IHS Working Paper 28
January 2021

Regime-dependent commodity price dynamics: A predictive analysis

Jesús Crespo-Cuaresma
Ines Fortin
Jaroslava Hlouskova
Michael Obersteiner

INSTITUTE FOR
ADVANCED STUDIES
VIENNA

Author(s)

Jesús Crespo-Cuaresma, Ines Fortin, Jaroslava Hlouskova, Michael Obersteiner

Editor(s)

Robert M. Kunst

Title

Regime-dependent commodity price dynamics: A predictive analysis

Funder(s)

FWF Project, 30915-G27

Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
T +43 1 59991-0
www.ihs.ac.at
ZVR: 066207973

Regime-dependent commodity price dynamics: A predictive analysis*

Jesus Crespo Cuaresma

Vienna University of Economics and Business, Vienna, Austria

International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria

Wittgenstein Center for Demography and Global Human Capital (IIASA,VID/OEAW,UniVie)

Austrian Institute of Economic Research (WIFO), Vienna, Austria

Ines Fortin

Macroeconomics and Economic Policy, Institute for Advanced Studies, Vienna, Austria

Jaroslava Hlouskova

Macroeconomics and Economic Policy, Institute for Advanced Studies, Vienna, Austria

International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria

Department of Economics, University of Economics in Bratislava, Slovakia

Michael Obersteiner

Environmental Change Institute, University of Oxford, Oxford, UK

International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria

*The authors gratefully acknowledge financial support from the Austrian Science Fund FWF (project number P 30915-G27) and helpful comments of Robert Kunst that lead to an improvement of the paper.

Abstract

We develop an econometric modelling framework to forecast commodity prices taking into account potentially different dynamics and linkages existing at different states of the world and using different performance measures to validate the predictions. We assess the extent to which the quality of the forecasts can be improved by entertaining different regime-dependent threshold models considering different threshold variables. We evaluate prediction quality using both loss minimization and profit maximization measures based on directional accuracy, directional value, the ability to predict adverse movements and returns implied by a trading strategy. Our analysis provides overwhelming evidence that allowing for regime-dependent dynamics leads to improvements in predictive ability for the Goldman Sachs Commodity Index, as well as for its five sub-indices (energy, industrial metals, precious metals, agriculture, livestock). Our results suggest the existence of a trade-off between predictive ability based on loss and profit measures, which implies that the particular aim of the prediction exercise carried out plays a very important role in terms of defining which set of models is the best to use.

Keywords: Commodity prices, forecasting, threshold models, forecast performance, states of economy.

JEL Classification: Q02, C53, F47.

1 Introduction

This study aims at creating an econometric modelling framework to forecast commodity prices, taking explicitly into account the potentially different dynamics and linkages existing in different states of the world and using different performance measures to validate the predictions. The literature on commodity price forecasts can be categorized into two broad groups depending on the approach they take. While some studies use asset prices as predictors of commodity prices, a more agnostic approach exploits statistical methods to search for the most effective set of predictors of commodity price changes. The more common approach based on asset prices, routinely used by central banks, creates predictions of commodity prices using futures prices. Recently, some authors argue that such a forecasting method rather provides noisy signals about future spot prices (see Hong and Yogo, 2012; Gorton and Rouwenhorst, 2006; Groen and Pesenti, 2011).

The early literature on commodity price modelling and forecasting builds upon large macroeconometric specifications (Just and Rausser, 1981), while modern methods rely on univariate and multivariate time series modelling which jointly assess the dynamics of macroeconomic variables and commodity prices (see for example Ahumada and Cornejo, 2015). Groen and Pesenti (2011) and Gargano and Timmermann (2014) provide relevant examples of the more agnostic and flexible approach to model building in the context of commodity price forecasting. In both studies, the authors assess whether forecasts of commodity prices based on a large pool of macroeconomic predictors, can systematically improve upon naive benchmarks. Groen and Pesenti (2011) study the predictability of ten commodity indices in an out-of-sample experiment. They conclude that neither commodity exchange rates nor a broad cross-section of macroeconomic variables produce overwhelmingly strong evidence of spot price predictability when compared with random walk or autoregressive benchmarks. Gargano and Timmermann (2014), on the other hand, examine the out-of-sample predictability of seven commodity indices over the period 1947–2010, using macroeconomic and financial variables. They find that commodity currencies have some predictive power at short (monthly and quarterly) forecast horizons, while growth in industrial production and the investment-capital ratio have some predictive power at longer (yearly) horizons, a result that resembles that by Chen et al. (2010). Gargano and Timmermann (2014) also observe that commodity price predictability varies substantially across economic states, being strongest during economic recessions. Other models are employed in more recent contributions to the literature, such as those by Xu (2017, 2018, 2020).

Exploiting the co-movement of prices across commodities has been also shown to contribute to the improvement of prediction quality in commodity prices (see Ahumada and Cornejo, 2016). The inclusion of price information from other commodity markets in predictive specifications of a given agricultural commodity price has therefore become a modelling strategy often used in the empirical literature. Xu (2020), for example, expands the information set of models for daily corn prices by including the price in other localities as additional variables in a multivariate time series model. In parallel, efforts to improve forecasts of commodity prices by explicitly modelling their volatility have also been carried out (see for example Bernard et al., 2008; Ramirez and Fadiga, 2003; or the recent contribution by Degiannakis et al., 2020).

In striving for modelling frameworks with good predictive accuracy for commodity prices, in this contribution we assess the extent to which the quality of the forecasts depends on the state of the economy. Issues related to optimizing out-of-sample prediction in the presence of structural breaks and parameter instability have been particularly prevalent in the modern forecasting literature (see for example Giacomini and Rossi, 2010). We aim at assessing whether, for example, models tend to provide more accurate predictions in calm than in turbulent times. First findings in this direction were provided by Gargano and Timmermann (2014), who observe that commodity price predictability is better during recessions than during expansions. In stock and bond markets, the importance of models that account for regime-dependent parameters has often been acknowledged. Recent studies (e.g., Guidolin and Timmermann, 2005; for excess stock and bond returns or Guidolin and Timmermann, 2009; for short-term interest rates) have found that regime switching models may prove extremely useful to forecast over intermediate horizons, using monthly data. Guidolin and Ono (2006) find overwhelming evidence of regime switching in the joint process for asset prices and macroeconomic variables. They also find that modelling explicitly the presence of such regimes improves considerably the out-of-sample performance of a model of the linkages between asset prices and the macroeconomy. Jacobsen et al. (2016) investigate stock return predictability and find a strong positive relation between industrial metals and equity returns in times of recessions and a negative relation during expansions. In this study, we entertain different regime-dependent models (threshold models), considering different threshold variables to capture states of the world.

In addition, we assess the quality of commodity forecasts not only with the mean squared error (MSE), the traditional forecast performance measure used in many studies including Gargano and Timmermann (2014), but also with measures that evaluate directional accuracy, directional value, the ability to predict adverse movements, and returns implied by a trading strategy based on commodity price forecasts. These additional measures (also called profit measures, as opposed to the loss measures like mean-squared error or mean absolute error) do not directly assess forecast accuracy but relate to other dimensions of forecasting quality and may be more relevant than accuracy for particular applications in policy and applied work.

We create models to predict commodity price dynamics as captured by the changes in an overall commodity price index, as well as in five sub-indices (energy, industrial metals, precious metals, agriculture, livestock), for short- and long-term forecast horizons, using monthly observations in the period 1980–2018. Our forecast models include threshold models that are based on different threshold variables and we consider the various performance measures discussed above. Based on the extensive empirical evidence we conclude the following. There is overwhelming evidence that allowing for regime-dependent dynamics leads to improvements in predictive ability for commodity prices. This is the case because the differences in the characteristics of the dynamics and the interactions with other variables are not constant over time, but differ depending on particular phenomena (for instance, periods of high and low volatility, good and bad economic times, times of high/low interest rates or inflation, etc.). If these regimes are well delimited, the stability of dynamics and interactions in particular regimes allow for better predictions. This is not too surprising, since regime-dependence should be explicitly taken into account if it is present. However, the nature of these improvements also differs across predictive measures and sectors.

Our results show that an interesting trade-off appears between loss and profit measures, which implies that the particular aim of the prediction exercise carried out plays a very important role in terms of defining which set of models is the best to use. The optimal specifications for applications where the metrics for success are related to systematically predicting the direction of change of commodity prices accurately may thus be systematically different from those aimed at providing point predictions with an absolute minimal distance to the realized values.

The paper is structured as follows. In Section 2 we present the forecast models, where we describe the threshold models, our main focus, in more detail. In Section 3 we introduce the commodity data and present basic descriptive statistics. We also describe the explanatory and threshold variables. We present forecast performance measures, including traditional and new measures, in Section 4. The following section presents and discusses the empirical results, and Section 6 concludes.

2 Methodology

In order to address our research question, which deals with how different states of the economy (like recessions/expansions, high/low volatility, high/low inflation, high/low interest rates, market sentiment, etc.) affect the price forecasting performance of different commodity classes, we assess a class of threshold models (both univariate and multivariate). These type of models allow the specification to change in different regimes (states of the world), whose occurrence depends on the value of a given threshold variable. In principle, there is a large universe of potential threshold variables that could be use as a trigger quantity which determines the regime where the process resides. It has often been observed, for example, that variables may behave differently in booming and declining markets. Hence, indicators describing different stages of the business cycle (e.g., business cycle indicators, economic sentiment indicators, inflation, interest rates, spreads between long- and short-term interest rates) are useful in defining the corresponding states of the economy. On the other hand, the behavior of economic variables may vary in periods of high and low risk, which are usually identified by a high or low volatility in the equity markets. The level of the oil price may also induce different types of dynamics in commodity prices. In addition to these threshold variables, which have already been used before in the literature, we examine whether the correlation between stock and government bond markets, as well as the correlation between stock and oil markets (which are relevant in portfolio diversification) may lead to differences in the quality of commodity price forecasting models. Finally, we are interested in whether the level of the target variable itself, i.e., the commodity index, may be useful to define different states of the world.

In our application, the set of variables that are assessed as potential drivers of the threshold-nonlinearity and thus define the states of the economy is given by: the composite leading indicator for the US (CLI), the consumer confidence indicator for the US (CCI), the US inflation rate (INF), the 3-months money market rate in the US (IR), the spread between long-term and short-term US interest rates (spread), the volatility of the US stock market (VOLA), the oil price (oil), the correlation between the US stock and government bond markets based on a six months rolling window (COR), the correlation between the world stock market and the oil price based on a six months rolling window (COR-oil), the S&P Coldman Sachs commodity index (GSCI) and its sub-indices, as well as first differences of these variables. For more details see Table 12.

As the set of specifications aimed at forecasting commodity prices, we consider a large battery of model classes, including autogressive models, Bayesian vector autoregresive models, GARCH models, and vector error correction models. In addition to these

Table 1: Model description

Abbreviations	Model description
AR(p)	Autoregression in levels with p lags
DAR(p)	Autoregression in first differences with p lags
s-AR(p)	Subset autoregression in levels with p lags
s-DAR(p)	Subset autoregression in first differences with p lags
ARCH(p,q)	Autoregression conditional heteroskedasticity in levels with p lags in mean equation and q lags in variance equation
DARCH(p,q)	Autoregression conditional heteroskedasticity in first differences with p lags in mean equation and q lags in variance equation
GARCH(p,q)	Generalized autoregression conditional heteroskedasticity in levels with p lags in mean equation and q lags in variance equation
DGARCH(p,q)	Generalized autoregression conditional heteroskedasticity in first differences with p lags in mean equation and q lags in variance equation
TAR(p,k)	Threshold autoregression in levels with p lags and with k-th lag in threshold variable
TDAR(p,k)	Threshold autoregression in first differences with p lags and with k-th lag in the threshold variable
VAR(p)	Vector autoregression in levels with p lags
DVAR(p)	Vector autoregression in first differences with p lags
VEC(p, c)	Vector error correction model with p lags and c cointegration relationships
s-VAR(p)	Subset vector autoregression in levels with p lags
s-DVAR(p)	Subset vector autoregression in first differences with p lags
BDVAR(p)	Bayesian vector autoregression in first differences with p lags
TVAR(p,k)	Threshold vector autoregression in levels with p lags and with k-th lag in threshold variable
TDVAR(p,k)	Threshold vector autoregression in first differences with p lags and with k-th lag in threshold variable
RW	Random walk

specifications, which do not allow for threshold effects, we consider univariate and multivariate two-regime threshold models. All these models are listed in Table 1. The simplest threshold model is the threshold autoregression in levels with p lags and with k lags in the threshold variable, TAR(p,k)

$$y_t = \begin{cases} \phi_{01} + \sum_{i=1}^{p} \phi_{i1} y_{t-i} + \varepsilon_t, & \text{for } z_{t-k} \leq \gamma_\phi \\ \\ \phi_{02} + \sum_{i=1}^{p} \phi_{i2} y_{t-i} + \varepsilon_t, & \text{for } z_{t-k} > \gamma_\phi \end{cases} \tag{1}$$

where y_t is the log of the Goldman Sachs commodity index (or its sub-index) at time t, $z \in \mathbf{Z} \cup \Delta\mathbf{Z}$, with $\mathbf{Z}$ being the set of above mentioned threshold variables, namely $\mathbf{Z} = \{y$, CLI, CCI, INF, IR, spread, VOLA, oil, COR, COR-oil$\}$ and $\Delta\mathbf{Z}$ is the set of their first differences, i.e., $\Delta\mathbf{Z} = \{\Delta y$, ΔCLI, ΔCCI, ΔINF, ΔIR, Δspread, ΔVOLA, Δoil, ΔCOR, ΔCOR-oil$\}$. Finally, $\varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2)$. The estimator of γ_ϕ is the value of z that minimizes the sum of squared residuals in the non-linear regression (1), i.e.

$$\hat{\gamma}_\phi = \arg\max_z \left\{ \sum \hat{\varepsilon}(z)^2 \right\} \tag{2}$$

Once the estimator of γ_ϕ is found, (1) can be estimated by OLS.

