlations with metal futures returns, particularly in recessions. We also find that the ADS index strongly predicts gold returns, indicated by an out-of-sample R^2 of 8.21%. Multiple predictive regressions confirm the strong predictive power of the ADS index.

Our study directly relates to the literature on stock return predictability. Initial studies used aggregated valuation ratios as the dividend–price ratio (e.g., Rozeff, 1984; Fama and French, 1988b), short-term interest rates (e.g., Campbell, 1987; Ang and Bekaert, 2007), and the consumption–wealth ratio (e.g., Lettau and Ludvigson, 2001). Campbell and Shiller (1988, 1998) introduced and comprehensively examined the earnings–price ratio, and provide evidence that this ratio especially predicts long-term stock returns. Goyal and Welch (2003, 2008) analyze numerous financial and macroeconomic variables, reexamine previous studies and identify the years of the oil crises as the main drivers for predictability. Campbell and Thompson (2008) show that there exists predictability when imposing two economically motivated restrictions. Rapach et al. (2010) and Rapach and Zhou (2013) document that combination forecasts might lead to out-of-sample improvements.

Our study also relates to the literature on commodities and commodity return predictability. Fama and French (1988a) analyze the behavior of metal spot and futures prices over business cycles and provide evidence that the prices are affected by the level of inventory and the business cycle stage. Numerous papers examine aggregated valuation ratios known from the stock return predictability literature, e.g. Bessembinder (1992), Bailey and Chan (1993), and Bjornson and Carter (1997). De Roon et al. (2000) use hedging pressure as predictive variable. Gargano and Timmermann (2014) use commodity spot indices to analyze the predictability of commodities over a longer time period. Nguyen et al. (2019) show that gold returns are predictable by the jump tail premium and variance risk premium. Jordan et al. (2018) analyze metal commodities in the G7 countries. Prokopczuk et al. (2018) examine the return predictability of commodities using spot prices of more than 140 years, and find evidence particularly for longer horizons.

The remainder of this paper is structured as follows. Section 2 describes the data, the computation of the variables, and the methodology. Section 3 provides the main empirical results. Section 4 shows the results related to the ADS index. Finally, Section 5 concludes.

2. Methodology

2.1. Data

We obtain our data from several sources. We use monthly settlement prices of five continuously rolled metal futures retrieved from Datastream: high grade copper (NHGCS00), gold (NGCCS00), palladium (NPACS20), platinum (NPLCS00), and silver (NSLCS00). The commodities are traded on exchanges in the U.S. and are denominated in U.S. Dollar (USD). Table 1 provides an overview about the metal futures. Our sample period spans from August 1988 to December 2017.

2.2. Variables

Metal futures excess returns. To base our analysis on the most conservative procedure, we compute the log-return on a fully collateralized futures contract (Gorton et al., 2013; Bakshi et al., 2017) as:

$$r_{t+1} = log\left(\frac{F_{t+1,T}}{F_{t,T}}\right) + rf_t, \tag{1}$$

where $F_{t+1,T}$ and $F_{t,T}$ are the settlement prices on the continuous futures contract with maturity T at the end of month $t+1$ and t, respectively. The interest rate on a fully collateralized position is denoted by rf_t. The corresponding futures excess return, i.e. er_{t+1}, is then defined as:

$$er_{t+1} = r_{t+1} - rf_t. \tag{2}$$

Predictors. To analyze the predictability of metal futures, we use 12 variables that appear to have predictive power for stock returns. We motivate the predictive power of these variables by the fact that the state of the economy has an influence on (metal) commodity prices

due to short-term imbalances between demand and supply for those as well as due to different financing costs in the short-term (Gargano and Timmermann, 2014). Moreover, due to the integration of equity and commodity markets, both markets might be affected by similar risk factors, which, in turn, are affected by the state of the economy in both the short- and long-term. Bessembinder (1992) and Hollstein et al. (2019) provide evidence for a good but not perfect integration of both markets.

Following Goyal and Welch (2008), we take the following *stock predictive variables*: the monthly dividend–payout ratio (de) as the difference between the log of dividends and the log of earnings,[1] the monthly default return spread (dfr) as the difference between long-term corporate bond returns and long-term government bond returns, the monthly default yield spread (dfy) as the difference between BAA- and AAA-rated corporate bond yields, the monthly dividend yield (dy) as the difference between the log of dividends and the log of lagged prices, the monthly earnings–price ratio (ep) as the difference between the log of earnings and the log of prices, the monthly inflation rate ($infl$) as the return on the U.S. consumer price index, the monthly long-term rate of returns on U.S. government bonds (ltr), the monthly long-term U.S. government bond yields (lty), the monthly stock variance ($svar$) as the sum of squared daily returns on the S&P 500, and the monthly term spread (tms) as the difference between long-term U.S. government bond yields and the 3-month U.S. Treasury bill rate.[2]

Moreover, we follow Gargano and Timmermann (2014) and use as *macroeconomic predictive variables* the change in industrial production ($\Delta Indpro$) as the log difference between monthly industrial production, and the unemployment rate ($unrate$), from the Federal Reserve Bank of St. Louis (FRED).[3] To gain insights about the behavior of metal commodities over business cycles, we use the ADS index from the Federal Reserve Bank of Philadelphia.

2.3. In- and out-of-sample return predictability

In-sample analysis. To evaluate the in-sample predictive power of a variable, we follow Rapach and Wohar (2006) and estimate the next year's excess return as:

$$er_{t+12} = \alpha + \beta X_t + \epsilon_{t+12}, \tag{3}$$

where er_{t+12} is the 12-months' ahead excess return from month t to $t+12$.[4] α, β, and ϵ_{t+12} are the intercept and the slope parameters, respectively, and the error term over the next year. X_t represents the predictor at month t.

In detail, to assess the in-sample predictability, we impose the null hypothesis of no predictability (H_0), i.e.,

$$er_{t+12} = \alpha + \epsilon_{t+12}, \tag{4}$$

defining the restricted model, where $\beta = 0$. Thus, under H_0 the excess return cannot be predicted using X_t, and we would expect that the slope estimate is not significantly different from zero. In that case, the excess return follows a random walk process, and the best estimate of the future excess return is just its historical mean. Under the alternative hypothesis of predictability (H_1), defined as the unrestricted model in Equation (3), the excess return can be predicted using X_t. Thus, we would expect that β is significantly different from zero.

To evaluate the significance of predictability, we use the bootstrap algorithm, proposed by Rapach and Wohar (2006), in order to obtain reliable statistical inferences.[5] As a consequence, we avoid the well known statistical issues of a small sample bias (Stambaugh, 1999) and serial correlation in the error terms (Richardson and Stock, 1989).

Out-of-sample analysis. Our out-of-sample analysis is based on the methodology used by, e.g., Goyal and Welch (2003, 2008). To obtain the first out-of-sample forecast, we proceed as follows: First, we estimate Equation (3) using an initial training window of 10 years.[6] We then use the obtained estimated parameters and the most recent observation of the predictor, to calculate the corresponding forecast. Third, we repeat that procedure by rolling the window by one observation ahead and estimate the next year's excess return.

To assess the out-of-sample predictability, we follow Campbell and Thompson (2008) and compute the out-of-sample R^2, i.e.,

$$R^2_{oos} = 1 - \frac{MSE_u}{MSE_r}, \tag{5}$$

where MSE_u and MSE_r are the mean squared errors of the unrestricted and restricted model, respectively. The R^2_{oos} represents a relative measure of two competing nested models. Thus, by using that measure we ask the question: How large is the improvement of the predictive power using the variable X_t in relation to the predictive power using the historical mean as naive benchmark? A variable is considered to have predictive power, if it is associated with a positive and significant R^2_{oos}. Since that measure represents a point estimate, we have to carefully assess the degree of significant predictability.[7] In doing so, we follow McCracken (2007) and compute the $MSE-F$ statistic, i.e.,

$$MSE-F = (N-k+1)\cdot\left(\frac{MSE_r - MSE_u}{MSE_u}\right), \tag{6}$$

where N is the number of out-of-sample forecasts. k is the degree of overlap, thus, in our case 12. In accordance to the previous section, under H_0 the restricted model performs at most as well as the unrestricted model ($MSE_r \leq MSE_u$).

To further analyze the predictive (in- and out-of-sample) performance of the variable X_t over time, we plot and analyze the cumulative differences in squared forecast errors (CDSFE). The in-sample performance is the difference between the cumulative squared demeaned excess return from the restricted model and the cumulative squared regression residual from the unrestricted model, whereas the out-of-sample performance is the difference between the cumulative squared forecast error from the restricted model and the cumulative squared forecast error of the unrestricted model.

Similar to the R^2_{oos}, the CDSFE also go along with a relative interpretation. Whenever the variable X_t has superior (inferior) predictive power relative to the historical mean benchmark, we expect an increase (a decrease) in the CDSFE. Thus, an absolute increase (decrease) in the CDSFE is not interpretable. Moreover, the CDSFE allow an analysis of the time-varying predictability of variable X_t to time periods of high and low predictability, respectively (Goyal and Welch, 2008).[8]

[1] The dividends (earnings) are usually defined as the 12-months moving sums of dividends payed on the S&P 500 stock index.

[2] We use the extended data set of Goyal and Welch (2008) that can be found at http://www.hec.unil.ch/agoyal/. Further variables known to predict stock returns are the dividend–price ratio and the 3-month U.S. Treasury bill rate. However, due to high correlations with dy (99%) and tms (−63%), respectively, we do not include these variables into our set of potential predictors.

[3] The corresponding tickers are "INDPRO" for industrial production and "UNRATE" for the unemployment rate.