We also consider threshold autoregressions in first differences with p lags and with a

k-th lag in threshold variable, TDAR(p, k)

$$\Delta y_t = \begin{cases} \theta_{01} + \sum_{i=1}^{p} \theta_{i1} \Delta y_{t-i} + \epsilon_t, & \text{for } z_{t-k} \leq \gamma_\theta \\ \\ \theta_{02} + \sum_{i=1}^{p} \theta_{i2} \Delta y_{t-i} + \epsilon_t, & \text{for } z_{t-k} > \gamma_\theta \end{cases} \tag{3}$$

where $\epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2)$, $\hat{\gamma}_\theta = \arg\max_z \left\{\sum \hat{\epsilon}(z)^2\right\}$ and $z \in \mathbf{Z} \cup \Delta\mathbf{Z}$.

In addition to univariate threshold models, we entertain multivariate threshold models, which generalize the class of threshold vector autoregression in levels with p lags and with a k-th lag threshold variable, TVAR(p, k). Let x_t be an $N-$dimensional vector, then the model under consideration is

$$x_{tn} = \begin{cases} \Psi_{0n1} + \sum_{l=1}^{p} \Psi_{ln1} x_{t-l} + \mu_{tn}, & \text{for } z_{t-k} \leq \gamma_\Psi \\ \\ \Psi_{0n2} + \sum_{l=1}^{p} \Psi_{ln2} x_{t-l} + \mu_{tn}, & \text{for } z_{t-k} > \gamma_\Psi \end{cases} \tag{4}$$

for $n = 1, \ldots, N$, where Ψ_{0n1} and Ψ_{0n2} are N-dimensional column vectors, Ψ_{ln1} and Ψ_{ln2} are $N \times N$ matrices, $\mu_t \sim \text{NID}(0, \Sigma_\mu)$, the S&P GS commodity index (or its sub-index) is the first element of x_t, i.e., $x_{t1} = y_t = \log(GSCI_t)$ and $z \in \mathbf{Z} \cup \Delta\mathbf{Z}$. Finally, γ_Ψ is estimated such that

$$\hat{\gamma}_\Psi = \arg\max_z \left\{\sum \hat{\mu}_1(z)^2\right\} \tag{5}$$

thus, the estimator of γ_Ψ is the value of z that minimizes the sum of squared residuals corresponding to the first equation in (4), i.e., the residuals corresponding to the commodity index. Vector x_t consists of the following macroeconomic and financial variables: the US composite leading indicator (CLI), the real effective exchange rate with respect to the US dollar (REER), the world stock index (stock), stock-to-use ratios, and additionally the S&P Goldman Sachs commodity index (GSCI) if the dependent variable is a commodity sub-index. Variables are logged, with the exception of the stock-to-use ratios.

Finally, we consider also a variation of threshold vector autoregression in first differences with p lags and with k-th lag in threshold variable, TDVAR(p, k) such as

$$\Delta x_{tn} = \begin{cases} \chi_{0n1} + \sum_{l=1}^{p} (\chi_{ln1})' \Delta x_{t-l} + u_{tn}, & \text{for } z_{t-k} \leq \gamma_\chi \\ \\ \chi_{0n2} + \sum_{l=1}^{p} (\chi_{ln2})' \Delta x_{t-l} + u_{tn}, & \text{for } z_{t-k} > \gamma_\chi \end{cases} \tag{6}$$

with parameter vectors and matrices defined analogously to those in the model above and

$u_t \sim \mathrm{NID}(0, \Sigma_u)$, $z \in \mathbf{Z} \cup \Delta\mathbf{Z}$. The threshold value γ_χ is estimated such that

$$\hat{\gamma}_\chi = \arg\max_z \left\{ \sum \hat{u}_1(z)^2 \right\} \tag{7}$$

Thus, the estimator of γ_χ is the value of z that minimizes the sum of squared residuals corresponding to the first equation in (6), i.e., the residuals corresponding to the commodity index in first differences ΔGSCI. As in (4), the regimes are implied by the first equation and taken as given for the remaining equations in (6).

In our empirical analysis, when we compare threshold and linear models, we consider up to three lags of the variables (with $p = 3$ being the maximum lag length) and up to twelve lags for the threshold variable under consideration (with $k = 12$ being the maximum lag length). Models are compared and selected according to out-of-sample performance measures.[1] When we compare threshold models with a larger set of models, namely with all the models listed in Table 1, the lag structure is determined in-sample making use of the Akaike information criterion.

3 Data

We use the family of S&P GSCI (Standard & Poors Goldman Sachs Commodity Index) indices to measure commodity prices. We use both the total aggregate commodity index (S&P GSCI) and five sub-indices that reflect the developments of certain components of the index, namely energy (63%), industrial metals (11%), precious metals (4%), agriculture (15%), and livestock (7%). In brackets we report the respective sector weights in the total commodity index, see Table 2. The S&P GSCI is regarded as a benchmark for investment in commodity markets and is designed to be a tradable index. It is calculated using a world production-weighted basis and includes physical commodities that are traded in liquid futures markets. The criteria for inclusion into the index are based on trading volume. In addition, the contracts must be denominated in US dollars and traded in an OECD country or on a trading facility that has its principal place of business in an OECD country. The current S&P GSCI comprises 24 commodities from all commodity sectors with a high exposure to energy (63%). These energy contracts include crude oil, heating oil, and gasoline traded in the US, as well as crude oil and gasoil traded in Europe. Table 11 in the appendix lists all contracts included in the S&P GSCI and their

[1] This implies that we explicitly consider all combinations of explanatory variables and all lags of the explanatory and threshold variables up to the specified maximal lag lengths and then choose the best model according to the given forecast performance measure.

Table 2: S&P GSCI Sector Weights

Sector	2019 Reference Percentage Sector Weight
Energy	62.63%
Industrial metals	11.16%
Precious metals	4.14%
Agriculture	15.41%
Livestock	6.65%
Total	99.99%

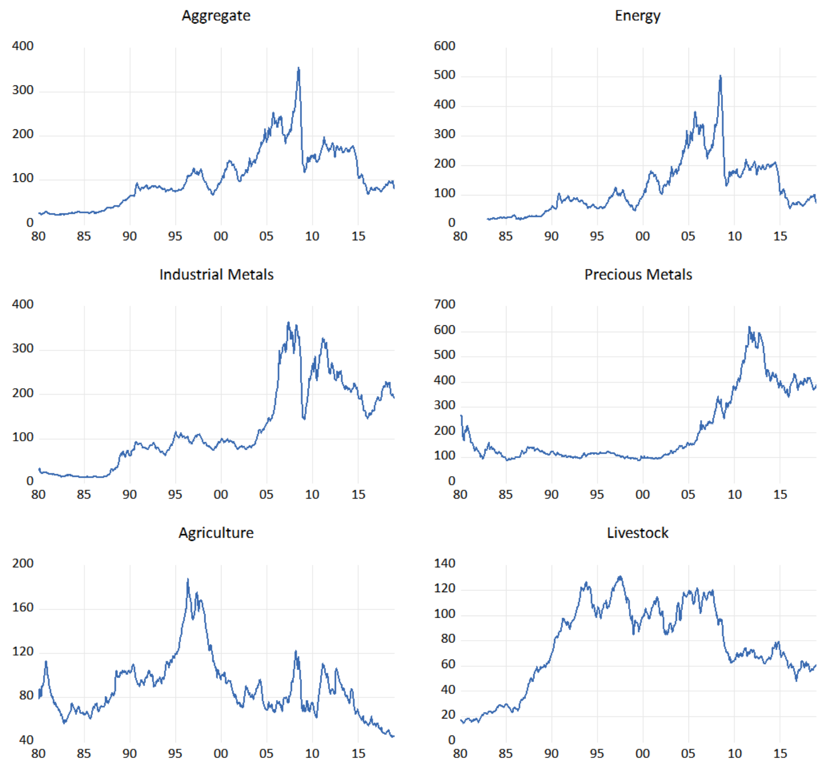
respective weights and trading places. We consider the class of total return indices.[2] For more information on the S&P GSCI see S&P Dow Jones (2019). We present graphs of the commodity price indices and their returns in Figures 1 and 2. Some descriptive statistics and correlations are given in Table 3. It can easily be seen that the price developments are quite heterogeneous across indices, with only the overall and the energy indices displaying rather similar price developments. The volatility in returns varies considerably as well, which has a direct impact on the forecasting accuracy of econometric models. The monthly returns of the energy index, for example, show a standard deviation of 7.7% over the total data sample (1980–2018), while the corresponding value for the livestock index is only 3.5%. Overall, the correlations between different commodity sector returns are low, so it makes sense to analyse the different sectors separately. There is only one exception: the overall and the energy indices are highly correlated (0.9).

As macroeconomic and finance variables in our models, we take the composite leading indicator for the US (CLI), the real effective exchange rate related to the US dollar (REER), and the world stock indicator (stock).[3] In addition, we employ fundamental variables summarizing the forces in the commodity market: stock-to-use ratios (stu) for oil (worldwide), wheat (US) and meat (US). More precisely, we use the worldwide oil stock-to-use ratio for the aggregate index and for the sub-indices energy, industrial metals and precious metals, we use the US wheat stock-to-use ratio for the sub-index agriculture, and we use the US meat stock-to-use ratio for the sub-index livestock. In those cases where we model commodity sub-indices, we also use the total commodity index as an additional variable. As threshold variables, we use the composite leading indicator for the US (CLI), the consumer sentiment indicator for the US (CCI), the US inflation measured by the

[2]The S&P GSCI total return indices reflect the performance of a total return investment in commodities, i.e., the contract daily return plus the daily interest on the funds hypothetically committed to the investment.

[3]In alternative modelling exercises, we also included the industrial production index as an additional variable, but removed it from the list of variables as it is heavily correlated with the CLI and did not help to improve the forecast performance substantially.

Figure 1: Commodity prices
The figure plots monthly values of the S&P GSCI Aggregate Commodity Index and the five sub-indices. The sample period is January 1980 to December 2018.

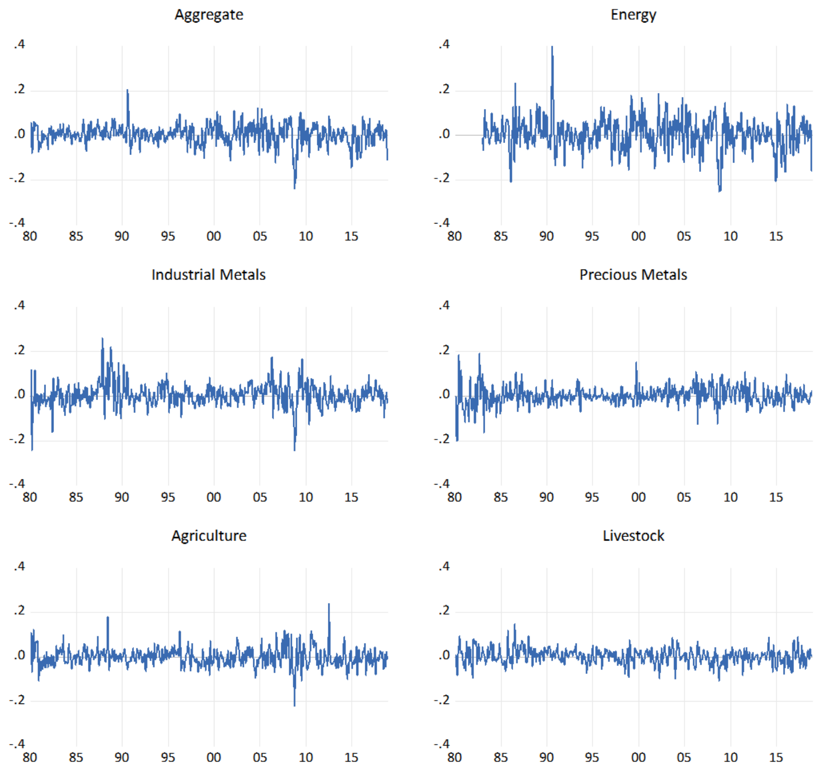


consumer price index (INF), the 3-months money market interest rate (IR), the spreads between long-term and short-term US interest rates (spread), the volatility of the S&P 500 (VOLA), the correlation between the US stock and government bond markets (COR), and the correlation between the global stock market and the oil market (COR-oil). The correlations are calculated between daily returns in the respective markets, over a rolling window of 130 trading days (i.e., approximately six months), recorded at the end of a given month. For details on all the data we use, see Table 12.

The data sample covers monthly observations for the period ranging from January 1980 through December 2018. We consider rolling-window estimation for our analysis, i.e., we keep the size of the estimation sample constant and equal to 20 years, and move forward the sample by one month, while re-estimating the model parameters. The out-of-sample period, in which we evaluate the forecast performance, ranges from January 2005 to December 2018.[4] Note that "best" models are chosen based on the individual

[4] As a robustness check, we also performed the analysis over the out-of-sample period January 2001 to December 2018 and obtained similar results.

Figure 2: Commodity returns
The figure plots monthly returns of the S&P GSCI Aggregate Commodity Index and the five sub-indices. The sample period is January 1980 to December 2018.



forecast performance of the models for all lags (up to specified maximum lags) and all combinations of variables under consideration.