[4] At each point in time t, we sum up the monthly excess returns from month t to $t+12$, thus, we end up with overlapping annual excess returns.

[5] A detailed description of the procedure can be found in Section 2.4.

[6] Here, we follow Çakmaklı and van Dijk (2016) and use a rolling training window of 10 years to capture the average length of a business cycle. Thus, we also take care about potential structural breaks in the time series.

[7] For more detailed information, see Section 2.5.

[8] The sign of the final value of the out-of-sample CDSFE is equal to the sign of the estimated R^2_{oos}. In our CDSFE plots, we standardize the in-sample CDSFE to zero at the date of the first out-of-sample forecast, by shifting the curve vertically downwards. Due to the apparent sensitivity of the forecasting accuracy of the $R^2_{oos}s$, it is necessary to additionally assess the degree of significance rather than relying on the absolute amount of the $R^2_{oos}s$ only.

2.4. Bootstrap procedure

To implement the bootstrap algorithm, we follow Rapach and Wohar (2006). In doing so, we assume a data generating process under the null hypothesis of no predictability, i.e.:

$$er_t = a_0 + \epsilon_{r,t}, \tag{7}$$

$$X_t = b_0 + b_1 X_{t-1} + \epsilon_{u,t}, \tag{8}$$

where er_t and X_t are the excess return and the predictive variable at month t, respectively. a_0, b_0 and b_1 are the intercept and slope parameters, respectively. $\epsilon_t = (\epsilon_{r,t}, \epsilon_{u,t})'$ is a vector of errors that are assumed to be independently and identically distributed. We assume that the predictive variable follows an AR(1) process (Goyal and Welch, 2008).

Next, we estimate Equations (7) and (8) via OLS and obtain the corresponding residuals, i.e., $\hat{\epsilon}_t = (\hat{\epsilon}_{r,t}, \hat{\epsilon}_{u,t})'$. Afterwards, we generate a series of pseudo errors $\{\hat{\epsilon}_t^*\}_{t=1}^{T+100}$, by drawing randomly with replacement $T + 100$ times from the OLS residuals. To retain the contemporaneous structure between the errors, we draw from the OLS residuals in tandem.

To compute our pseudo sample of $T + 100$ observations for er_t and X_t, i.e., $\{er_t^*, X_t^*\}_{t=1}^{T+100}$, we proceed as follows. First, we define $\hat{a}_0$, $\hat{b}_0$, and $\hat{b}_1^A$ as the OLS estimates of the intercept and slope parameters in Equations (7) and (8), respectively, where the bias adjustments in Shaman and Stine (1988) are used. Second, we take the estimates and $\{\hat{\epsilon}_t^*\}_{t=1}^{T+100}$, and using Equations (7) and (8) to compute our pseudo sample. Here, we set the initial values in Equation (8) equal to zero. Third, we drop the first 100 observations of our pseudo sample to obtain the same size as our original sample.

Next, using our pseudo sample, we compute the (in-sample) t-statistic of the unrestricted model, and the (out-of-sample) $MSE - F$ statistic. We repeat the algorithm 1,000 times obtaining an empirical distribution of the respective statistic. Finally, to compute the p-values, we also calculate the in- and out-of-sample statistics using the original sample. The corresponding p-value is defined as the proportion of the respective bootstrapped statistic that is greater than the real statistic.

2.5. MSFE-adjusted test statistic

The R^2_{oos} is a point estimate, thus, the forecast accuracy is sensitive, among others, to the sample size (Zhu and Zhu, 2013). To test whether the unrestricted and the restricted models are statistically different, we can use the MSFE-adjusted test statistic, developed by Clark and West (2007). The statistic is an adjusted version of the Diebold and Mariano (1995) and West (1996) statistic and examines the null hypothesis that $R^2_{oos} \leq 0$. Thus, the statistic is applicable for nested models. The asymptotic distribution of the nested model forecasts can be well approximated by the standard normal distribution. Moreover, in finite samples the MSFE-adjusted test statistic also performs quite well (Rapach and Zhou, 2013).

Following Rapach and Zhou (2013), we divide the number of total observations T into an in-sample estimation period comprising the first R observations and an out-of-sample period comprising the last $N = T - R$ observations, where $s = 1, ..., N$. To compute the MSFE-adjusted test statistic, we first define:

$$\tilde{d}_{R+s} = \hat{\epsilon}^2_{r,R+s} - [\hat{\epsilon}^2_{u,R+s} - (\bar{er}_{R+s} - \hat{er}_{u,R+s})^2], \tag{9}$$

where, $\hat{\epsilon}^2_r$ and $\hat{\epsilon}^2_u$ are the squared out-of-sample errors from the restricted and unrestricted model, respectively. $\bar{er}$ is the average excess return, and $\hat{er}_u$ the forecast of the excess return of the unrestricted model based on predictor X_t. Finally, we regress $\tilde{d}_{R+s}$ on a unit vector of length N without intercept. The MSFE-adjusted test statistic is then equal to the corresponding t-statistic, considered as one-sided test.

Table 2
Correlation matrix metal futures excess returns. This table reports the correlations among all annual metal futures excess returns. All data are sampled at the monthly frequency.

Commodity	Copper	Gold	Palladium	Platinum	Silver
Copper					
Gold	0.47				
Palladium	0.54	0.10			
Platinum	0.70	0.54	0.59		
Silver	0.55	0.75	0.43	0.61	

3. Results

3.1. Summary statistics

Before discussing our main empirical results, it is instructive to look at some summary statistics and correlations. Table 1 provides summary statistics on the annual metal futures excess returns. We observe that the average return ranges between 2% for platinum and 8% for palladium, which is consistent with former studies of, e.g., Gorton et al. (2013). We also notice that gold exhibits the smallest standard deviation of 14%, whereas copper has a standard deviation of 28%.

Table 2 shows correlations between the metal futures excess returns. We observe that gold and silver exhibit a high correlation of 0.75, further, silver and platinum of 0.61, and gold and platinum of 0.54, which is not surprising given that these commodities belong to the class of precious metals. Moreover, copper exhibits a high correlation with platinum of 0.70. Tables 3 and 4 report summary statistics on and correlations between the predictive variables, which are consistent with, e.g., Goyal and Welch (2008).

3.2. Return predictability

We start by analyzing the performance of variables predicting the next year's excess return on the basis of univariate regressions. Table 5 reports the in-sample and out-of-sample results.

In-sample results. We find an extensive degree of predictability across all metal commodities. In particular, de, dfy, and ep exhibit significant predictive power across all metals, indicated by their significant t-statistics, among others, of 4.46, 4.85, and −8.13 in the case of platinum. It is also worth analyzing the predictive power of the individual variables. We observe highly significant in-sample R^2s of 5.55%, 6.48%, and 16.33%, respectively.

Further variables showing substantial predictive power are, among others, $\Delta Indpro$ in the case of gold ($R^2 = 1.50\%$), dy in the case of palladium ($R^2 = 5.63\%$), $infl$ in the case of platinum ($R^2 = 3.05\%$), lty in the case of gold ($R^2 = 2.86\%$), $svar$ in the case of platinum ($R^2 = 5.37\%$), tms in the case of silver ($R^2 = 5.60\%$), and $unrate$ in the case of silver ($R^2 = 3.35\%$). We take note that dfr and ltr do not have significant power in predicting next year's metal futures excess returns.

Out-of-sample results. Now, we translate our analysis to the out-of-sample predictability and examine whether the strong in-sample predictability also holds out-of-sample. Table 5 shows a high degree of return predictability by the variables. This is true for all variables, except dfr and ltr. Among many others, we find strong predictive power for tms in the case of copper ($R^2_{oos} = 5.06\%$), dfy in the case of gold ($R^2_{oos} = 10.99\%$), ep in the case of platinum ($R^2_{oos} = 13.75\%$), dy in the case of palladium ($R^2_{oos} = 19.58\%$), and lty in the case of gold ($R^2_{oos} = 37.90\%$). These results are interesting, given that, e.g., Goyal and Welch (2008) argue that the historical mean is a tough benchmark to beat in the case of stock return predictability.

Overall, we find a substantial degree of predictability across all metal commodities for almost all variables. We do not only find evidence for in-sample predictability, but also strong evidence for out-of-sample predictability. The results are consistent with those in Fama and

Table 3
Summary statistics predictive variables. This table summarizes (non-annualized) key statistics about the predictive variables. "de" denotes the dividend–payout ratio, "ΔIndpro" the growth of industrial production. "dfr" is the default return spread as the difference between long-term U.S. corporate bond returns and long-term U.S. government bond returns. "dfy" is the default yield spread as the difference between U.S. BAA- and AAA-rated corporate bond yields. "dy" the dividend yield, "ep" the earnings–price ratio, "infl" the inflation rate, "ltr" the long-term U.S. government bond returns, "lty" the long-term U.S. government bond yields, and "svar" the stock variance. "tms" is the term spread as the difference between the long-term yield on U.S. government bonds and the 3-month Treasury bill rate. "unrate" is the unemployment rate. "Mean", "Std Dev", "Skew", and "Kurt" denote the mean, standard deviation, skewness, and kurtosis, respectively. The next three columns show the first-order autoregressive coefficient and the p-value of the Jarque-Bera and Augmented Dicky Fuller test, respectively. "First Obs." and "Nobs" denote the first observation of the time series and the number of observations. All data are sampled at the monthly frequency.

Predictor	Mean	Std Dev	Skew	Kurt	AR(1)	JB p-value	ADF p-value	First obs.	Nobs
de	-0.8024	0.4042	2.7791	13.6839	0.9847	<0.01	<0.01	31.08.1988	353
ΔIndpro	0.0015	0.0063	-1.6432	12.3151	0.2256	<0.01	<0.01	31.08.1988	353
dfr	0.0000	0.0157	-0.4845	11.3621	0.0275	<0.01	<0.01	31.08.1988	353
dfy	0.0095	0.0039	3.2061	17.1986	0.9638	<0.01	<0.01	31.08.1988	353
dy	-3.8813	0.2924	0.1077	2.4746	0.9857	<0.1	0.66	31.08.1988	353
ep	-3.0854	0.3696	-1.9176	9.1496	0.9757	<0.01	<0.1	31.08.1988	353
infl	0.0021	0.0033	-0.9453	8.4420	0.4723	<0.01	<0.01	31.08.1988	353
ltr	0.0070	0.0285	0.0073	5.3642	0.0386	<0.01	<0.01	31.08.1988	353
lty	0.0534	0.0196	0.1278	2.1312	0.9887	<0.01	<0.01	31.08.1988	353
svar	0.0025	0.0045	7.3680	76.5853	0.7072	<0.01	<0.01	31.08.1988	353
tms	0.0234	0.0128	-0.1236	1.9339	0.9763	<0.01	0.48	31.08.1988	353
unrate	0.0596	0.0152	0.9967	3.1889	0.9969	<0.01	0.98	31.08.1988	353

Table 4
Correlation matrix predictive variables. This table reports the correlations among all predictive variables. "de" denotes the dividend–payout ratio, and "ΔIndpro" the growth of industrial production. "dfr" is the default return spread as the difference between long-term U.S. corporate bond returns and long-term U.S. government bond returns. "dfy" is the default yield spread as the difference between U.S. BAA- and AAA-rated corporate bond yields. "dy" is the dividend yield, "ep" the earnings–price ratio, "infl" the inflation rate, "ltr" the long-term U.S. government bond returns, "lty" the long-term U.S. government bond yields, and "svar" the stock variance. "tms" is the term spread as the difference between the long-term yield on U.S. government bonds and the 3-month Treasury bill rate. "unrate" is the unemployment rate. All data are sampled at the monthly frequency.

Predictor	de	ΔIndpro	dfr	dfy	dy	ep	infl	ltr	lty	svar	tms	unrate
de												
ΔIndpro	-0.24											
dfr	0.15	0.05										
dfy	0.61	-0.43	0.11									
dy	0.47	-0.11	0.05	0.24								
ep	-0.72	0.18	-0.17	-0.46	0.27							
infl	-0.06	0.03	-0.02	-0.20	0.06	0.11						
ltr	-0.01	-0.04	-0.50	0.01	0.04	0.06	-0.20					
lty	0.12	0.12	-0.03	-0.25	0.39	0.17	0.23	0.00				
svar	0.30	-0.24	-0.24	0.59	-0.01	-0.29	-0.31	0.15	-0.10			
tms	0.39	0.01	0.09	0.27	0.19	-0.28	-0.08	-0.11	-0.15	0.11		
unrate	0.30	0.01	0.09	0.37	0.39	-0.03	-0.06	0.01	-0.18	0.08	0.69	

French (1989) who argue that tms, dfy, and dy are appropriate predictors, because tms is related to short-term business cycles, whereas dfy and dy to long-term business cycles.

3.3. Model selection approach

Next, we examine the return predictability based on a model selection approach. In particular, we ask the question: Is it possible to improve the return predictability when aggregating the information of variables? The results are presented in Table 6. We do not only analyze the full sample, but also three further sub-samples. First, we analyze the post financialization time period, i.e., the time after December 2000.[9] Second, we analyze expansions and recessions separately. According to Gorton and Rouwenhorst (2006) and Gorton et al. (2013), commodities behave differently over the business cycle, and commodity futures returns shall be better predictable during recessions.[10]

In-sample results. To select specific variables for predicting the next year's futures excess return, we proceed as follows. First, we run a kitchen sink regression by including all variables. In doing so, we are able to identify the variables that have significant predictive power for future excess returns, at least at the 10% significance level. Afterwards, employed with the significant variables, we run a multiple predictive regression and extract the adjusted in-sample R^2, i.e., $\bar{R}^2$. To determine the corresponding significance, we use an F-test.

Table 6 shows a solid predictability across all metal futures over all time periods. Analyzing the full sample, we find $\bar{R}^2 s$ ranging from 7.46% for silver to 19.51% for platinum. By analyzing the post finan-

[9] Here, the release of the Commodity Futures Modernization Act (CFMA) took place.

[10] Due to limited data availability, we cannot provide out-of-sample results for recessions.

Table 5
Return predictability: univariate regressions. This table reports the regression results of monthly excess returns [name in row] on a constant and the lagged predictive variable [name in column]. We predict the next year's excess return. Statistical inferences are based on a bootstrapped distribution following Rapach and Wohar (2006). "de" denotes the dividend–payout ratio, "ΔIndpro" the growth of industrial production. "dfr" is the default return spread as the difference between long-term U.S. corporate bond returns and long-term U.S. government bond returns. "dfy" is the default yield spread as the difference between U.S. BAA- and AAA-rated corporate bond yields. "dy" the dividend yield, "ep" the earnings–price ratio, "infl" the inflation rate, "ltr" the long-term U.S. government bond returns, "lty" the long-term U.S. government bond yields, and "svar" the stock variance. "tms" is the term spread as the difference between the long-term yield on U.S. government bonds and the 3-month Treasury bill rate. "unrate" is the unemployment rate. R^2 and R^2_{oos} are the in-sample and out-of-sample R^2, respectively. We report the t-statistics in parentheses. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

Commodity	Statistic	de	ΔIndpro	dfr	dfy	dy	ep	infl	ltr	lty	svar	tms	unrate
Copper	R^2	**6.23*****	0.31	**1.00***	**9.21*****	0.01	**6.68*****	**1.49****	0.10	**1.18***	**3.06*****	**3.54*****	**0.86***
	R^2_{oos}	**2.47*****	-1.87	-0.76	**5.87*****	-7.68	**0.86****	0.11	-1.78	**10.99*****	-9.30	**5.06*****	-6.62
	t-stat	(4.75)	(-1.03)	(1.85)	(5.86)	(0.21)	(-4.93)	(-2.26)	(-0.59)	(-2.01)	(3.27)	(3.53)	(1.72)
Gold	R^2	**3.07*****	**1.50****	0.12	**9.59*****	**1.88*****	**8.42*****	0.62	0.00	**2.86*****	**4.71*****	**2.69*****	**1.20****
	R^2_{oos}	-2.26	**0.88****	-0.94	**10.99*****	-3.97	**0.79****	-1.03	-1.12	**37.90*****	-5.12	-2.09	-1.33
	t-stat	(3.28)	(-2.27)	(0.63)	(6.00)	(-2.55)	(-5.58)	(-1.45)	(0.10)	(-3.16)	(4.09)	(3.06)	(2.03)
Palladium	R^2	**5.92*****	**1.47****	0.65	**2.47****	**5.63*****	**2.44*****	**1.55****	0.20	0.00	0.66	0.03	**2.51*****
	R^2_{oos}	**2.82*****	-14.89	-3.76	-9.61	**19.58*****	-5.88	**0.76***	-3.47	-20.84	-22.74	-1.35	**2.69*****
	t-stat	(4.05)	(1.97)	(1.31)	(2.57)	(3.94)	(-2.55)	(-2.03)	(-0.72)	(0.09)	(1.32)	(-0.29)	(2.59)
Platinum	R^2	**5.55*****	0.02	0.22	**6.48*****	**3.52*****	**16.33*****	**3.05*****	0.08	0.04	**5.37*****	**1.38****	0.01
	R^2_{oos}	**1.16****	-2.83	-1.96	**5.10*****	**1.15*****	**13.75*****	**1.16****	-1.21	**11.02*****	**0.34***	-2.80	-6.35
	t-stat	(4.46)	(0.28)	(0.86)	(4.85)	(-3.52)	(-8.13)	(-3.26)	(-0.51)	(-0.36)	(4.39)	(2.17)	(0.21)
Silver	R^2	**3.17*****	0.06	0.00	**4.85*****	0.16	**4.98*****	**2.41*****	0.47	0.71	**3.53*****	**5.60*****	**3.35*****
	R^2_{oos}	**2.59*****	-0.81	-1.72	**4.79*****	-1.22	-1.15	0.13	-1.15	**14.29*****	-4.03	**3.45*****	**3.19*****
	t-stat	(3.33)	(-0.45)	(-0.09)	(4.16)	(-0.73)	(-4.21)	(-2.89)	(1.27)	(-1.56)	(3.52)	(4.48)	(3.43)

Table 6
Return predictability: model selection approach. This table reports the regression results of monthly excess returns [name in row] on a constant and the lagged predictive variable(s) based on a model selection approach. We predict the next year's excess return. Statistical inferences are based on an F-test (in-sample), and on the MSFE-adjusted test statistic with robust Newey and West (1987) standard errors (12 lags) following Clark and West (2007) (out-of-sample). For the in-sample analysis, we run a multiple predictive regression containing all significant variables (at least at the 10% significance level) determined by a prior kitchen sink regression. For the out-of-sample analysis, we first run a kitchen sink regression to determine the significant variables (at least at the 10% significance level). Subsequently, we use a mean forecast combination approach using all significant variables. $\bar{R}^2$ and R^2_{oos} are the in-sample adjusted and out-of-sample R^2, respectively. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

Commodity	Statistic	Full sample	Post financialization	Expansion	Recession
Copper	$\bar{R}^2$	**11.05*****	**2.34****	**42.30*****	**79.63*****
	R^2_{oos}	3.13	-5.06	12.56	
Gold	$\bar{R}^2$	**9.33*****	**28.31*****	**49.09*****	**44.96*****
	R^2_{oos}	**14.38****	**18.57***	**18.00****	
Palladium	$\bar{R}^2$	**17.52*****	**19.05*****	**28.79*****	**0.00*****
	R^2_{oos}	14.26	-13.92	-12.20	
Platinum	$\bar{R}^2$	**19.51*****	**26.46*****	**34.53*****	**64.62*****
	R^2_{oos}	**2.31***	**3.30***	3.81	
Silver	$\bar{R}^2$	**7.46*****	**12.49*****	**40.97*****	**40.47*****
	R^2_{oos}	14.17	-1.06	**14.77****	

cialization period, we notice an improvement in the predictability for all metals, except for copper. The $\bar{R}^2 s$ span from 12.49% for silver to 28.31% for gold. When differentiating between expansions and recessions, we observe a substantial improvement in return predictability, except for palladium in recessions. In particular, we find a superior performance for copper and platinum in recessions, indicated by $\bar{R}^2 s$ of 79.63% and 64.62%, respectively.