4 Forecast evaluation

The evaluation of different commodity price forecasts are carried out not only employing traditional loss measures, like mean absolute error (MAE) and mean squared error (MSE), but also profit-based measures like directional accuracy (DA), directional value (DV) and directional value of turning points (TP). The latter might be more relevant under certain conditions.[5] The directional accuracy indicator, or hit rate, is a binary variable measuring whether the direction of a price change was correctly forecast. The directional value incorporates the economic value of directional forecasts by assigning to each correctly

[5]Granger and Pesaran (2000) report on the fact that the forecast evaluation literature appears to be biased towards statistical accuracy measures, while neglecting measures based on the economic importance for the forecasts.

Table 3: Summary statistics for commodity returns.
The table reports the mean, standard deviation, skewness, and kurtosis for monthly commodity returns over the sample period January 1980 to December 2018. Commodity returns are computed from the S&P GSCI commodity indices. The last column shows returns of the world stock market index.

	all	energy	ind. met.	prec. met.	agriculture	livestock	stock
Descriptive statistics							
Mean (%)	0.3771	0.6246	0.5491	0.1750	-0.0300	0.3361	0.6706
Std. (%)	4.8416	7.6745	5.4517	4.3801	4.3028	3.5431	3.6639
Skew.	-0.6128	0.1151	0.1907	0.0174	0.5327	0.0220	-0.8491
Kurt.	5.6013	5.7475	6.6736	6.1019	6.5898	3.7059	7.3947
Correlation matrix							
energy	0.9406	1					
ind. met.	0.3947	0.2819	1				
prec. met.	0.2031	0.1398	0.2738	1			
agriculture	0.3274	0.1419	0.2362	0.1843	1		
livestock	0.1953	0.0780	0.0672	-0.0534	0.0611	1	
stock	0.2712	0.1922	0.3039	0.1306	0.1888	0.1180	1

predicted change its magnitude. The directional accuracy of turning points describes the ability to predict adverse movements, i.e., turning points. Note, however, that it is difficult to compute a reliable value of this measure for different regimes, which arise naturally when we work with threshold models.[6] Therefore, we do not use turning points in the comparison of threshold and linear models, where the analysis of regime-based performance is essential. However, we use this performance measure when analyzing overall performance differences.

The loss-based and profit-based performance measures are formally defined as follows,

$$\begin{aligned}
AE_{t+h,h} &= \left|\log \hat{P}_{t+h|t} - \log P_{t+h}\right| \\
SE_{t+h,h} &= \left(\log \hat{P}_{t+h|t} - \log P_{t+h}\right)^2 \\
DA_{t+h,h} &= I\left(\text{sgn}(P_{t+h} - P_t) = \text{sgn}(\hat{P}_{t+h|t} - P_t)\right) \\
DV_{t+h,h} &= |P_{t+h} - P_t| \; DA_{t+h,h} \\
TP_{t+h,h} &= \begin{cases} DA_{t+h,h} & \text{if } \; \text{sgn}(P_{t+h} - P_t) \times \text{sgn}(P_t - P_{t-h}) = -1 \\ \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{8}$$

[6]Note that three consecutive time points are needed to calculate the TP. There are usually not that many turning points in general, and there tend to be even less in each regime. It may easily happen that the three consecutive time points required to calculate the measure are not in the same regime.

where P_t is the price of the commodity index at time t, $\hat{P}_{t+h|t}$ is the forecast of the price of the commodity index for time $t+h$ conditional on the information available at time t, i.e., h is the forecast horizon, and $I(\cdot)$ is the indicator function.

In addition, we consider forecast ability measures based on the returns implied by a simple 'buy low, sell high' trading strategy, which is based on buying the commodity index if its price is forecast to rise and selling it when its price is forecast to fall. This strategy is described (for foreign exchange rates), e.g., in Gençay (1998).[7] Predicted upward movements of the commodity index with respect to the actual value (positive returns) are executed as long positions, while predicted downward movements (negative returns) are executed as short positions. The following discrete return $r_{t+h,h}$ is implied by the 'buy low, sell high' trading strategy,

$$r_{t+h,h} = \begin{cases} \frac{1}{P_t}\left(P_t - P_{t+h}\right) = 1 - \frac{P_{t+h}}{P_t}, & \text{if } \hat{P}_{t+h|t} < P_t \\ & \text{commodity index is bought at } t+h \\ & \\ \frac{1}{P_t}\left(P_{t+h} - P_t\right) = \frac{P_{t+h}}{P_t} - 1, & \text{if } \hat{P}_{t+h|t} > P_t \\ & \text{commodity index is sold at } t+h \end{cases}$$

The aggregate performance measures for each model are calculated over the out-of-

[7]Notice, that while the 'buy low, sell high' trading strategy is not a feasible trading strategy for physical commodities, as it would require calculating spot returns net of the cost of carry such as storage costs and insurance, it may well be implemented for investable indices like the GSCI indices. See Miffre (2016) for an overview on strategies in commodity markets.

sample period for a given forecasting horizon as follows,

$$
\begin{aligned}
MSE_h &= \sum_{j=0}^{T_2-T_1} \frac{SE_{T_1+j,h}}{T_2-T_1+1} \\
MAE_h &= \sum_{j=0}^{T_2-T_1} \frac{AE_{T_1+j,h}}{T_2-T_1+1} \\
DA_h &= 100 \sum_{j=0}^{T_2-T_1} \frac{DA_{T_1+j,h}}{T_2-T_1+1} \\
DV_h &= 100 \frac{\sum_{j=0}^{T_2-T_1} DV_{T_1+j,h}}{\sum_{j=0}^{T_2-T_1} |S_{T_1+j} - S_{T_1+j-h}|} \\
&= 100 \frac{\sum_{j=0}^{T_2-T_1} |\hat{S}_{T_1+j} - S_{T_1+j-h}| DA_{T_1+j,h}}{\sum_{j=0}^{T_2-T_1} |S_{T_1+j} - S_{T_1+j-h}|} \\
TP_h &= 100 \frac{\sum_{j=0}^{T_2-T_1} TP_{T_1+j,h}}{\sum_{j=0}^{T_2-T_1} TP^{actual}_{T_1+j,h}}
\end{aligned}
$$

where

$$
TP^{actual}_{t+h,h} = \begin{cases} 1 & \text{if } \operatorname{sgn}(P_{t+h} - P_t) \times \operatorname{sgn}(P_t - P_{t-h}) = -1 \\ \\ 0 & \text{otherwise} \end{cases}
$$

$$
R_h = 100 \left[\left(\sum_{j=0}^{T_2-T_1} \frac{r_{T_1+j,h}}{T_2-T_1+1} + 1 \right)^{12/h} - 1 \right]
$$

where T_0 = January 1980, T_1 = January 2005, and T_2 = December 2018.

5 Results

In analyzing the value of threshold models in commodity price forecasting, we focus on different metrics. First, we compare threshold models with linear models, which is the most natural comparison to find out about the value of threshold models as predictive instrument. In this context, we also analyse the differences across threshold models implied by the different threshold variables. We use different performance measures to evaluate the forecasting performance and consider both total and regime-based performance measures. In addition to pure comparison issues, we look at threshold variables and selected explanatory variables in the best threshold models, and also discuss the pattern of performance criteria found for the two regimes. Furthermore, we look at the sector-specific performance of best threshold models. Finally, we compare threshold models with a much larger set of models to find out whether threshold models tend to outperform specifications created out of this expanded set of covariates, and consider an additional performance measure related to turning points.

Threshold models and linear models

Our primary focus is to compare the performance of best threshold models (for a given threshold variable) with the performance of linear models. Threshold models include (vector) autoregression threshold models in levels and differences (TAR, TDAR, TVAR, and TDVAR) and linear models contain (vector) autoregression models in levels and differences (AR, DAR, VAR, and DVAR), see Table 1. Usually we look at threshold models where the threshold variable is in levels *or* in differences and do not explicitly differentiate between the two. Sometimes, however, it makes more sense to perform the analysis separately, looking at the threshold variable being either in levels or in differences. For example, in the case of volatility we may be more interested in what happens in periods of high and low risk, and thus look at levels, as compared to what happens in periods of changing risk (implied by the use of first differences). On the other hand, in the case of oil it might be more reasonable to look at differences than at levels, due to potential structural breaks in price level dynamics over the whole sample period.

Before we look at threshold versus linear models, we investigate the performance of threshold variables other than the dependent variable itself and examine whether different threshold variables imply large differences in the forecasting performance of their corresponding specifications. We therefore compare the performance of the best threshold model when the threshold variable coincides with the lagged commodity price index variable (self-exciting models) with that of the best threshold model when the threshold variable is one of the other nine threshold variables listed in Table 12. With this exercise, we assess whether states of the world defined by the commodity price itself are informative enough to capture the complex economic environment implied by various different threshold variables. Our results suggest that the use of other threshold variables different from the commodity price index adds predictive information to our models. Figure 3 shows how many threshold models (from the maximum number of nine threshold variables) outperform the self-exciting model, for different performance measures, different forecast horizons and the various commodity sectors. The self-exciting model is mostly outperformed by the other threshold models, sometimes even by the majority of these. In all but three cases (out of a total of 120), the self-exciting model is outperformed by at least one threshold model. The self-exciting model is only better than any other threshold model for precious metals and $h = 3$ considering the return, and $h = 6$ considering the directional value, as well as for the livestock and $h = 3$ considering the directional value. We show later (see the following subsection) that, in general, the forecasting performance implied by the different threshold models is rather similar. For the overall GSCI index, for example, nearly all (i.e., eight or nine) threshold models outperform the self-exciting specification for forecast horizons of three, six and twelve months, irrespective of which performance measure we consider. This implies that explicitly acknowledging information like economic sentiment, uncertainty, interest rates, oil prices or correlation can help improve commodity price forecasting. Results are somewhat less clear-cut for energy, industrial metals and agriculture, and they are the least strong for precious metals and livestock. Even in these two sectors, however, in most cases the best threshold models are not the self-exciting ones. Note that for livestock and a forecast horizon of three months, only one single threshold variable helps to improve the forecast in terms of MAE, MSE and DA, namely the correlation between stock and bond markets.

In contrasting threshold and linear models, we first compare the performance of the best threshold model with the performance of the "corresponding" linear model. By "corresponding" linear model we mean that linear model that uses exactly the same variables and the same lag structure as the threshold model. In more detail, we compare the "total" performance (i.e., performance over the whole out-of-sample period) and also the performance in the two regimes separately, where the regimes for the corresponding linear model are implied by the regimes of the best threshold model for a specific threshold variable. We observe that best threshold models with respect to specific threshold variables mostly outperform the corresponding linear models, and also the best linear models. In addition, threshold models outperform the corresponding linear models in at least one regime; mostly, however, in both regimes. We present one exemplary table, Table 4, which shows the performance of the best threshold model and the performance of the corresponding linear model for the aggregate GSCI, for the threshold variable "spread" (difference between long- and short-term US interest rates) in levels or differences. We have chosen this model based on the best short-term forecasting performance (MSE, 1-month ahead) for the overall GSCI index. We observe that in all cases (for horizons of one, three, six and twelve months ahead) the total performance (with respect to all considered performance measures) of the best threshold model is better than the total performance of the corresponding linear model. Except for one case, the total performance of the best threshold model is also better than the total performance of the best linear model. When comparing the regime-based performance of the best threshold model and the regime-based performance of the corresponding linear model, then the best threshold model outperforms the corresponding linear model in both regimes in most of the cases (16 out of 20). The best threshold model is never outperformed by the corresponding linear model in both regimes. Similar observations follow for Table 6 for the aggregate GSCI with the threshold variable being volatility of the US stock market (in levels), which we discuss below in more detail concerning a different aspect.

Second, we check whether the total performance of the best threshold model is better than the total performance of the best linear model (out of *all* linear models). Note that the best threshold model (among all threshold variables) always outperforms the best linear model, if we look at mean values of the performance criteria over the out-of-sample period. Even more, in basically all cases the best threshold model for *any* given threshold variable outperforms the best linear model; the minimum number of threshold models outperforming the best linear model is four, see Figure 4. In order to shed more light on this observation, we additionally consider the complete distribution of the difference between the squared errors (between the best linear and the best threshold models) observed in the out-of-sample period, not only their mean (which is the basis of the MSE). Figure 5 shows boxplots[8] of this difference across all commodity sectors, for forecast horizons of one and twelve months. The average difference is always clearly positive, which confirms our previous observation that the best threshold model outperforms the best linear model on average, i.e., using the mean squared error. A striking further observation is that the support of the distribution varies substantially across the different commodity sectors, and the pattern observed is similar for forecast horizons of one and twelve months. The interquartile range, for example, is largest in the energy sector and smallest in the livestock sector, for both forecast horizons. In particular, the error range for energy is approximately three times that for livestock when forecasting one month ahead, and this ratio is even as large as about ten when forecasting twelve months ahead. This implies that the difference between best threshold and best linear models is rather large in the energy sector and thus threshold models may indeed help to reduce the forecast error in forecasting energy prices, while in livestock the two competing models perform rather similarly or, put differently, threshold models do not provide so much additional value. The industrial metals and agriculture sector also display rather large

[8]The box portion of a boxplot represents the first and third quartiles (middle 50% of the data), and the difference between them represents the interquartile range q. The median is depicted using a line through the center of the box, while the mean is drawn using a symbol. The staples show the values that are outside the first and third quartiles, but within the inner fences that are defined as the first quartile minus $1.5q$ and the third quartile plus $1.5q$. The shaded region displays approximate confidence intervals for the median (under certain restrictive statistical assumptions), i.e., $median \pm 1.57q/n$, where n is the number of observations. Shading may be useful in comparing differences in medians: if the shades of two boxes do not overlap, then the medians are, roughly, significantly different at a 95% confidence level.

differences in terms of a larger interquartile range, similar to energy, while the sector precious metals shows a rather small difference, like livestock. On the other hand, Figure 6 presents boxplots for energy sector depicting errors that form the MAE, MSE, DV and return of the best linear and best threshold models for forecast horizon of twelve months. Again, the larger mass in the positive region suggests the superiority of the best threshold models over the best linear models, i.e., the differences are consistent across predictive measures.