Out-of-sample results. Analogously to the in-sample analysis, we proceed out-of-sample similarly. In the first step, we run a kitchen sink regression to determine the variables that are significant at least the 10% significance level. Afterwards, we use a mean forecast combination approach to determine the out-of-sample forecast and the corresponding R^2_{oos}.

Following Rapach et al. (2010), forecast combinations might lead to an improvement of the out-of-sample predictability. On the one hand, forecast combinations aggregate information of multiple variables, and thus, providing more stable out-of-sample forecasts by reducing the forecast volatility. On the other hand, forecast combinations incorporate information about the state of the real economy. The authors argue that mean forecast combinations provide evidence for a superior performance, despite its simplicity.[11] In doing so, we compute the combined out-of-sample forecast as:

[11] Further alternatives are the median and trimmed mean forecast combination approaches. We obtain similar results, when using these approaches rather than the mean forecast combination approach.

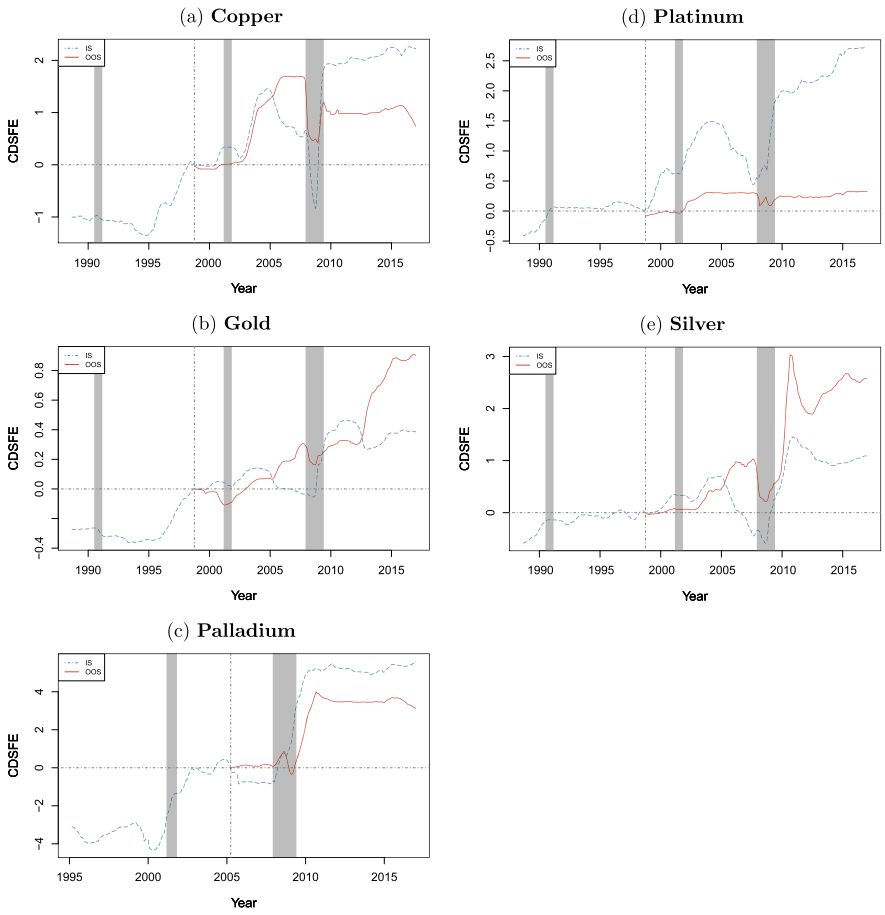

Fig. 1. Return predictability (model selection approach). This figure plots the in- and out-of-sample performances predicting the next year's excess return based on a model selection approach. For the in-sample analysis, we run a multiple predictive regression containing all significant variables (at least at the 10% significance level) determined by a prior kitchen sink regression. For the out-of-sample analysis, we first run a kitchen sink regression to determine the significant variables (at least at the 10% significance level). Subsequently, we use a mean forecast combination approach using all significant variables. On the ordinate, there are the cumulative differences in squared forecast errors (CDSFE). The in-sample performance is the difference between the cumulative squared demeaned excess return from the restricted model and the cumulative squared regression residual from the unrestricted model, whereas the out-of-sample performance is the difference between the cumulative squared forecast error from the restricted model and the cumulative squared forecast error of the unrestricted model. The grey bars indicate the U.S. recessions, published by the NBER. The sample period spans from August 1988 to December 2017. All data are sampled at the monthly frequency.

$$\widehat{er}^{c,oos}_{t+12} = \frac{1}{M}\sum_{m=1}^{M} \widehat{er}^{m,oos}_{t+12}, \tag{10}$$

where $\widehat{er}^{c,oos}_{t+12}$ is the combined out-of-sample forecast, and $\widehat{er}^{m,oos}_{t+12}$ the individual out-of-sample forecast of the (at the at least 10% level significant) predictor m, where $m = 1, ..., M$. To determine the significance of the R^2_{oos}, we use the MSFE-adjusted test statistic of Clark and West (2007).[12]

In Table 6, we find a substantial improvement of the out-of-sample predictability, in particular for gold, indicated by $R^2_{oos}s$ of 14.38% and 18.57% analyzing the full sample and the post financialization period, respectively. Moreover, platinum provides evidence for a strong return predictability, indicated by $R^2_{oos}s$ of 2.31% and 3.30%, respectively. In the case of silver, we observe a notable predictive power in expansions ($R^2_{oos} = 14.77\%$).

Overall, the model selection approaches provide evidence for an improved performance predicting the next year's futures excess return. The findings suggest that aggregated information might lead to a superior performance, especially in the case of gold and platinum.

[12] Further information about the MSFE-adjusted test statistic can be found in Section 2.5.

CDSFE. Fig. 1 shows the in- and out-of-sample CDSFE plots for each metal commodity based on the model selection approaches. Here, the dashed (blue) curve represents the in-sample performance, whereas the solid (red) curve the out-of-sample one. For all metal commodities, we observe both an increasing in- and out-of-sample performance over time, indicating the superior performance of the unrestricted model relative to the historical mean as naive benchmark. We also notice that during recessions a decline is observable, however, short-lasting only.

Overall, the CDSFE plots provide evidence for return predictability over the entire time period, on average. In particular in the time periods between crises, we observe years of high and stable predictability. This finding is a result of the aggregation of the information incorporated in different variables, justifying the application of the mean forecast combination approach. Accordingly, our findings do not confirm the previous results of Goyal and Welch (2008) who argue that return predictability is mainly driven by crises.

3.4. Economic value analysis

Next, we examine whether return predictability also translates to economic gains. The results are presented in Table 7 for different sample periods. In doing so, we assume an investor with mean–variance preferences who decides to allocate a fraction ω_t of her wealth to the

Table 7
Economic value. This table reports utility gains based on the mean forecast combination approach, assuming that the combined forecast predicts excess returns. The historical mean return serves as naive benchmark. ΔCER is the (annualized) utility gain relative to a naive strategy that assumes that excess returns are unpredictable. All data are sampled at the monthly frequency.

Panel A: $\gamma = 3$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.04	0.48	1.99	–
Gold	ΔCER	2.18	1.55	3.92	–
Palladium	ΔCER	1.42	-0.36	0.09	–
Platinum	ΔCER	1.00	0.09	1.75	–
Silver	ΔCER	0.65	0.54	4.41	–

Panel B: $\gamma = 6$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.65	0.23	0.90	–
Gold	ΔCER	1.93	2.98	1.99	–
Palladium	ΔCER	0.62	-0.16	0.03	–
Platinum	ΔCER	0.38	0.04	0.81	–
Silver	ΔCER	-4.97	0.12	2.04	–

Panel C: $\gamma = 9$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.53	0.15	0.58	–
Gold	ΔCER	0.78	2.29	1.26	–
Palladium	ΔCER	0.39	-0.10	0.02	–
Platinum	ΔCER	0.23	0.03	0.52	–
Silver	ΔCER	-3.78	0.07	1.27	–

Panel D: $\gamma = 12$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.43	0.11	0.43	–
Gold	ΔCER	0.04	1.64	0.92	–
Palladium	ΔCER	0.29	-0.07	0.01	–
Platinum	ΔCER	0.16	0.02	0.38	–
Silver	ΔCER	-3.01	0.05	0.91	–

risky portfolio and the remainder, i.e. $1 - \omega_t$, to the risk-free asset. The investor's objective function reads as follows:

$$\max_{\omega_t} E_t \left(R_{p,t+12} - \frac{\gamma}{2} \sigma^2_{p,t+12} \right), \tag{11}$$

where $E_t(\cdot)$ is the expectation operator, $\sigma^2_{p,t+12}$ the conditional variance of the portfolio from t to $t+12$, and γ is the coefficient of relative risk-aversion. $R_{p,t+12}$ is the next-period's (simple) return on the investor's portfolio. To address the fact that our analysis is based on log rather than simple returns, we use a second-order Taylor expansion to convert the returns.[13] Thus, we can express the objective function as follows:

$$\max_{\omega_t} E_t \left(r_{p,t+12} - \frac{\gamma - 1}{2} \sigma^2_{p,t+12} \right), \tag{12}$$

where $r_{p,t+12}$ is the log-return on the portfolio, and σ^2_{t+12} is estimated using a five-year rolling window.

Optimizing Equation (12), we obtain the optimal weight invested in the risky asset (Jordan et al., 2014):

$$\omega_t = \frac{E_t(er_{t+12} + \frac{1}{2}\sigma^2_{t+12})}{\gamma E_t(\sigma^2_{t+12})} = \frac{E_t(er_{t+12})}{\gamma E_t(\sigma^2_{t+12})} + \frac{1}{2\gamma}. \tag{13}$$

Equation (13) shows that the optimal weight depends on the expected futures excess returns, the coefficient of relative risk-aversion, and the expected variance.

[13] The second-order Taylor expansion leads to the following relationship: $r_t \approx R_t - \frac{1}{2}\sigma^2_t$, where r_t, R_t, and σ^2_t are the log-return, simple return, and variance at time t, respectively. Accordingly, we use the log-return and variance to express the simple return.

Table 8
Economic value and transaction costs. This table reports utility gains based on the mean forecast combination approach, assuming that the combined forecast predicts excess returns. We assume transaction costs of 50 basis points per transaction proportional to the asset's traded size. The historical mean return serves as naive benchmark. ΔCER is the (annualized) utility gain relative to a naive strategy that assumes that excess returns are unpredictable. All data are sampled at the monthly frequency.

Panel A: $\gamma = 3$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.08	0.48	2.00	–
Gold	ΔCER	2.14	1.50	3.82	–
Palladium	ΔCER	1.39	-0.37	0.09	–
Platinum	ΔCER	0.97	0.09	1.69	–
Silver	ΔCER	0.61	0.50	4.35	–

Panel B: $\gamma = 6$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.67	0.24	0.90	–
Gold	ΔCER	1.90	2.94	1.93	–
Palladium	ΔCER	0.61	-0.16	0.03	–
Platinum	ΔCER	0.37	0.05	0.79	–
Silver	ΔCER	-5.01	0.09	1.99	–

Panel C: $\gamma = 9$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.54	0.16	0.58	–
Gold	ΔCER	0.74	2.27	1.22	–
Palladium	ΔCER	0.39	-0.11	0.02	–
Platinum	ΔCER	0.22	0.03	0.50	–
Silver	ΔCER	-3.81	0.05	1.24	–

Panel D: $\gamma = 12$

Commodity	Statistic	Full sample	Post financial.	Expansion	Recession
Copper	ΔCER	-0.44	0.12	0.43	–
Gold	ΔCER	0.02	1.61	0.89	–
Palladium	ΔCER	0.28	-0.08	0.01	–
Platinum	ΔCER	0.15	0.02	0.37	–
Silver	ΔCER	-3.03	0.04	0.89	–

For each month in our out-of-sample analysis, we compute the weight ω_t and also the realized return of the portfolio. We impose the restriction that whenever the forecast of the market excess return in Equation (13) equals zero, we set the portfolio weight equal to $1/(2\gamma)$. Further, following Campbell and Thompson (2008) and Jordan et al. (2017), we impose the restriction that ω_t has to be between 0 and 1.5.[14] Finally, we compute the certainty equivalent return (CER) as:

$$CER = \bar{R}_p - \frac{\gamma}{2}\sigma^2_p, \tag{14}$$

where $\bar{R}_p$ is the average return on the portfolio, and σ^2_p is the variance of the portfolio returns. Further, we define the utility gain (ΔCER) as the difference between the CER of a strategy assuming that excess returns are predictable using X_t, and the CER of the benchmark strategy that assumes that returns are unpredictable.

Table 7 provides the results for different coefficients of relative risk-aversion. Assuming $\gamma = 3$, we find positive utility gains for all metals ranging from 0.65% p.a. for silver to 2.18% p.a. for gold. Copper represents an exception by showing a slightly negative ΔCER for the full sample. Examining the post financialization period, we find that all metals, except palladium, provide evidence for positive utility gains up to 1.55% p.a. for gold. Moreover, in expansions, all metal commodities exhibit positive utility gains up to 4.41% p.a. for silver.

Table 8 reports the utility gains taking transaction costs into account. Here, we follow Balduzzi and Lynch (1999) and assume trans-

[14] In doing so, we avoid short-selling and an extensive leverage of the risky asset.

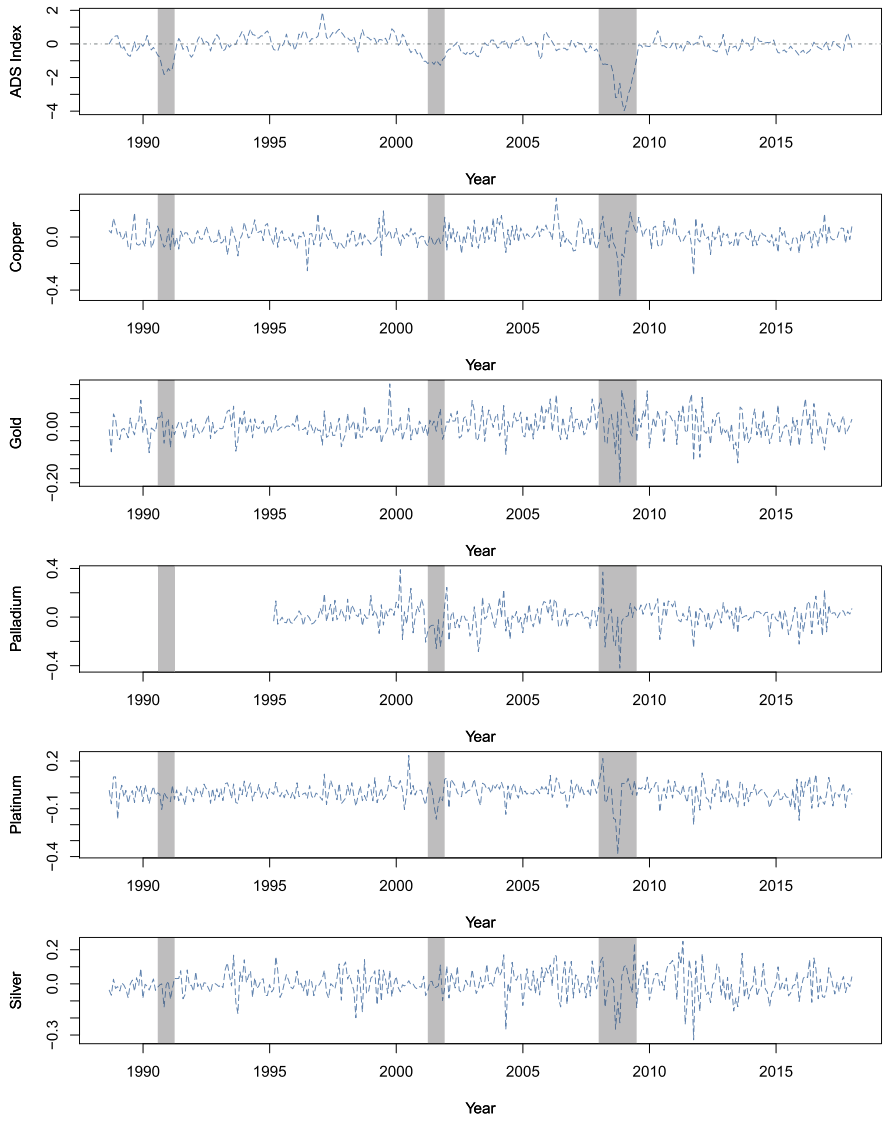

Fig. 2. Development of ADS index and metal futures excess returns. This figure plots the development of the ADS index and the development of the monthly metal futures excess returns over time. The grey bars indicate the U.S. recessions, published by the NBER. The sample period spans from August 1988 to December 2017. All data are sampled at the monthly frequency.

action costs of 50 basis points per transaction proportional to the asset's traded size $|\omega_{t+12} - \omega_{t^+}|$, where ω_{t^+} is the portfolio weight before re-balancing at $t + 12$. We observe that our main results are almost unchanged when taking transaction costs into account. E.g., an investor relaying on the aggregated information when predicting gold excess returns, would earn an utility gain of now 2.14% p.a. (rather than 2.18% p.a.).

Overall, the results provide evidence that return predictability also translates to economic gains. When relying on the aggregated information, investors might earn substantial utility gains. We notice that transaction costs do not systematically affect our results.

4. Results and discussion

In this section, we analyze the relationship between metal futures excess returns and the ADS index. In the first step, we introduce the ADS index, and in the second one we examine the informative value. Finally, we use multiple regressions to analyze the robustness of the predictive power of the ADS index.

4.1. ADS index

The ADS index is developed by Aruoba et al. (2009) and represents a measure of macroeconomic activity, and thus of business conditions in real-time. The importance to have aggregated business conditions in real-time arises from the fact that economic agents make decisions in real-time. This includes, among many others, policy makers, central banks, and investors.