The best threshold model practically always outperforms the random walk and the autoregressive model, the two standard univariate predictive benchmarks. In total, our design is composed of 1,200 combinations (five performance measures, four forecast horizons, six commodity sectors, and ten threshold variables). In only nine cases out of these (0.75%) the random walk outperforms the best threshold model. This happens for forecast horizons of twelve and six months (in six and three cases, respectively). The performance of the autoregressive benchmark is similarly weak. In only three cases (0.25%) it performs better than the best threshold model (twice for the twelve-months forecast horizon and once for the one-month horizon).

Threshold models and threshold variables/explanatory variables

Analysing the best threshold models with respect to threshold variables across the commodity sectors, a certain pattern can be seen for all sectors but energy. For industrial metals, the threshold variable of the best threshold model with respect to the loss-based measures is CLI, and with respect to the profit-based measures it is oil. The most frequently occurring threshold variable (in best threshold models) for precious metals is oil for all performance measures but MSE, while for MSE most of the best threshold models rely on the threshold variable COR. The majority of best threshold models for agriculture and livestock employ the threshold variable COR. Finally, in the aggregate sector the best threshold models are often based on the threshold variables CLI and CCI, while there is no obvious pattern for the energy sector (regarding the threshold variable of the best threshold models). See Figure 7, which shows a ranking of the nine threshold variables with respect to MSE and return.

In general, the forecasting performance of different best threshold models (implied by the different threshold variables) does not vary substantially. Table 5 provides some information on the variability across best threshold models related to the best linear model. In particular, the table reports the average deviation of best threshold models given a certain threshold variable from the *best overall* threshold model ("average deviation"), in proportion to the deviation of the best linear model from the best threshold model ("linear deviation"). Note that the best threshold model is always better than the best linear model, the "average" threshold model, however, may be worse than the best linear model (implied by a ratio larger than one in the table). The latter is rarely the case, however. In almost all cases (117 out of 120) the average deviation is smaller than the linear deviation (reflected by a number in the table that is smaller than one), and often to a very large extent. In roughly two thirds of all cases the average deviation is less than half the linear deviation, implying that, in general, threshold models seem to perform (similarly) well and considerably better than the best linear model.

We examine the nature of the variables included in the set of best threshold models, so as to assess the relative importance of different theoretical drivers of commodity price dynamics. In the group of best linear models, one group of commodity indices can be found which are similar among themselves but different from the other indices with respect to the variables included in the best models. This group includes the aggregate sector, the energy sub-sector and the industrial metals sub-sector. The remaining indices (precious metals, agriculture, livestock) are different from this group but also different from each other. In this (first) group the CLI indicator is important for forecasting while the oil stock-to-use ratio does not seem to help forecasting. By contrast, the real effective exchange rate (REER), the world stock index and the aggregate GSCI index (for the sub-sectors) are sometimes included and sometimes not, depending on the forecast horizon and performance criterion used. The REER appears to be more relevant for longer forecast horizons (included for twelve months forecast horizon for all five performance criteria for industrial metals, and for the twelve months forecast horizon for four out of five performance criteria for energy). For precious metals, the CLI is not important at all as a predictor (not included in any of the twenty best models), while for livestock both CLI and the world stock market index are very relevant.

With respect to the best threshold models, the pattern is somehow similar. There is the same group of commodity indices (aggregate sector, energy, industrial metals) where the CLI is an important predictor, the oil stock-to-use ratio is not very important, and the real effective exchange rate, the world stock index, and the GSCI aggregate index (for sub-sectors) are sometimes included in the best predictive specifications, but sometimes not. Figure 8 shows how often a given explanatory variable is included in the best threshold model, considering the total of ten best threshold models (for each threshold variable under consideration: commodity price, VOLA, CLI, CCI, IR, INF, oil, spread, COR, COR-oil), i.e., the maximum possible number is ten. For the sector precious metals the most important variables are REER and the oil stock-to-use ratio. For the sectors agriculture and livestock the most important variable is the world stock index, followed by CLI.

Threshold models and performance criteria

We investigate whether there is a structural pattern as to when, i.e., in which regime, the threshold model performs better with respect to the different performance criteria. In some situations, loss measures (MAE, MSE) and profit-based measures (DA, DV, return) behave differently when comparing between regimes. For instance, in threshold models with volatility as the threshold variable, loss measures (MAE, MSE) seem to perform better in times of low volatility (than in times of high volatility) while profit-based measures (DA, DV, return) seem to perform better in times of high volatility (than in times of low volatility). This is most pronounced for the aggregate GSCI index, energy and industrial metals, while for the other sectors the evidence is mixed. We now look at the threshold variable volatility in levels (not in levels or differences), as we are explicitly interested in contrasting times of low and high economic uncertainty. Table 6 presents the case of the aggregate GSCI with the threshold variable volatility (in levels), where red shading indicates better performance between the two regimes implied by the threshold model. The results suggest that, for all forecast horizons, commodity prices can be forecasted more accurately in times of low volatility than in times of high volatility; however, directional accuracy, directional value and the returns of a simple trading strategy (i.e., all profit measures) are higher in times of larger volatility. While the first observation can probably be explained through lower price variability and thus better forecasting ability in times of low uncertainty, the second observation may be related to the chances of making more profits in large volatility markets when the direction of price change is forecast correctly.

If the threshold variable is inflation (in levels) forecast accuracy is better in times of low inflation than in times of high inflation for forecast horizons of one, three and six months, while there are not really strong differences for the 12-months forecast horizon. This is true for the aggregate commodity index and the energy subsector and may be explained in the following way: if one assumes that energy prices, and also total commodity prices, depend positively on inflation, then lower inflation may go hand in hand with smaller price variability and thus with better forecasting ability, which is true up to the six-months-ahead forecast. Profit-based measures (for the aggregate index and energy), on the other hand, seem to behave in the opposite way: they perform better in times of high inflation than in times of low inflation (for forecast horizons of up to six months). Price forecasting for industrial metals works differently in the sense that for this sector all performance measures, including MAE and MSE, perform better in times of low inflation for a twelve-month forecast horizon.

If the threshold variable is the consumer confidence indicator or the correlation between stock and bond markets (both in levels) then the loss measures seem to perform better in the second regime (higher confidence, larger correlation) while the profit-based measures perform better in the first regime (lower confidence, smaller correlation). This observation is true for the overall commodity index and the energy sector, for the other sectors the evidence is mixed.

There is a somewhat remarkable observation for the aggregate commodity index and the composite leading indicator (in levels) as the threshold variable. In this case, all performance measures, including loss and profit measures, perform better in times of a booming economy for shorter forecast horizons (one month and three months), but perform better in times of a shrinking economy for longer forecast horizons (six and twelve months). This pattern cannot be observed for any commodity subsector.

Threshold models across sectors

When we compare threshold models across commodity sectors, our findings suggest that those sectors that lead to more accurate predictions (i.e., their loss measures, MAE and MSE, are the lowest ones) yield the lowest profit, and the other way round. Figure 9 presents MSE and returns of the set of best threshold models for the different commodity sectors. For example, prices of livestock, precious metals and agricultural commodities can be predicted comparatively well, while they provide low returns. On the other hand, prices of energy and industrial metals lead to the highest prediction errors but yield the largest returns.[9] This observation holds over all forecast horizons. A possible explanation could be that larger deviations of the forecasts from its realizations are needed in order to increase the implied profit, even if some deviations are in the "wrong" direction, as long as this happens in a sufficiently low number of cases. This pattern holds, albeit to a lesser quantitative degree, also for the other profit-based measures. Directional accuracy and directional deviation seem to be higher for commodity sectors which are harder to predict. An overview of all performance measures across all sectors and forecast horizons is presented in Figure 10.

Considering best threshold models, both loss measures, MAE and MSE, and the return display a clear structure relating to the forecast horizon. The loss measures increase, i.e., forecast accuracy decreases, with an increasing forecast horizon, so commodity prices can

[9]Note that the different forecasting accuracy across commodity sectors corresponds to the different variability in returns, as suggested before. Commodity sectors with smaller variability are easier to predict (e.g., according to MSE) than those with larger variability, see Table 3, Figure 9 and Figure 10.

be better predicted in the short term than in the long term. For example, the MSE in forecasting aggregate commodity prices increases from 0.28% when forecasting one month ahead to 6.47% when forecasting twelve months ahead. Also according to the return we observe the best performance for the shortest forecast horizon, with decreasing performance for increasing forecast horizons. While the return implied by a simple trading strategy for the aggregate commodity index is 31.5% when forecasting one month ahead, the corresponding return is only 12.4% when forecasting twelve months ahead.[10] The observed patterns (for MAE, MSE, return) with respect to the forecast horizon hold for all commodity sectors. For the other two profit-based measures (DA, DV) the behavior with respect to the forecast horizon is not the same across commodity sectors. While for precious metals and agriculture the directional accuracy and directional value grow with increasing forecast horizons, the picture is mixed for energy, industrial metals, livestock and for the aggregate sector. Mostly, however, the directional accuracy (value) is largest when forecasting twelve months ahead. See Table 8 and Figure 9.

Table 7 indicates that the commodity sector whose returns dominate those of the others is most of the time the industrial metals sector. Exceptions are the energy sector for return and DV for $h = 1$ and the sector of agriculture for DA and DV for $h = 12$. The sector with the best loss-based performance is livestock. The smallest loss-based performance occurs for livestock in case of one month forecast horizon, namely 2.42% for MAE and 0.1% for MSE, and the largest profit-based performance occurred for agriculture for twelve months forecast horizon, namely 80.95% for DA and 89.41% for DV and for energy sector where the return of 46.74% occurred in the case of one month forecast horizon.

Threshold models and larger class of models

In addition to standard linear autoregressive models, we consider a much larger class of models in order to find out whether threshold models also outperform other specifications. This class includes different univariate GARCH models, vector error correction models and Bayesian models (see Table 1). For this much larger class of models, it is not computationally feasible to consider a large number of different lag combinations and choose the best model according to the implied out-of-sample performance measure, as we did before. Now the lag structure is determined in sample using the Akaike information

[10]Notice that the return we report does not account for potential trading costs. The returns from actual trading strategies related to different forecast horizons which include trading costs may be different, and the current pattern with respect to the forecast horizon may not be preserved.

criterion.[11] We also use an additional performance measure, namely the proportion of correctly forecasted turning points (TP). As discussed before, see Section 4, this measure cannot be reliably computed for the two regimes separately and thus has not been used in the previous analysis.

Our results show that threshold models are the best models in the vast majority of cases. In only 9 out of a total of 144 cases (six commodity sectors, six performance measures and four forecast horizons) the threshold model is beaten by an alternative model. In five cases, the best model is the vector error correction model, in three cases the (subset) vector autoregression, and once the Bayesian vector autoregression. These exceptions seem to be randomly scattered across sectors, performance measures and forecast horizons with probably one exception: when forecasting agriculture commodity prices twelve months ahead, the vector autoregression model beats the threshold model if we look at DA, DV and return. Note that the threshold model is always the best model if we consider MSE and TP, i.e., over all sectors and forecast horizons.

None of the best models found now can keep up with the best models found before, however. In all cases without any exception, the best model determined in our previous analysis, which is always a threshold model, outperforms the best model found now, including the cases when the best model now is not a threshold model (see Table 9).[12] As best threshold models for different threshold variables do often perform similarly (well), as found in our previous analysis, not only the best threshold model but often also other threshold models (with different threshold variables) outperform the corresponding best model found now.

Both loss measures and the return show a clear pattern with respect to the forecast horizon: forecast accuracy decreases with an increasing forecast horizon, and so does the return. The proportion of correctly forecast turning points, which was not analyzed before, does not show a uniform pattern with respect to the forecast horizon. However, it is clearly largest for the 12-months forecast horizon for the total commodity index, for energy and industrial metals, while it is largest for the 1-month forecast horizon for the remaining sectors (precious metals, agriculture, livestock). With respect to differences across sectors, TP does not show the same pattern over different forecast horizons, but the overall index and the energy subsector are among the lowest TPs for forecast horizons up to six months. When forecasting twelve months ahead, this situation reverses, and the overall index and energy are actually among the best (ranking third and second)

[11] Note that the in-sample lag determination is in some sense more restrictive and may provide (slightly) inferior forecast models, if we compare the simple linear models and the threshold models used before.

[12] This comparison does not include best models with respect to the proportion of correct turning points, as this measure was not used before.

according to TP (see Table 9).

For the aggregate index and energy all best threshold models (but one) with respect to TP rely on a threshold variable that is connected to oil (OIL or COR-oil). All models for the aggregate index, and for energy except for one case, contain the oil stock-to-use ratio. Best models for precious metals according to TP are either based on a threshold variable related to oil or the oil stock-to-use ratio is among the explanatory variables. The same holds for industrial metals and livestock. For all indices best models according to TP rely on an oil related threshold variable, for a forecast horizon of twelve months. For all indices (but agriculture) is the REER included in the best model (according to TP) for a 12-months forecast horizon.

To sum up, our analysis emphasizes the importance of modelling regime-dependent dynamics and linkages in commodity prices in order to achieve out-of-sample predictive gains. Although the potential of nonlinear specifications to improve forecasts differs across the different sub-indices considered, modelling threshold effects tends to systematically improve predictions. However, depending on whether the aim of the prediction exercise is to minimize loss or to maximize profits, the particular structure of the optimal threshold model may be very different.