Using a variety of information, the authors are able to track business conditions over time. In doing so, they use a dynamic factor model to deal with potential co-movements of business cycles with related variables. Moreover, they employ business conditions indicators, measured at low and high frequencies, to extract relevant information, e.g., about asset prices, the term premium, the payroll employment, initial jobless claims, and the GDP. The index is zero on average, thus, a positive (negative) value represents business conditions above (below) the average conditions.

4.2. Informative value of the ADS index for metal futures returns

Fig. 2 shows the monthly development of the ADS index and of the metal futures excess returns. We observe that particularly in crises, the

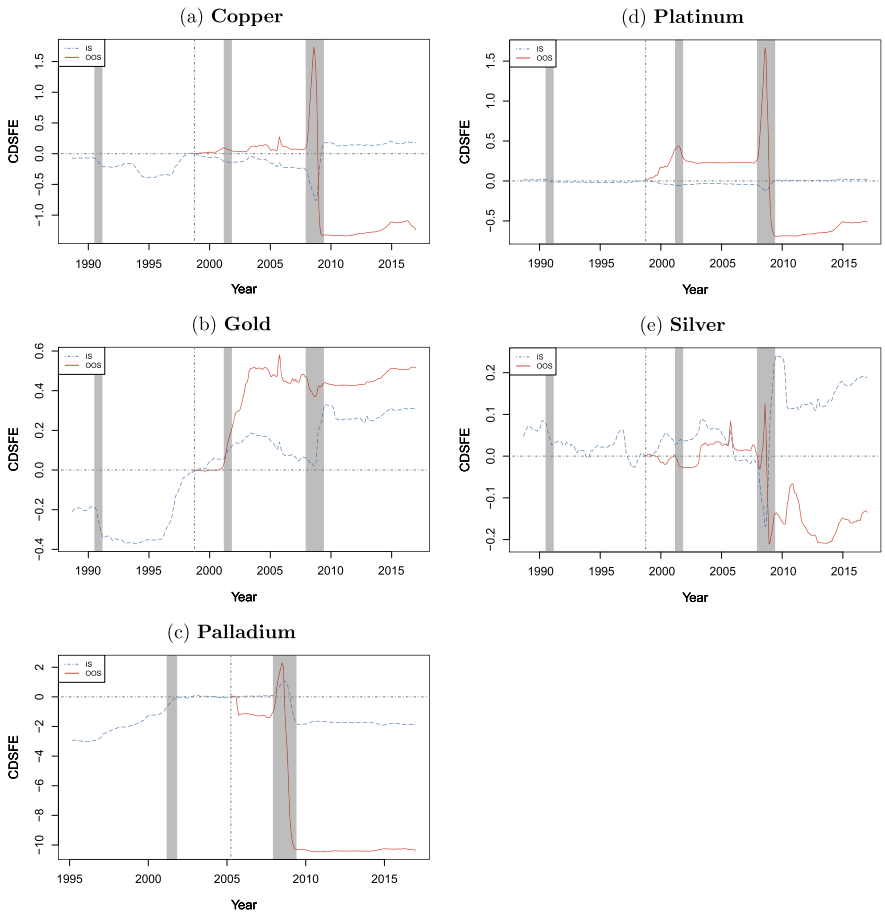

Fig. 3. Return predictability (ADS index). This figure plots the in- and out-of-sample performances predicting the next year's excess return based on the ADS index. On the ordinate, there are the cumulative differences in squared forecast errors (CDSFE). The in-sample performance is the difference between the cumulative squared demeaned excess return from the restricted model and the cumulative squared regression residual from the unrestricted model, whereas the out-of-sample performance is the difference between the cumulative squared forecast error from the restricted model and the cumulative squared forecast error of the unrestricted model. The grey bars indicate the U.S. recessions, published by the NBER. The sample period spans from August 1988 to December 2017. All data are sampled at the monthly frequency.

metal excess returns move into the same direction as the ADS index. To deepen the analysis, Panel A of Table 9 reports the correlations between the ADS index and the metal excess returns over different sample periods.

In the case of copper, platinum, and silver we observe small negative correlations with the ADS index in the range of −0.01 and −0.06 over the full sample, the post financialization period, and in expansions. Gold exhibits a strong negative correlation of −0.28 (full sample), −0.13 (post financialization), and −0.41 in recessions. In contrast, palladium shows a slight positive correlation over the post financialization period (0.06), but strong positive correlations over the full sample (0.21) and in expansions (0.54). Interestingly, all metal commodities have in common a strong negative correlation with the ADS index in recessions ranging from −0.45 for gold and platinum, to −0.60 for palladium. Thus, the results show that the ADS index has most explanatory power for metal future excess returns during recessions.

Panel B of Table 9 reports the in-sample and out-of-sample results predicting the next year's excess return based on the ADS index. In doing so, we use the same methodology as in our previous section, however, using the ADS index as predictive variable. In-sample, we observe significant R^2s in the case of copper ($R^2 = 0.89\%$), palladium ($R^2 = 2.50\%$), and gold ($R^2 = 7.69\%$). For platinum and silver, we find positive but insignificant R^2s. Thus, the ADS index seems to have in-sample predictive power for at least three of five metal commodities.

Table 9
Correlations and return predictability of ADS index. This table reports results related to the ADS index. Panel A shows the correlations between the ADS index and the annual metal futures excess returns over different time periods. Panel B shows the regression results of monthly excess returns [name in column] on a constant and the lagged ADS index. We predict the next year's excess return. Statistical inferences are based on a bootstrapped distribution following Rapach and Wohar (2006). R^2 and R^2_{oos} are the in-sample and out-of-sample R^2, respectively. We report the t-statistics in parentheses. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

Panel A: correlations

Time period	Copper	Gold	Palladium	Platinum	Silver
Full sample	-0.03	-0.28	0.21	0.02	-0.05
Post financial.	-0.01	-0.13	0.06	-0.05	-0.02
Expansion	-0.04	-0.41	0.54	-0.04	-0.06
Recession	-0.53	-0.45	-0.60	-0.45	-0.57

Panel B: return predictability (full sample)

Statistic	Copper	Gold	Palladium	Platinum	Silver
R^2	**0.89***	**7.69*****	**2.50****	0.06	0.72
R^2_{oos}	-5.26	**8.21*****	-47.42	-3.68	-0.74
t-stat	(-1.74)	(-5.31)	(2.59)	(-0.47)	(-1.57)

Analyzing the return predictability out-of-sample, we find that the ADS index has strong predictive power for gold excess returns, indicated

Table 10
Multiple regressions: stock variables, macroeconomic variables, and ADS index. This table reports the regression results of monthly excess returns [name in column] on a constant and the lagged predictive variables [name in rows]. We predict the next year's excess return. Statistical inferences are based on a bootstrapped distribution following Rapach and Wohar (2006). "de" denotes the dividend–payout ratio, "ΔIndpro" the growth of industrial production. "dfr" is the default return spread as the difference between long-term U.S. corporate bond returns and long-term U.S. government bond returns. "dfy" is the default yield spread as the difference between U.S. BAA- and AAA-rated corporate bond yields. "dy" the dividend yield, "ep" the earnings–price ratio, "infl" the inflation rate, "ltr" the long-term U.S. government bond returns, "lty" the long-term U.S. government bond yields, and "svar" the stock variance. "tms" is the term spread as the difference between the long-term yield on U.S. government bonds and the 3-month Treasury bill rate. "unrate" is the unemployment rate. R^2 and R^2_{oos} are the in-sample and out-of-sample R^2, respectively. We report the t-statistics in parentheses. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

	Panel A: with ADS index					Panel B: without ADS index				
	Copper	Gold	Palladium	Platinum	Silver	Copper	Gold	Palladium	Platinum	Silver
Constant	-0.26	**-0.70*****	0.84	**-1.61*****	**-1.01*****	-0.26	**-0.70*****	0.73	**-1.61*****	**-1.01*****
	(-0.69)	(-3.90)	(1.44)	(-6.37)	(-3.15)	(-0.69)	(-3.86)	(1.17)	(-6.23)	(-3.15)
de	0.42	-0.02	-0.27	-0.17	-0.17	0.17	0.08	**-1.23***	-0.37	-0.25
	(1.01)	(-0.08)	(-0.43)	(-0.59)	(-0.49)	(0.42)	(0.38)	(-1.90)	(-1.31)	(-0.72)
ΔIndpro	-3.44	2.51	-6.60	-1.56	-0.67	3.98	-0.26	**13.78*****	**4.60****	1.66
	(-0.98)	(1.47)	(-1.31)	(-0.65)	(-0.22)	(1.47)	(-0.20)	(3.41)	(2.47)	(0.72)
dfr	0.94	0.06	-2.14	-1.07	-0.53	0.88	0.09	-2.58	-1.12	-0.55
	(0.76)	(0.10)	(-1.29)	(-1.26)	(-0.50)	(0.70)	(0.14)	(-1.45)	(-1.29)	(-0.52)
dfy	**32.42*****	**7.84****	**27.92****	**26.60*****	**14.84****	**22.17*****	**11.67*****	2.14	**18.09*****	**11.62***
	(4.11)	(2.03)	(2.41)	(4.93)	(2.17)	(3.02)	(3.26)	(0.19)	(3.57)	(1.85)
dy	-0.39	-0.12	0.79	-0.04	0.03	-0.17	-0.20	**1.68*****	0.14	0.10
	(-0.96)	(-0.60)	(1.31)	(-0.15)	(0.10)	(-0.41)	(-1.02)	(2.66)	(0.51)	(0.30)
ep	0.34	-0.03	-0.59	-0.30	-0.22	0.11	0.06	**-1.52****	**-0.49***	-0.29
	(0.85)	(-0.15)	(-0.96)	(-1.07)	(-0.61)	(0.27)	(0.29)	(-2.37)	(-1.75)	(-0.83)
infl	-2.70	-0.28	-10.55	**-7.26****	-5.88	-4.56	0.42	**-14.19***	**-8.81****	-6.46
	(-0.54)	(-0.12)	(-1.54)	(-2.14)	(-1.37)	(-0.91)	(0.17)	(-1.95)	(-2.56)	(-1.51)
ltr	0.28	0.09	**-1.47***	-0.28	0.56	0.09	0.16	**-1.65***	-0.43	0.50
	(0.46)	(0.28)	(-1.74)	(-0.66)	(1.04)	(0.15)	(0.51)	(-1.83)	(-1.01)	(0.94)
lty	-0.22	0.52	1.51	**2.95*****	**1.55***	-0.19	0.52	1.64	**2.97*****	**1.56***
	(-0.20)	(0.99)	(0.78)	(4.00)	(1.66)	(-0.18)	(0.97)	(0.80)	(3.94)	(1.66)
svar	-0.15	-0.94	8.20	1.52	2.24	-2.08	-0.22	2.83	-0.09	1.63
	(-0.03)	(-0.40)	(1.27)	(0.47)	(0.54)	(-0.44)	(-0.09)	(0.41)	(-0.03)	(0.40)
tms	**3.95****	1.04	**-9.62*****	0.63	**3.02****	**3.95****	1.03	**-10.22*****	0.63	**3.02****
	(2.28)	(1.22)	(-3.91)	(0.53)	(2.01)	(2.25)	(1.21)	(-3.89)	(0.52)	(2.01)
unrate	**-3.30****	0.65	**4.88****	0.26	1.44	**-2.86***	0.48	**6.42*****	0.63	1.58
	(-2.03)	(0.81)	(2.24)	(0.23)	(1.02)	(-1.74)	(0.60)	(2.77)	(0.55)	(1.12)
ADS index	**0.13*****	**-0.05****	**0.38*****	**0.11*****	0.04					
	(3.27)	(-2.50)	(6.10)	(3.97)	(1.18)					
R^2	**13.01*****	**14.55*****	**27.93*****	**26.01*****	**9.76*****	**10.43*****	**13.18*****	**17.47*****	**22.67*****	**9.65*****
R^2_{oos}	-18.38	**39.17*****	-100.44	-8.50***	**13.10*****	-24.13	**37.91*****	-88.80	-15.89	**9.21*****

by an R^2_{oos} of 8.21%. The results suggest that the ADS index is a reliable predictor, at least for gold excess returns.

Fig. 3 plots the in- and out-of-sample performances related to the ADS index. In particular in the case of gold, we observe a superior predictive power, indicated by increasing CDSFE curves. Moreover, in the case of copper, palladium, platinum, and silver we notice a sharp decline during the global financial crisis, indicating an inferior performance of the unrestricted relative to the restricted model.

4.3. Multiple regressions

For robustness, to assess the predictive power of the ADS index in the case of multiple predictive regressions, we use Equation (3), however, we now define X_t as the vector of predictors. In doing so, we are able to analyze whether the ADS index still has significant predictive power for future excess returns or whether other predictive variables are capturing its predictive power. Thus, we proceed as follows: we run the multiple regressions once with and once without the ADS index. First, we use stock and macroeconomic variables jointly. Afterwards, we consider both kinds of predictive variables separately.

Table 10 reports the results of multiple regressions using stock and macroeconomic variables jointly. Panel A presents the results when using the ADS index as an additional variable, whereas Panel B reports the results of the regressions without ADS index. The results confirm the strong predictive power of the ADS index. The R^2s and the $R^2_{oos}s$ in Panel A are larger than in Panel B. The multiple regressions yield significant R^2s for all metal futures ranging from 9.76% for silver to 27.93% for palladium. In the out-of-sample analysis, we detect significant predictive power for gold and silver excess returns, indicated by $R^2_{oos}s$ of 39.17% and 13.10%. Moreover, the strong predictive power of the ADS index is also supported by its statistically significant t-statistics for all metals, except silver. Also dfy reveals strong predictive power for future excess returns, indicated by statistically significant t-statistics for all metals, which is supporting our previous findings of the univariate regressions (see Table 5).

The results in Table 11 confirm the findings, when using stock variables only. The R^2s are statistically significant for all metals. The out-of-sample results show significant predictive power for gold, silver, and platinum excess returns, indicated by $R^2_{oos}s$ of 36.89%, 16.61%, and 4.15%. Table 12 provides somewhat weaker evidence, when using macroeconomic variables. Overall, the analysis reveals significant predictive power in-sample for gold, palladium, and silver excess returns, whereas out-of-sample for gold and silver excess returns.

In total, the multiple regressions provide evidence for a strong predictive power of the ADS index for future metal excess returns, both in-sample and out-of-sample. To emphasize is the predictive power in particular for gold and silver excess returns. The results support our main findings.

5. Conclusion

This paper performs a comprehensive study of metal futures excess return predictability using 12 variables that are supposed to predict

Table 11
Multiple regressions: stock variables and ADS index. This table reports the regression results of monthly excess returns [name in column] on a constant and the lagged predictive variables [name in rows]. We predict the next year's excess return. Statistical inferences are based on a bootstrapped distribution following Rapach and Wohar (2006). "de" denotes the dividend–payout ratio. "dfr" is the default return spread as the difference between long-term U.S. corporate bond returns and long-term U.S. government bond returns. "dfy" is the default yield spread as the difference between U.S. BAA- and AAA-rated corporate bond yields. "dy" the dividend yield, "ep" the earnings–price ratio, "infl" the inflation rate, "ltr" the long-term U.S. government bond returns, "lty" the long-term U.S. government bond yields, and "svar" the stock variance. "tms" is the term spread as the difference between the long-term yield on U.S. government bonds and the 3-month Treasury bill rate. R^2 and R^2_{oos} are the in-sample and out-of-sample R^2, respectively. We report the t-statistics in parentheses. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

	Panel A: with ADS index					Panel B: without ADS index				
	Copper	Gold	Palladium	Platinum	Silver	Copper	Gold	Palladium	Platinum	Silver
Constant	**-0.65****	**-0.63*****	**1.43*****	**-1.58*****	**-0.84*****	**-0.58***	**-0.65*****	**1.75*****	**-1.51*****	**-0.81*****
	(-2.04)	(-4.05)	(2.77)	(-7.32)	(-3.07)	(-1.81)	(-4.17)	(3.05)	(-6.79)	(-2.96)
de	0.38	0.03	-0.52	-0.21	-0.21	0.28	0.06	**-1.13***	-0.32	-0.26
	(0.94)	(0.17)	(-0.84)	(-0.75)	(-0.60)	(0.68)	(0.32)	(-1.68)	(-1.11)	(-0.73)
dfr	0.65	0.19	-2.22	-1.12	-0.49	0.97	0.10	-1.26	-0.78	-0.35
	(0.53)	(0.32)	(-1.34)	(-1.33)	(-0.46)	(0.78)	(0.16)	(-0.69)	(-0.91)	(-0.33)
dfy	**28.46*****	**8.81****	**33.60*****	**26.72*****	**16.33****	**17.04****	**12.20*****	-4.64	**14.71*****	**11.27***
	(3.68)	(2.33)	(2.97)	(5.09)	(2.45)	(2.48)	(3.66)	(-0.41)	(3.09)	(1.92)
dy	-0.41	-0.16	**1.11***	0.00	0.09	-0.32	-0.18	**1.69*****	0.09	0.13
	(-1.02)	(-0.81)	(1.88)	(-0.00)	(0.26)	(-0.79)	(-0.95)	(2.61)	(0.34)	(0.37)
ep	0.29	0.02	-0.80	-0.33	-0.24	0.20	0.05	**-1.35****	-0.43	-0.28
	(0.72)	(0.11)	(-1.32)	(-1.23)	(-0.69)	(0.49)	(0.25)	(-2.03)	(-1.53)	(-0.81)
infl	-3.83	0.04	-9.32	**-7.27****	-5.50	-5.54	0.55	**-13.63***	**-9.07*****	-6.26
	(-0.77)	(0.02)	(-1.36)	(-2.16)	(-1.29)	(-1.11)	(0.23)	(-1.80)	(-2.62)	(-1.47)
ltr	0.05	0.16	-1.28	-0.29	0.62	-0.02	0.18	-1.12	-0.36	0.60
	(0.08)	(0.53)	(-1.52)	(-0.69)	(1.19)	(-0.03)	(0.60)	(-1.20)	(-0.84)	(1.13)
lty	0.36	0.43	0.76	**2.88*****	1.27	0.39	0.42	1.34	**2.91*****	1.28
	(0.35)	(0.85)	(0.40)	(4.10)	(1.42)	(0.37)	(0.83)	(0.63)	(4.00)	(1.43)
svar	-0.37	-0.63	6.73	1.27	2.02	-1.51	-0.29	2.35	0.07	1.52
	(-0.08)	(-0.27)	(1.04)	(0.40)	(0.49)	(-0.32)	(-0.13)	(0.33)	(0.02)	(0.37)
tms	1.57	**1.54****	**-6.35*****	0.77	**4.01*****	**2.18***	**1.36****	**-4.40****	1.41	**4.28*****
	(1.19)	(2.40)	(-3.29)	(0.86)	(3.53)	(1.65)	(2.14)	(-2.08)	(1.55)	(3.80)
ADS index	**0.10*****	**-0.03***	**0.35*****	**0.10*****	0.04					
	(3.10)	(-1.88)	(7.63)	(4.80)	(1.59)					
R^2	**12.08*****	**14.27*****	**26.70*****	**26.36*****	**10.02*****	**9.79*****	**13.61*****	**10.06*****	**21.45*****	**9.60*****
R^2_{oos}	-5.69***	**36.89*****	-78.20	**4.15*****	**16.61*****	-7.34**	**36.46*****	-81.30	**2.40*****	**16.11*****

Table 12
Multiple regressions: macroeconomic variables and ADS index. This table reports the regression results of monthly excess returns [name in column] on a constant and the lagged predictive variables [name in rows]. We predict the next year's excess return. Statistical inferences are based on a bootstrapped distribution following Rapach and Wohar (2006). "ΔIndpro" denotes the growth of industrial production. "unrate" is the unemployment rate. R^2 and R^2_{oos} are the in-sample and out-of-sample R^2, respectively. We report the t-statistics in parentheses. *, **, *** indicate the significance at the 10%, 5%, and 1% significance levels, respectively. All data are sampled at the monthly frequency.