6 Conclusions

In this paper we present overwhelming evidence that allowing for regime-dependent dynamics in models for commodity prices leads to improvements in predictive ability. This follows from the fact that the characteristics of the dynamics of commodity prices and their interactions with other variables are not constant over time, but differ depending on particular phenomena (for instance, periods of high and low volatility in the equity markets, good and bad economic times, times of high/low interest rates or inflation, etc.). If these regimes are properly delimited, the stability of dynamics and interactions in particular regimes allow for better predictions. However, the nature of these improvements also differs across predictive measures and sectors.

We assess the quality of commodity forecasts with a variety of different performance measures: in addition to the mean squared error, the traditional forecast performance measure used in many studies, we also consider measures that evaluate directional accuracy, directional value, the ability to predict adverse movements, and returns implied by a trading strategy based on commodity price forecasts. These additional profit-based measures do not directly assess forecast accuracy but relate to other dimensions of forecasting quality and may be more relevant than accuracy for particular applications in policy and applied work. We create an econometric modelling framework to predict commodity price dynamics as captured by the changes in an overall commodity price index, the S&P Goldman Sachs Commodity Index, as well as in five sub-indices (energy, industrial metals, precious metals, agriculture, livestock). We consider short-term and long-term forecast horizons (ranging from one month to twelve months) and use monthly observations in the period 1980–2018. Our forecast models include threshold models that are based on different threshold variables.

We provide a rich set of empirical results. In addition to the forecast performance comparison of threshold and linear models we investigate the threshold variables and explanatory variables that imply "best" models, the structural pattern of evaluation criteria across different regimes, and best sector-specific forecast performance. For instance, we observe that in threshold models with volatility in equity markets defining the states of the economy, loss measures seem to perform better in times of low volatility (than in times of high volatility) while profit-based measures seem to perform better in times of high volatility (than in times of low volatility). This observation is most pronounced for the overall GSCI commodity index and for the two sub-indices energy and industrial metals. Our results suggest that an interesting trade-off appears between loss and profit measures, which implies that the particular aim of the prediction exercise carried out plays a very important role in terms of defining which set of models is the best to use. The optimal specifications for applications, where the metrics for success are related to systematically predicting the direction of change of commodity prices accurately, may thus be systematically different from those aimed at providing point predictions with an absolute minimal distance to the realized values.

Exploiting the potential for improving predictive ability in order to refine the specification and estimation of models may be a potentially fruitful avenue of future research. In particular, entertaining estimation methods that differ from least squares (and thus do not build on the minimization of in-sample squared errors) or Bayesian methods with suitable prior specifications could lead to further improvements in the prediction of commodity prices. Enlarging the set of possible models to account for nonlinearities to include smooth transition in the parameters appears also as a natural next step that builds upon the results presented in this study.

Figure 3: Best threshold model for TV other than dependent variable versus best threshold model with TV=dep

The graphs show a comparison of the best threshold model for a given threshold variable other than the dependent variable with the best threshold model with the threshold variable being the dependent variable. The numbers indicate the number of threshold variables where the best model for this threshold variable outperforms the best exciting threshold model, where the threshold variables is in levels or differences (both for TV=dep and TV=other TV). The maximum number possible is nine.

		MAE	MSE	DA	DV	return
all	1m	3	4	8	5	6
	3m	8	9	9	9	9
	6m	8	9	9	9	8
	12m	8	9	9	9	9
energy	1m	5	2	9	4	6
	3m	8	6	9	7	8
	6m	9	5	3	9	8
	12m	9	6	1	8	1
industrial metals	1m	7	9	8	4	2
	3m	9	8	2	5	4
	6m	4	9	2	3	1
	12m	2	2	6	6	7
precious metals	1m	8	8	1	1	3
	3m	8	9	5	1	0
	6m	0	2	3	0	2
	12m	4	4	3	1	3
agriculture	1m	5	8	5	2	2
	3m	8	7	7	6	6
	6m	8	9	7	7	3
	12m	7	4	7	3	4
livestock	1m	2	1	3	6	8
	3m	1	1	1	0	2
	6m	2	2	1	3	5
	12m	6	4	2	5	9

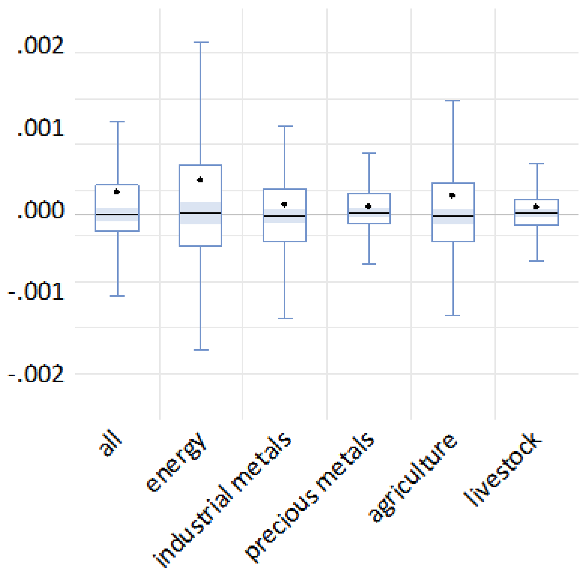
Table 4: Performance of best threshold model, TV=spread in levels or differences, and of corresponding linear model for the aggregate index.
The four-digit combination of ones and zeros below the model shows the inclusion (1) of the explanatory variables CLI, REER, stock index, and oil stock-to-use ratio, and spread (Δspread) indicates that the threshold variable is in levels (differences). Petrol shading indicates that the best threshold model outperforms the best linear model. Light petrol shading shows better total performance between best threshold model and corresponding linear model. Red shading indicates better regime-based performance between best threshold model and corresponding linear model. Regime 1 is defined by $(\Delta)\text{spread}_{t-k} \leq \gamma$ while regime 2 is defined by $(\Delta)\text{spread}_{t-k} > \gamma$.

		MAE	MSE	DA	DV	return
1-month horizon		TDVAR(1,8)	TDVAR(1,5)	TDVAR(3,2)	TDVAR(1,1)	TDVAR(1,1)
		1100, spread	1110, spread	1000, spread	1100, spread	1100, spread
threshold	total	4.06	0.28	68.45	76.85	29.75
	regime 1	4.59	0.23	68.97	80.42	34.98
	regime 2	3.84	0.29	68.18	51.11	2.06
linear	total	4.21	0.30	63.69	68.19	20.97
	regime 1	4.83	0.27	65.52	70.72	24.85
	regime 2	3.94	0.31	62.73	49.91	-0.07
3-months horizon		TDVAR(1,12)	TDVAR(1,12)	TDVAR(2,1)	TDVAR(2,5)	TDVAR(2,5)
		1000, spread	1000, spread	1110, spread	1110, spread	1110, spread
threshold	total	8.70	1.46	67.86	78.34	20.70
	regime 1	15.16	4.58	66.67	72.50	17.98
	regime 2	7.83	1.04	75.00	80.20	21.34
linear	total	9.22	1.71	61.90	56.16	9.72
	regime 1	16.74	6.18	59.03	61.31	9.51
	regime 2	8.20	1.11	79.17	54.52	9.77
6-months horizon		TDVAR(3,2)	TVAR(2,12)	TDVAR(2,2)	TDVAR(2,5)	TDVAR(2,2)
		1110, Δspread	0010, spread	1010, spread	1100, Δspread	1010, spread
threshold	total	13.24	4.24	67.26	78.88	13.87
	regime 1	25.47	11.94	65.71	74.20	10.74
	regime 2	11.50	3.25	67.67	94.00	14.71
linear	total	14.43	4.76	64.88	63.34	10.27
	regime 1	25.29	11.48	62.86	62.64	4.41
	regime 2	12.88	3.90	65.41	65.61	11.84
12-months horizon		TDVAR(2,2)	TVAR(2,10)	TVAR(2,1)	TVAR(2,3)	TVAR(2,1)
		1101, Δspread	1100, spread	1010, spread	1010, Δspread	1100, spread
threshold	total	20.13	7.85	68.45	75.07	8.87
	regime 1	24.29	6.81	66.67	65.88	9.67
	regime 2	19.54	8.13	79.17	79.91	4.04
linear	total	22.35	10.82	54.76	58.38	-0.27
	regime 1	24.80	11.14	51.39	43.37	-1.18
	regime 2	22.00	10.73	75.00	66.29	5.13

Figure 4: Best threshold model versus best linear model
The graphs show a comparison of the best threshold model (for a given threshold variable, including the dependent variable) with the best linear model. The numbers indicate how many of the best threshold models outperform the best linear model, where the threshold variable is in levels or differences. The maximum possible number is ten.

		MAE	MSE	DA	DV	return
all	1m	10	10	9	10	4
	3m	10	10	10	10	10
	6m	10	10	7	10	10
	12m	8	10	10	10	10
energy	1m	10	10	10	10	10
	3m	10	10	10	10	10
	6m	10	10	10	10	10
	12m	10	10	10	10	10
industrial metals	1m	10	10	10	10	10
	3m	10	10	10	10	10
	6m	10	10	10	10	10
	12m	10	10	9	9	8
precious metals	1m	10	9	10	10	10
	3m	10	10	10	10	10
	6m	10	10	10	10	10
	12m	10	10	10	10	10
agriculture	1m	9	8	10	7	6
	3m	10	10	10	10	10
	6m	10	10	10	10	10
	12m	10	10	10	10	10
livestock	1m	5	10	10	10	10
	3m	10	10	10	10	10
	6m	10	10	10	10	10
	12m	10	10	10	10	10

Figure 5: Difference between MSE for best linear and best threshold models
The graphs show boxplots of the differences between the MSE for the best linear model and the MSE for the best threshold model, for forecast horizons of 1 (left) and 12 (right) months. The differences are taken such that a higher mass in the positive region (or a positive mean) indicates a better threshold model.



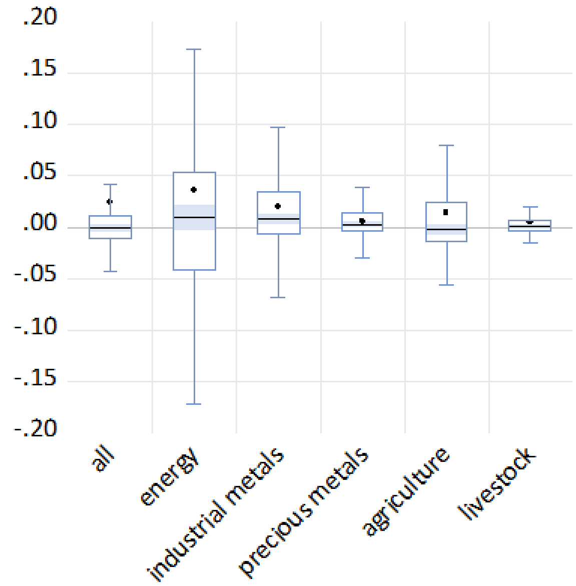


Figure 6: Difference between MAE, MSE, DV and return for best linear and best threshold models for the energy sector
The graph shows boxplots of the differences between the MAE, MSE, DV and return for the best linear and best threshold models, for the energy sector and a forecast horizon of 12 months. The differences are taken such that for each measure a higher mass in the positive region (or a positive mean) indicates a better threshold model. The DV was divided by 100 in order to be better comparable to the other measures.

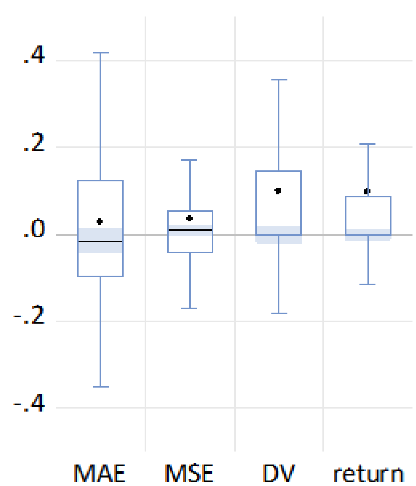


Table 5: Deviation of average threshold model from best threshold model divided by deviation of best linear model from best threshold model
Each number is calculated as the average deviation of a threshold model (across different threshold variables) from the best threshold model divided by the deviation of the best linear model from the best threshold model. Deviations are taken in absolute values, so the numbers are always positive. Note that the best threshold model is always better than the best linear model, the average threshold model, however, may be worse than the best linear model (implied by a ratio larger than one). The smaller the ratio the better the average threshold model compared to the best linear model. Threshold variables are in levels or differences. Light petrol shading indicates smallest deviation of average threshold model compared to best linear model, red shading indicates largest deviation, across commodity sectors.

		MAE	MSE	DA	DV	return
1-month horizon	all	0.39	0.44	0.80	0.38	1.15
	energy	0.38	0.44	0.46	0.52	0.54
	industrial metals	0.40	0.33	0.68	0.36	0.32
	precious metals	0.42	0.50	0.50	0.44	0.43
	agriculture	0.70	0.62	0.41	0.92	0.85
	livestock	1.16	0.66	0.58	0.63	0.56
3-months horizon	all	0.28	0.48	0.40	0.33	0.44
	energy	0.41	0.38	0.42	0.26	0.21
	industrial metals	0.47	0.40	0.41	0.51	0.51
	precious metals	0.67	0.48	0.45	0.40	0.40
	agriculture	0.58	0.39	0.24	0.34	0.28
	livestock	0.68	0.71	0.43	0.42	0.31
6-months horizon	all	0.59	0.51	1.04	0.34	0.37
	energy	0.46	0.47	0.49	0.37	0.50
	industrial metals	0.59	0.57	0.43	0.46	0.55
	precious metals	0.45	0.35	0.55	0.35	0.50
	agriculture	0.67	0.54	0.38	0.34	0.34
	livestock	0.52	0.70	0.58	0.36	0.43
12-months horizon	all	0.70	0.58	0.32	0.30	0.38
	energy	0.59	0.43	0.25	0.23	0.15
	industrial metals	0.54	0.53	0.57	0.48	0.50
	precious metals	0.30	0.38	0.41	0.35	0.55
	agriculture	0.47	0.42	0.43	0.39	0.51
	livestock	0.36	0.48	0.33	0.28	0.42

Table 6: Performance of best threshold model for threshold variable being volatility in levels and of corresponding linear model for the aggregate index.
The four-digit combination of ones and zeros below the model shows the inclusion (1) of the explanatory variables CLI, REER, stock index, and oil stock-to-use ratio. Petrol shading indicates that the best threshold model outperforms the best linear model. Light petrol shading shows better total performance between best threshold model and corresponding linear model. Red shading indicates better performance between the two regimes for the best threshold model. Regime 1 is defined by $\text{VOLA}_{t-k} \leq \gamma$ while regime 2 is defined by $\text{VOLA}_{t-k} > \gamma$.