	Panel A: with ADS index					Panel B: without ADS index				
	Copper	Gold	Palladium	Platinum	Silver	Copper	Gold	Palladium	Platinum	Silver
Constant	-0.07	-0.02	**-0.16***	0.01	**-0.14*****	-0.07	-0.02	**-0.16***	0.01	**-0.14*****
	(-1.09)	(-0.54)	(-1.78)	(0.24)	(-2.57)	(-1.09)	(-0.57)	(-1.72)	(0.23)	(-2.58)
ΔIndpro	0.45	**3.37****	-1.48	2.33	1.73	-2.61	**-2.78****	**7.34****	0.50	-1.09
	(0.13)	(1.99)	(-0.27)	(0.88)	(0.58)	(-1.07)	(-2.33)	(1.98)	(0.27)	(-0.53)
unrate	1.58	0.65	**4.50*****	0.04	**2.77*****	**1.77***	**1.04****	**3.87*****	0.16	**2.94*****
	(1.53)	(1.33)	(2.98)	(0.05)	(3.19)	(1.74)	(2.09)	(2.59)	(0.21)	(3.44)
ADS index	-0.04	**-0.08*****	**0.11****	-0.02	-0.03					
	(-1.18)	(-4.99)	(2.22)	(-0.95)	(-1.29)					
R^2	0.73	**8.65*****	**4.66*****	-0.59	**3.05*****	0.61	**2.18*****	**3.22*****	-0.56	**2.86*****
R^2_{oos}	-13.08	**7.99*****	-44.19	-11.36	**1.15*****	-8.13	-0.52**	-12.33	-9.37	**2.41*****

stock returns. We also focus on the identification of years of high and low predictability. We find a substantial degree of predictability both in- and out-of-sample. Mean forecast combinations provide evidence for an improved out-of-sample predictability. Gold returns appear to be best predictable. A timing strategy leads to utility gains of 2.18% p.a.

Moreover, we analyze the ADS index, which captures business conditions in real-time, to examine the potential effects on metal returns and on the behavior over business cycles. We provide evidence that the ADS index incorporates relevant information for metal returns. It turns out to be a strong predictor for gold returns.

Declarations

Author contribution statement

Björn Tharann: Conceived and designed the analysis; Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? Rev. Financ. Stud. 20 (3), 651–707.

Aruoba, S.B., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. J. Bus. Econ. Stat. 27 (4), 417–427.

Bailey, W., Chan, K.C., 1993. Macroeconomic influences and the variability of the commodity futures basis. J. Finance 48 (2), 555–573.

Bakshi, G., Gao, X., Rossi, A.G., 2017. Understanding the sources of risk underlying the cross section of commodity returns. Manag. Sci. 65 (2), 619–641.

Balduzzi, P., Lynch, A.W., 1999. Transaction costs and predictability: some utility cost calculations. J. Financ. Econ. 52 (1), 47–78.

Belousova, J., Dorfleitner, G., 2012. On the diversification benefits of commodities from the perspective of euro investors. J. Bank. Finance 36 (9), 2455–2472.

Bessembinder, H., 1992. Systematic risk, hedging pressure, and risk premiums in futures markets. Rev. Financ. Stud. 5 (4), 637–667.

Bessler, W., Wolff, D., 2015. Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies. J. Bank. Finance 60, 1–20.

Bjornson, B., Carter, C.A., 1997. New evidence on agricultural commodity return performance under time-varying risk. Am. J. Agric. Econ. 79 (3), 918–930.

Çakmaklı, C., van Dijk, D., 2016. Getting the most out of macroeconomic information for predicting excess stock returns. Int. J. Forecast. 32 (3), 650–668.

Campbell, J.Y., 1987. Stock returns and the term structure. J. Financ. Econ. 18 (2), 373–399.

Campbell, J.Y., Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. J. Finance 43 (3), 661–676.

Campbell, J.Y., Shiller, R.J., 1998. Valuation ratios and the long-run stock market outlook. J. Portf. Manag. 24 (2), 11–26.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. 21 (4), 1509–1531.

Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econom. 138 (1), 291–311.

Cochrane, J.H., 2008. The dog that did not bark: a defense of return predictability. Rev. Financ. Stud. 21 (4), 1533–1575.

De Roon, F.A., Nijman, T.E., Veld, C., 2000. Hedging pressure effects in futures markets. J. Finance 55 (3), 1437–1456.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.

Dow, C.H., 1920. Scientific Stock Speculation. Magazine of Wall Street.

Erb, C.B., Harvey, C.R., 2006. The strategic and tactical value of commodity futures. Financ. Anal. J. 62 (2), 69–97.

Fama, E.F., French, K.R., 1988a. Business cycles and the behavior of metals prices. J. Finance 43 (5), 1075–1093.

Fama, E.F., French, K.R., 1988b. Dividend yields and expected stock returns. J. Financ. Econ. 22 (1), 3–25.

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. J. Financ. Econ. 25 (1), 23–49.

Gargano, A., Timmermann, A., 2014. Forecasting commodity price indexes using macroeconomic and financial predictors. Int. J. Forecast. 30 (3), 825–843.

Gorton, G.B., Hayashi, F., Rouwenhorst, K.G., 2013. The fundamentals of commodity futures returns. Rev. Finance 17 (1), 35–105.

Gorton, G.B., Rouwenhorst, K.G., 2006. Facts and fantasies about commodity futures. Financ. Anal. J. 62 (2), 47–68.

Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Manag. Sci. 49 (5), 639–654.

Goyal, A., Welch, I., 2008. A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. 21 (4), 1455–1508.

Hammoudeh, S., Malik, F., McAleer, M., 2011. Risk management of precious metals. Q. Rev. Econ. Finance 51 (4), 435–441.

Hammoudeh, S., Santos, P.A., Al-Hassan, A., 2013. Downside risk management and VaR-based optimal portfolios for precious metals, oil and stocks. N. Am. J. Econ. Finance 25, 318–334.

Hollstein, F., Prokopczuk, M., Tharann, B., 2019. The economic sources of return anomalies: evidence from commodity futures markets. Working Paper.

Jordan, S.J., Vivian, A., Wohar, M.E., 2017. Forecasting market returns: bagging or combining? Int. J. Forecast. 33 (1), 102–120.

Jordan, S.J., Vivian, A., Wohar, M.E., 2018. Stock returns forecasting with metals: sentiment vs. fundamentals. Eur. J. Finance 24 (6), 458–477.

Jordan, S.J., Vivian, A.J., Wohar, M.E., 2014. Forecasting returns: new European evidence. J. Empir. Finance 26, 76–95.

Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. J. Finance 56 (3), 815–849.

McCracken, M.W., 2007. Asymptotics for out of sample tests of Granger causality. J. Econom. 140 (2), 719–752.

Newey, W.K., West, K.D., 1987. Hypothesis testing with efficient method of moments estimation. Int. Econ. Rev. 28 (3), 777–787.

Nguyen, D.B.B., Prokopczuk, M., Wese Simen, C., 2019. The risk premium of gold. J. Int. Money Financ. Forthcoming.

Prokopczuk, M., Tharann, B., Wese Simen, C., 2018. Predictability in commodity markets: evidence from more than a century. Working Paper.

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23 (2), 821–862.

Rapach, D.E., Wohar, M.E., 2006. In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. J. Empir. Finance 13 (2), 231–247.

Rapach, D.E., Zhou, G., 2013. Forecasting stock returns. In: Handbook of Economic Forecasting, vol. 2 (Part A), pp. 328–383.

Richardson, M., Stock, J.H., 1989. Drawing inferences from statistics based on multiyear asset returns. J. Financ. Econ. 25 (2), 323–348.

Rozeff, M.S., 1984. Dividend yields are equity risk premiums. J. Portf. Manag. 11 (1), 68–75.

Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. J. Am. Stat. Assoc. 83 (403), 842–848.

Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54 (3), 375–421.

Symeonidis, L., Prokopczuk, M., Brooks, C., Lazar, E., 2012. Futures basis, inventory and commodity price volatility: an empirical analysis. Econ. Model. 29 (6), 2651–2663.

West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica 64 (5), 1067–1084.

Zhang, Z., Zhang, H.-K., 2016. The dynamics of precious metal markets VaR: a GARCHEVT approach. J. Commod. Mark. 4 (1), 14–27.

Zhu, X., Zhu, J., 2013. Predicting stock returns: a regime-switching combination approach and economic links. J. Bank. Finance 37 (11), 4120–4133.