		MAE	MSE	DA	DV	return
1-month horizon		TDVAR(1,2)	TDVAR(1,4)	TDVAR(3,4)	TVAR(3,4)	TDVAR(3,4)
		1001	1100	1100	1100	1100
threshold	total	4.10	0.29	69.05	75.86	28.63
	regime 1	3.87	0.26	68.49	75.73	25.83
	regime 2	4.57	0.50	72.73	76.49	48.66
linear	total	4.23	0.30	64.88	67.60	24.44
	regime 1	3.82	0.29	65.07	65.26	22.59
	regime 2	5.07	0.42	63.64	78.52	37.40
3-months horizon		TDVAR(2,4)	TDVAR(2,9)	TDVAR(3,12)	TDVAR(1,4)	TDVAR(1,4)
		1000	1011	1000	1010	1010
threshold	total	8.88	1.55	66.67	75.20	19.05
	regime 1	8.63	1.23	65.52	71.83	15.63
	regime 2	10.64	3.74	73.91	91.35	45.15
linear	total	9.15	1.85	63.10	65.01	13.33
	regime 1	8.88	1.32	64.14	59.17	8.94
	regime 2	11.00	5.54	56.52	92.99	47.90
6-months horizon		TDVAR(3,10)	TDVAR(2,9)	TDVAR(3,4)	TDVAR(2,11)	TDVAR(2,4)
		1100	1011	1000	1110	1100
threshold	total	14.23	4.43	67.86	74.11	11.27
	regime 1	13.11	3.82	65.99	68.15	6.51
	regime 2	19.17	8.68	80.95	96.80	45.51
linear	total	14.46	4.92	66.67	64.02	9.67
	regime 1	13.09	4.22	65.99	66.70	7.37
	regime 2	20.54	9.83	71.43	53.80	25.53
12-months horizon		TDVAR(3,6)	TVAR(3,11)	TVAR(2,4)	TVAR(2,4)	TVAR(2,4)
		1000	0101	1010	1010	1010
threshold	total	21.14	8.36	69.64	75.42	9.34
	regime 1	18.10	7.47	67.35	68.84	6.95
	regime 2	39.35	11.41	85.71	88.44	26.03
linear	total	21.65	12.70	54.76	58.38	3.77
	regime 1	18.08	8.46	52.38	54.45	2.51
	regime 2	43.06	27.20	71.43	66.15	12.58

Figure 7: Best threshold variables according to MSE and return
The graph indicates which threshold variables yield the best (1), second best (2), and so on, to the worst (9) performance according to MSE and return. The threshold variable is in levels or differences.

		MSE									return								
		VOLA	CLI	CCI	IR	INFL	OIL	spread	COR	COR_oil	VOLA	CLI	CCI	IR	INFL	OIL	spread	COR	COR_oil
all	1m	5	3	6	8	4	9	1	2	7	3	1	9	7	4	5	2	6	8
	3m	7	2	4	1	5	9	3	8	6	5	2	4	6	3	8	1	9	7
	6m	6	4	3	1	2	9	8	7	5	8	2	1	4	7	9	5	6	3
	12m	4	2	1	5	3	7	6	8	9	8	1	2	5	3	4	9	7	6
energy	1m	7	4	9	3	8	5	1	6	2	2	6	8	9	4	7	5	3	1
	3m	8	2	4	1	6	9	5	7	3	4	8	3	1	9	6	7	2	5
	6m	5	3	4	1	2	8	7	9	6	3	4	1	8	7	9	2	5	6
	12m	5	1	3	2	4	7	6	9	8	7	5	9	6	1	8	4	2	3
industrial metals	1m	8	3	6	7	1	9	3	2	5	7	2	8	9	4	3	6	1	5
	3m	5	2	8	1	4	7	9	3	6	7	8	6	3	9	1	4	5	2
	6m	4	1	6	2	9	5	8	3	7	7	3	5	1	9	2	4	6	8
	12m	2	3	7	1	6	4	5	8	9	1	7	4	3	9	5	6	8	2
precious metals	1m	9	7	6	4	1	8	5	2	2	8	2	7	5	3	1	4	9	6
	3m	3	9	7	8	2	4	5	1	6	6	2	5	3	9	1	8	4	7
	6m	2	5	9	3	4	7	6	1	8	7	3	5	1	8	2	6	4	9
	12m	3	5	7	1	9	2	6	4	8	9	8	6	1	5	2	4	3	7
agriculture	1m	8	5	3	4	6	9	2	1	7	5	7	4	2	6	8	1	3	9
	3m	6	7	2	4	5	8	3	1	9	9	6	4	5	2	7	8	1	3
	6m	6	8	1	3	7	5	4	2	9	5	8	2	4	9	7	6	3	1
	12m	8	7	1	4	9	6	5	2	3	5	9	1	7	4	8	6	2	3
livestock	1m	5	9	4	8	5	5	1	2	2	6	7	8	3	9	5	4	1	2
	3m	2	4	8	9	3	7	5	1	6	9	2	5	3	6	4	8	1	7
	6m	8	5	3	2	9	4	6	1	7	9	1	8	7	3	5	6	2	4
	12m	8	2	7	4	6	5	9	1	3	9	5	6	7	8	1	4	3	2

Figure 8: Inclusion of explanatory variables in best threshold model
The graphs shows the number of times a given explanatory variable (CLI, REER, stock, stu, GSCI aggregate) is included in the best TM (aggregated over the 10 different threshold variables), where the TV is in levels or differences. The maximum number possible is ten.

		CLI					REER					stock					stu					GSCI aggregate				
		MAE	MSE	DA	DV	return	MAE	MSE	DA	DV	return	MAE	MSE	DA	DV	return	MAE	MSE	DA	DV	return	MAE	MSE	DA	DV	return
all	1m	10	10	10	9	10	2	5	3	5	2	1	3	3	1	2	4	2	1	4	4					
	3m	9	8	10	10	10	1	1	6	3	4	3	3	8	6	5	1	2	3	4	4					
	6m	4	6	10	8	10	6	3	4	3	1	4	3	6	5	5	2	4	2	2	1					
	12m	4	6	8	10	10	8	5	1	4	5	3	2	3	2	2	3	4	1	1	1					
energy	1m	10	9	10	10	10	4	4	3	4	4	5	4	2	4	4	2	1	4	5	5	1	2	4	1	2
	3m	8	7	10	9	10	3	2	5	3	3	6	4	5	2	3	2	1	6	6	6	5	4	6	7	4
	6m	0	6	9	8	10	3	6	6	4	3	6	3	9	6	4	1	2	3	3	4	5	5	2	2	1
	12m	2	3	8	9	7	5	5	0	3	1	0	0	4	1	2	7	3	0	0	1	2	2	3	3	4
industrial metal	1m	10	10	10	10	10	5	3	2	4	3	2	3	7	4	4	0	1	3	5	4	1	2	7	5	2
	3m	10	10	10	10	10	2	2	4	1	1	2	3	3	4	5	0	0	3	0	0	1	1	5	3	3
	6m	10	10	10	10	10	4	5	3	5	4	5	4	5	5	4	0	1	1	1	1	2	4	3	3	3
	12m	8	8	9	10	10	7	5	8	1	3	6	5	2	5	4	0	2	0	2	3	6	8	7	4	6
precious metals	1m	0	3	8	4	2	4	1	3	8	4	9	7	2	4	5	2	3	8	5	6	4	3	1	3	2
	3m	1	1	8	7	4	6	5	5	5	6	7	2	3	3	3	3	4	7	9	6	4	4	2	4	3
	6m	0	2	3	4	4	6	4	7	5	5	6	3	2	2	2	5	4	6	5	3	7	9	4	3	4
	12m	4	4	2	6	3	5	5	10	6	7	5	5	4	3	5	5	5	3	4	3	7	7	1	1	2
agriculture	1m	3	7	4	4	4	7	7	5	3	4	5	1	8	7	7	3	1	3	2	2	1	5	2	6	6
	3m	2	7	9	9	10	6	4	6	5	6	7	7	8	8	6	2	2	5	7	8	1	5	1	6	5
	6m	4	4	5	4	6	6	5	5	4	6	6	9	7	5	7	3	5	6	9	7	3	7	2	4	6
	12m	6	7	8	7	7	5	3	5	3	4	5	5	8	4	6	8	5	6	4	3	4	6	7	5	6
livestock	1m	6	5	5	8	6	4	6	6	6	5	2	2	6	6	6	4	3	6	5	6	5	6	4	6	5
	3m	7	7	10	8	9	0	1	5	2	3	8	10	10	10	10	0	1	4	5	3	2	0	6	3	4
	6m	9	8	9	8	9	1	1	3	1	1	8	10	10	10	10	2	2	2	1	2	1	0	2	0	1
	12m	4	5	7	6	6	2	3	2	1	1	10	10	8	9	10	3	3	3	3	1	2	1	1	1	0

Table 7: Performance of best threshold model across GSCI sectors.

		MAE	MSE	DA	DV	return
$h = 1$	best TM	2.42	0.10	72.62	79.53	46.74
	TV	ΔCOR	Δspread	Δspread	VOLA	ΔCOR-oil
	sector	livestock	livestock	industrial met.	energy	energy
$h = 3$	best TM	4.84	0.38	76.79	88.57	31.67
	TV	COR	COR	IR	Δoil	Δoil
	sector	livestock	livestock	industrial met.	industrial met.	industrial met.
$h = 6$	best TM	6.58	0.64	78.57	86.62	24.26
	TV	IR	COR	CLI	ΔIR	ΔIR
	sector	livestock	livestock	industrial met.	industrial met.	industrial met.
$h = 12$	best TM	9.84	1.55	80.95	89.41	17.01
	TV	CLI	ΔCOR	CCI	CCI	VOLA
	sector	livestock	livestock	agriculture	agriculture	industrial met.

Figure 9: Returns and MSE of best threshold models for different GSCI sectors
The graph shows the returns (left) and MSE (right) of best threshold models for different GSCI sectors and different forecast horizons, where the threshold variable is in levels or differences.

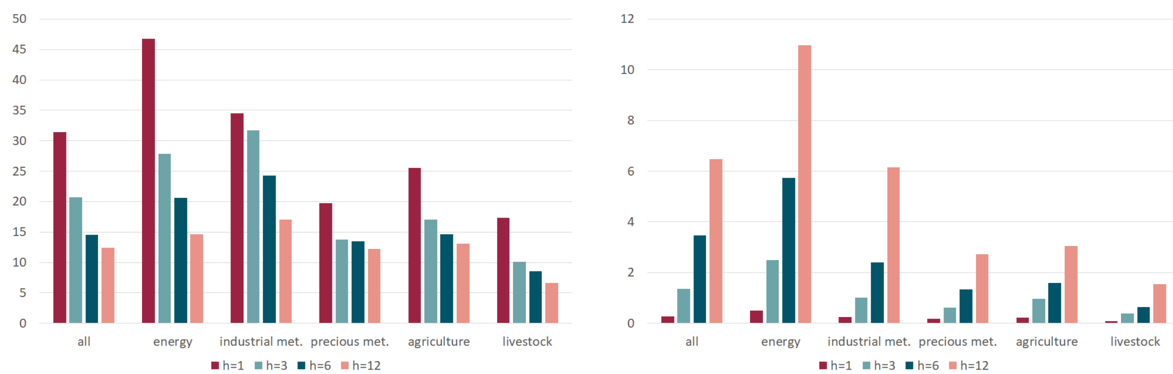


Table 8: Performance of best threshold models for different GSCI sectors
The table shows the forecast performance of best threshold models for different GSCI sectors and different forecast horizons, where the TV is in levels or differences. Light petrol shading indicates best performance across GSCI sectors, red shading indicates worst performance.

		MAE	MSE	DA	DV	return
1-month horizon	all	4.00	0.28	70.24	76.92	31.46
	energy	5.39	0.50	69.05	79.53	46.74
	industrial metals	3.72	0.26	72.62	78.89	34.47
	precious metals	3.23	0.17	66.67	72.90	19.75
	agriculture	3.56	0.23	67.86	76.38	25.54
	livestock	2.42	0.10	71.43	76.93	17.30
3-months horizon	all	8.70	1.37	69.05	78.56	20.70
	energy	11.48	2.50	71.43	78.83	27.81
	industrial metals	7.42	1.02	76.79	88.57	31.67
	precious metals	5.93	0.61	70.24	76.25	13.82
	agriculture	7.25	0.97	68.45	77.47	17.07
	livestock	4.84	0.38	67.26	74.01	10.10
6-months horizon	all	13.24	3.46	70.24	82.01	14.58
	energy	17.83	5.74	69.05	83.41	20.58
	industrial metals	11.57	2.40	78.57	86.62	24.26
	precious metals	8.94	1.35	75.00	79.80	13.50
	agriculture	9.56	1.60	73.21	81.26	14.66
	livestock	6.58	0.64	72.02	78.43	8.59
12-months horizon	all	19.27	6.47	71.43	83.62	12.41
	energy	25.02	10.97	70.83	82.14	14.67
	industrial metals	18.05	6.16	78.57	84.91	17.01
	precious metals	13.10	2.73	76.19	85.10	12.20
	agriculture	13.41	3.04	80.95	89.41	13.12
	livestock	9.84	1.55	73.21	80.60	6.64

Figure 10: Loss and profit measures for different GSCI sectors
The graphs show the MAE, MSE, DA, DV and return (left to right, top to bottom) for different GSCI sectors and different forecast horizons.

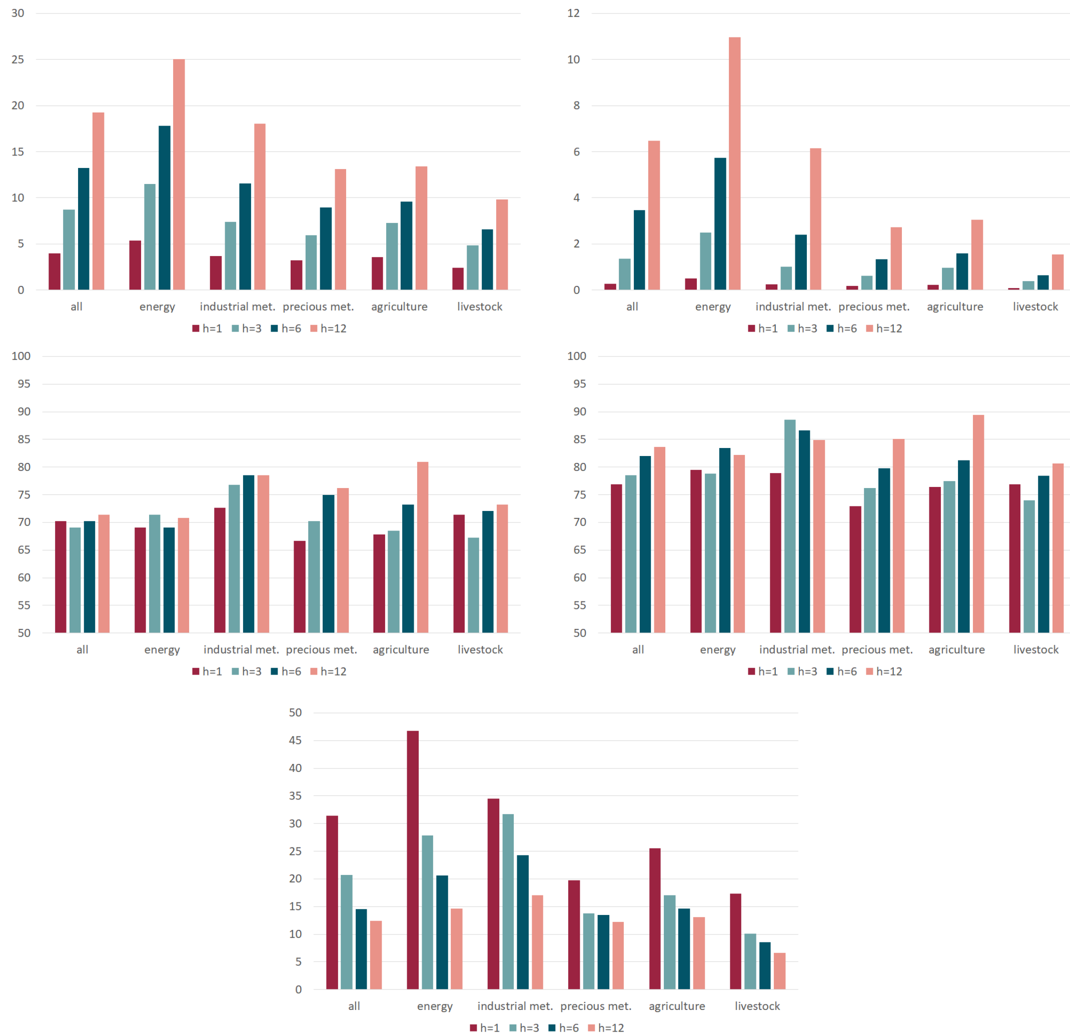


Table 9: Best models in smaller and larger class of models

The table shows the performance criteria of best models in "TMs versus linear models" and of best models in "TMs versus a larger class of models" for different GSCI sectors and different forecast horizons, where the TV is in levels or differences. The best model in the smaller class of models (left panel) is always better than, or at least as good as, the best model in the larger class of models (right panel). In the smaller class of models best models are always threshold models, in the larger class of models, in nine out of the total of 144 cases the best model is not a threshold model. Light petrol shading indicates the cases when the best model is not a threshold model.

		TMs versus linear models					TMs versus larger class of models					
		MAE	MSE	DA	DV	return	MAE	MSE	DA	DV	return	TP
$h = 1$	all	4.00	0.28	70.24	76.92	31.46	4.07	0.28	67.86	74.91	27.44	20.25
	energy	5.39	0.50	69.05	79.53	46.74	5.61	0.54	64.88	75.08	37.62	24.38
	industrial met.	3.72	0.26	72.62	78.89	34.47	3.76	0.27	69.05	77.85	32.23	26.85
	precious met.	3.23	0.17	66.67	72.90	19.75	3.36	0.18	60.71	66.96	14.59	33.74
	agriculture	3.56	0.23	67.86	76.38	25.54	3.70	0.25	66.07	73.30	22.17	33.13
	livestock	2.42	0.10	71.43	76.93	17.30	2.43	0.10	66.67	74.38	15.77	33.56
$h = 3$	all	8.70	1.37	69.05	78.56	20.70	8.70	1.46	66.07	78.34	20.70	20.48
	energy	11.48	2.50	71.43	78.83	27.81	12.03	2.69	69.05	75.30	26.38	25.84
	industrial met.	7.42	1.02	76.79	88.57	31.67	7.59	1.19	72.02	80.89	25.75	28.38
	precious met.	5.93	0.61	70.24	76.25	13.82	6.32	0.65	65.48	69.48	11.85	25.51
	agriculture	7.25	0.97	68.45	77.47	17.07	7.77	1.06	64.29	72.28	14.29	28.57
	livestock	4.84	0.38	67.26	74.01	10.10	5.28	0.45	63.69	70.07	8.29	29.35
$h = 6$	all	13.24	3.46	70.24	82.01	14.58	13.96	4.27	67.26	76.63	12.18	20.97
	energy	17.83	5.74	69.05	83.41	20.58	18.33	7.20	67.86	76.91	16.57	23.68
	industrial met.	11.57	2.40	78.57	86.62	24.26	12.44	3.21	76.19	83.61	22.32	25.58
	precious met.	8.94	1.35	75.00	79.80	13.50	9.37	1.45	72.02	73.13	11.23	30.16
	agriculture	9.56	1.60	73.21	81.26	14.66	10.81	2.08	67.86	75.56	12.36	28.07
	livestock	6.58	0.64	72.02	78.43	8.59	6.88	0.85	63.69	68.86	5.76	25.00
$h = 12$	all	19.27	6.47	71.43	83.62	12.41	20.44	8.42	66.07	73.30	7.34	35.00
	energy	25.02	10.97	70.83	82.14	14.67	26.92	12.66	65.48	72.98	11.75	36.36
	industrial met.	18.05	6.16	78.57	84.91	17.01	20.73	7.47	76.19	82.65	15.20	39.39
	precious met.	13.10	2.73	76.19	85.10	12.20	14.18	3.09	73.21	75.50	10.67	18.61
	agriculture	13.41	3.04	80.95	89.41	13.12	15.57	3.87	75.00	79.12	9.71	29.55
	livestock	9.84	1.55	73.21	80.60	6.64	10.25	1.79	67.26	72.78	5.17	31.58

Table 10: Deviation between best models in class with linear models and best models in larger class of models
The table shows the difference (in percentage points) between the forecast performance of best models in "TMs versus linear models" and the forecast performance of best models in "TMs versus a larger class of models" for different GSCI sectors and different forecast horizons, where the TV is in levels or differences. In the smaller class of models best models are always threshold models, in the larger class of models, in nine out of the 120 cases shown in this table the best model is not a threshold model. The difference is presented such that a positive value implies a better best model in the class with linear models. Thus the best model in the class with linear models is always better or at least as good as the best model in the larger class of models. Light petrol shading indicates the cases when the best model in the larger class of model is not a threshold model.

		MAE	MSE	DA	DV	return
1-month horizon	all	0.08	0.00	2.38	2.01	4.02
	energy	0.22	0.04	4.17	4.45	9.12
	industrial metals	0.04	0.01	3.57	1.04	2.24
	precious metals	0.13	0.01	5.96	5.94	5.16
	agriculture	0.14	0.02	1.79	3.08	3.37
	livestock	0.01	0.00	4.76	2.55	1.53
3-months horizon	all	0.00	0.09	2.98	0.22	0.00
	energy	0.55	0.19	2.38	3.53	1.43
	industrial metals	0.17	0.17	4.77	7.68	5.92
	precious metals	0.39	0.04	4.76	6.77	1.97
	agriculture	0.52	0.09	4.16	5.19	2.78
	livestock	0.44	0.07	3.57	3.94	1.81
6-months horizon	all	0.72	0.81	2.98	5.38	2.40
	energy	0.50	1.46	1.19	6.50	4.01
	industrial metals	0.87	0.81	2.38	3.01	1.94
	precious metals	0.43	0.10	2.98	6.67	2.27
	agriculture	1.25	0.48	5.35	5.70	2.30
	livestock	0.30	0.21	8.33	9.57	2.83
12-months horizon	all	1.17	1.95	5.36	10.32	5.07
	energy	1.90	1.69	5.35	9.16	2.92
	industrial metals	2.68	1.31	2.38	2.26	1.81
	precious metals	1.08	0.36	2.98	9.60	1.53
	agriculture	2.16	0.83	5.95	10.29	3.41
	livestock	0.41	0.24	5.95	7.82	1.47

Appendix A: Data description

Table 11: Contracts included in the S&P GSCI in 2019
(RPDW = Reference Percentage Dollar Weight, see S&P Dow Jones, 2019.)

Commodity	Trading facility	2019 RPDW	Sector
Chicago Wheat	CBT	2.77%	agriculture
Kansas Wheat	KBT	1.15%	agriculture
Corn	CBT	4.36%	agriculture
Soybeans	CBT	3.14%	agriculture
Coffee	ICE - US	0.72%	agriculture
Sugar	ICE - US	1.54%	agriculture
Cocoa	ICE - US	0.32%	agriculture
Cotton	ICE - US	1.41%	agriculture
Lean Hogs	CME	1.91%	agriculture
Live Cattle	CME	3.48%	agriculture
Feeder Cattle	CME	1.27%	agriculture
WTI Crude Oil	NYM / ICE	26.42%	energy
Heating Oil	NYM	4.45%	energy
RBOB Gasoline	NYM	4.48%	energy
Brent Crude Oil	ICE - UK	18.61%	energy
Gasoil	ICE - UK	5.56%	energy
Natural Gas	NYM / ICE	3.11%	industrial metals
Aluminum	LME	3.89%	industrial metals
Copper	LME	4.45%	industrial metals
Nickel	LME	0.76%	industrial metals
Lead	LME	0.78%	industrial metals
Zinc	LME	1.28%	industrial metals
Gold	CMX	3.72%	precious metals
Silver	CMX	0.42%	precious metals

Table 12: Data Description and Sources

TR = Thomson Reuters, DS = Datastream, own calc. = own calculations, S&P = Standard and Poors, GSCI = Goldman Sachs Commodity Index, b = barrel, d = day, SD = standard deviation, BLS = Bureau of Labor Statistics, pp = percentage points. All variables with the unit "index" are indexed at 2000:1=100. The volatility is calculated as the standard deviation of the daily returns in a given month, annualized. The correlations are calculated between returns in the respective markets over the last 65 days ($\sim$ 3 months) or over the last 130 days ($\sim$ 6 months), recorded at the end of a given month. If daily data are available for a given variable the monthly values are computed as the averages of the daily values in a given month.

Abbreviation	Variable	Unit	Note	Source	Code	Start date	Frequency
Commodity							
GSCI	S&P GSCI	Index	Total return index	TR DS: S&P	GSCITOT	1980:1	m
GSCI-energy	S&P GSCI Energy	Index	Total return index	TR DS: S&P	GSENTOT	1982:12	m
GSCI-industrial	S&P GSCI Industrial Metals	Index	Total return index	TR DS: S&P	GSINTOT	1980:1	m
GSCI-precious	S&P GSCI Precious Metals	Index	Total return index	TR DS: S&P	GSPMTOT	1980:1	m
GSCI-agri	S&P GSCI Agriculture	Index	Total return index	TR DS: S&P	GSAGTOT	1980:1	m
GSCI-live	S&P GSCI Livestock	Index	Total return index	TR DS: S&P	GSLVTOT	1980:1	m
GSCI-agri-live	S&P GSCI Agriculture & Livestock	Index	Total return index	TR DS: S&P	GSALTOT	1980:1	m
GSCI-non-energy	S&P GSCI Non-Energy	Index	Total return index	TR DS: S&P	GSNETOT	1980:1	m
explanatory variables: macro/finance							
CLI	US Composite Leading Indicator	Index	Amplitude adjusted, seasonally adjusted	TR DS: OECD	USOL2000Q	1980:1	m
REER	US real effective exchange rate	Index		TR DS: OECD	USOCC011	1980:1	m
stock	world stock market index	Index		TR DS: DS	TOTMKWD	1980:1:1	d
explanatory variables: fundamental							
oil-stu	oil stock-to-use ratio, total world	ratio	linear interpolation from annual	own calc., OPEC		1980:1	m (a)
wheat-stu	US wheat stock-to-use ratio	%	linear interpolation from annual	USDA (FAS)		1980:1	a
meat-stu	US meat stock-to-use ratio	%	lin interp from annual, meat: beef & veal	USDA (FAS)		1980:1	a
Threshold variables							
CLI	US Composite Leading Indicator	Index	Amplitude adjusted, seasonally adjusted	TR DS: OECD	USOL2000Q	1980:1	m
CCI	US consumer confidence index	Index	Seasonally adjusted	TR DS: Conference Board	USCNFCONQ	1980:1	m
VOLA	US Stock Market Volatility	%	SD of daily STOCK returns in one month, ann	own calc., TR DS		1980:1	m
COR	Cor betw. US stock & bond markets, 6m	Cor	Correlation between stock and bond, 6m	own calc., TR DS		1980:6	m
COR-oil	Cor betw. world stock & oil markets, 6m	Cor	Correlation between stock and oil, 6m	own calc., TR DS		1980:6	m
oil	oil price (Brent)	USD/b	Crude Oil BFO M1 Europe FOB $/BBl, Brent	TR DS: TR	OILBREN	1980:1	m
INF	US inflation (consumer price index)	%	All urban sample: all items	TR DS: BLS	USCPANNL	1980:1	m
IR	US interbank rate, 3 months	%		TR DS: Reuters	USINTER3	1980:1	m
spread	diff. betw. long- and short-term US int. rates	pp	IR-long minus IR	own calc., TR DS		1980:1	m
Auxiliary variables							
stock	US stock market index	Index	S&P 500	TR DS: S&P	S&PCOMP	1980:1:1	d
bond	US government bond market index	Index	US tracker all Lives DS government index	TR DS: DS	TUSGVAL(RI)	1980:1:1	d
stock	world stock market index	Index		TR DS: DS	TOTMKWD	1980:1:1	d
IR-long	US treasury constant maturity, 10 years	%		TR DS: US Fed	FRTCM10	1980:1:1	d

Appendix B: Tables for a larger number of models

Table 13: Summary of forecast performance of best models for the GSCI aggregate index. stu represents the oil stock-to-us ratio for the total world.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
GSCI	4.07	0.28	67.86	74.91	27.44	0.37	20.25
regime 1: $z_{t-k} \leq \gamma$	5.06	0.44	63.41	62.52			
regime 2: $z_{t-k} > \gamma$	3.94	0.26	69.29	69.74			
	TDVAR(1)	TDVAR(1)	TVAR(2)	TVAR(2)	s-DVAR(2)	s-DVAR(2)	TVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	
			REER	REER	REER	REER	REER
			stock	stock			stock
				stu	stu	stu	stu
threshold variable z_t	spred_{t-12}	spred_{t-12}	ΔINF_{t-9}	ΔINF_{t-9}			Δoil_{t-11}
3-months horizon	MAE	MSE	DA	DV	return	SR	TP
GSCI	8.70	1.46	66.07	78.34	20.70	0.43	20.48
regime 1: $z_{t-k} \leq \gamma$	15.16	4.58	68.75	70.84	17.98	0.41	
regime 2: $z_{t-k} > \gamma$	7.83	1.04	65.44	77.70	21.34	0.43	
	TDVAR(1)	TDVAR(1)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
			REER	REER	REER	REER	
			stock	stock	stock	stock	
							stu
threshold variable z_t	spred_{t-12}	spred_{t-12}	spread_{t-5}	spred_{t-5}	spred_{t-5}	spred_{t-5}	Δoil_{t-11}
6-months horizon	MAE	MSE	DA	DV	return	SR	TP
GSCI	13.96	4.27	67.26	76.63	12.18	0.35	20.97
regime 1: $z_{t-k} \leq \gamma$	26.78	3.01		84.70	10.84	0.27	
regime 2: $z_{t-k} > \gamma$	10.95	7.59		64.71	12.50	0.37	
	TAR(3)	TVAR(2)	DVAR(3)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TVAR(3)
			CLI	CLI	CLI	CLI	CLI
			REER	REER	REER	REER	
		stock	stock	stock	stock	stock	
							stu
threshold variable z_t	ΔVOLA_{t-1}	IR_{t-7}		CCI_{t-7}	spread_{t-5}	spread_{t-5}	Δoil_{t-11}
12-months horizon	MAE	MSE	DA	DV	return	SR	TP
GSCI	20.44	8.42	66.07	73.30	7.34	0.30	35.00
regime 1: $z_{t-k} \leq \gamma$	25.74	9.57	77.46	63.88	10.84	0.54	
regime 2: $z_{t-k} > \gamma$	19.19	3.15	57.73	80.22	-7.12	-0.54	
	TDAR(2)	TAR(2)	TDVAR(2)	TVAR(3)	TVAR(2)	TVAR(2)	TVAR(3)
			CLI				CLI
				REER			REER
			stock	stock	stock	stock	stock
			stu				stu
threshold variable z_t	Δoil_{t-1}	INF_{t-1}	oil_{t-12}	VOLA_{t-11}	IR_{t-12}	IR_{t-12}	COR-oil_{t-9}

Table 14: Summary of forecast performance of best models for the GSCI energy index. stu represents the oil stock-to-us ratio for the total world.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
energy	5.61	0.54	64.88	75.08	37.62	0.37	24.38
regime 1: $z_{t-k} \leq \gamma$	6.14	0.78	63.19	66.82	31.75	0.33	
regime 2: $z_{t-k} > \gamma$	4.93	0.47	75.00	74.10	78.15	0.61	
	TDVAR(2)	TDVAR(2)	TDVAR(3)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
	REER		REER	REER	REER	REER	
				stock	stock	stock	
		stu					stu
	GSCI	GSCI		GSCI	GSCI	GSCI	GSCI
threshold variable z_t	ΔCLI_{t-11}	ΔCLI_{t-11}	VOLA_{t-4}	VOLA_{t-4}	VOLA_{t-4}	VOLA_{t-4}	ΔIR_{t-12}
3-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
energy	12.03	2.69	69.05	75.30	26.38	0.41	25.84
regime 1: $z_{t-k} \leq \gamma$	11.60	3.76	66.67	82.27	28.59	0.32	
regime 2: $z_{t-k} > \gamma$	15.04	2.36	69.14	66.15	26.30	0.42	
	TDVAR(2)	TDVAR(2)	TVAR(3)	TDVAR(2)	TVAR(3)	TVAR(3)	TVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
			REER		REER	REER	REER
			stock		stock	stock	
		stu	stu		stu	stu	stu
		GSCI					
threshold variable z_t	VOLA_{t-4}	ΔCLI_{t-11}	oil_{t-3}	CCI_{t-7}	oil_{t-3}	oil_{t-3}	Δoil_{t-6}
6-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
energy	18.33	7.20	67.86	76.91	16.57	0.38	23.68
regime 1: $z_{t-k} \leq \gamma$		7.44	67.35	83.77	12.51	0.32	
regime 2: $z_{t-k} > \gamma$		7.12	71.43	62.78	46.98	0.71	
	s-DVAR(2)	TVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)
	CLI		CLI	CLI	CLI	CLI	CLI
	REER						
		stock					stock
	stu						stu
			GSCI		GSCI	GSCI	GSCI
threshold variable z_t		ΔCCI_{t-12}	VOLA_{t-4}	CCI_{t-7}	VOLA_{t-4}	VOLA_{t-4}	oil_{t-3}
12-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
energy	26.92	12.66	65.48	72.98	11.75	0.39	36.36
regime 1: $z_{t-k} \leq \gamma$	28.17	15.21	50.91	74.33	5.59	0.20	
regime 2: $z_{t-k} > \gamma$	21.14	12.29	72.57	42.96	14.74	0.48	
	TVAR(3)	TVAR(3)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TVAR(3)
		CLI	CLI	CLI	CLI	CLI	CLI
	REER	REER					REER
			stock	stock	stock	stock	stock
							stu
				GSCI			
threshold variable z_t	INF_{t-5}	$\Delta\text{spread}_{t-2}$	CLI_{t-8}	IR_{t-1}	CLI_{t-8}	CLI_{t-8}	COR-oil_{t-9}

Table 15: Summary of forecast performance of best models for the GSCI industrial metals index. stu represents the oil stock-to-us ratio for the total world.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
industrial metals	3.76	0.27	69.05	77.85	32.23	0.45	26.85
regime 1: $z_{t-k} \leq \gamma$	3.73	0.20	70.73	75.21	34.49	0.47	
regime 2: $z_{t-k} > \gamma$	3.77	0.31	64.44	73.26	26.30	0.38	
	TDVAR(1)	TDVAR(1)	TVAR(3)	TVAR(3)	TVAR(3)	TVAR(3)	TVAR(2)
	CLI	CLI	CLI	CLI	CLI	CLI	
	REER		REER	REER	REER	REER	REER
	stock						
			stu	stu	stu	stu	
			GSCI	GSCI	GSCI	GSCI	GSCI
threshold variable z_t	ΔCCI$_{t-6}$	INF$_{t-7}$	COR-oil$_{t-1}$	COR-oil$_{t-1}$	COR-oil$_{t-1}$	COR-oil$_{t-1}$	ΔVOLA$_{t-4}$
3-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
industrial metals	7.59	1.19	72.02	80.89	25.75	0.54	28.38
regime 1: $z_{t-k} \leq \gamma$	6.35	0.63	68.83		14.73	0.39	
regime 2: $z_{t-k} > \gamma$	7.52	1.25	71.22		27.59	0.58	
	TDVAR(2)	TDVAR(2)	TVAR(3)	VEC(3,1)	TDVAR(2)	TDVAR(2)	TDVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	
			REER	REER			REER
	stock	stock		stock	stock	stock	
							stu
				GSCI			GSCI
threshold variable z_t	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	COR-oil$_{t-1}$		ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	VOLA$_{t-7}$
6-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
industrial metals	12.44	3.21	76.19	83.61	22.32	0.56	25.58
regime 1: $z_{t-k} \leq \gamma$	10.92	2.71	75.41	75.22	17.67	0.56	
regime 2: $z_{t-k} > \gamma$	13.30	3.49	76.63	81.73	25.01	0.63	
	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TDVAR(2)	TVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
	stock	stock					stock
							stu
threshold variable z_t	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	ΔCOR$_{t-6}$	COR-oil$_{t-1}$
12-months horizon	**MAE**	**MSE**	**DA**	**DV**	**return**	**SR**	**TP**
industrial metals	20.73	7.47	76.19	82.65	15.20	0.52	39.39
regime 1: $z_{t-k} \leq \gamma$	18.98	6.53	85.71	73.42	11.18	0.47	
regime 2: $z_{t-k} > \gamma$	23.89	9.17	74.29	76.06	32.03	0.74	
	TDVAR(1)	TDVAR(1)	TVAR(2)	TVAR(2)	TVAR(2)	TVAR(2)	TVAR(3)
			CLI	CLI	CLI	CLI	CLI
							REER
	stock	stock			stock	stock	stock
	stu	stu					stu
		GSCI					GSCI
threshold variable z_t	ΔIR$_{t-6}$	ΔIR$_{t-6}$	ΔCOR-oil$_{t-2}$	ΔCOR-oil$_{t-2}$	ΔIR$_{t-6}$	ΔIR$_{t-6}$	COR-oil$_{t-1}$

Table 16: Summary of forecast performance of best models for the GSCI precious metals index. stu represents the oil stock-to-us ratio for the total world.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
precious metals	3.36	0.18	60.71	66.96	14.59	0.27	33.74
regime 1: $z_{t-k} \leq \gamma$	3.69	0.31	64.52		13.76	0.29	
regime 2: $z_{t-k} > \gamma$	3.06	0.15	56.00		16.73	0.25	
	TDVAR(2)	TDVAR(1)	TVAR(1)	BDVAR(2)	TDVAR(1)	TDVAR(1)	TDVAR(3)
			CLI				CLI
	REER		REER	REER	REER	REER	REER
		stock	stock	stock			stock
			stu		stu	stu	stu
		GSCI					
threshold variable z_t	ΔCLI_{t-12}	spread_{t-9}	COR_{t-11}		INF_{t-9}	INF_{t-9}	Δoil_{t-6}
3-months horizon	MAE	MSE	DA	DV	return	SR	TP
precious metals	6.32	0.65	65.48	69.48	11.85	0.36	25.51
regime 1: $z_{t-k} \leq \gamma$	6.44	0.61	63.22	76.73	17.12	0.54	
regime 2: $z_{t-k} > \gamma$	6.18	0.84	67.90	63.36	3.47	0.10	
	TDVAR(3)	TDAR(1)	TDVAR(1)	TVAR(3)	TDVAR(3)	TDVAR(3)	TDVAR(2)
			CLI				
	REER			REER			
				stock	stock	stock	stock
			stu	stu			stu
threshold variable z_t	ΔINF_{t-11}	CLI_{t-1}	ΔIR_{t-11}	CLI_{t-12}	oil_{t-10}	oil_{t-10}	Δ CLI_{t-12}
6-months horizon	MAE	MSE	DA	DV	return	SR	TP
precious metals	9.37	1.45	72.02	73.13	11.23	0.46	30.16
regime 1: $z_{t-k} \leq \gamma$	14.60	1.27	78.10	76.09	16.91	0.72	
regime 2: $z_{t-k} > \gamma$	7.99	2.47	61.90	64.62	2.09	0.09	
	TDVAR(3)	TDAR(1)	TDVAR(1)	TVAR(3)	TDVAR(1)	TDVAR(1)	TDVAR(3)
	REER		REER	REER	REER	REER	
	stock			stock			stock
				stu			stu
	GSCI		GSCI		GSCI	GSCI	GSCI
threshold variable z_t	COR_{t-9}	CLI_{t-1}	oil_{t-10}	CLI_{t-12}	oil_{t-10}	oil_{t-10}	CCI_{t-6}
12-months horizon	MAE	MSE	DA	DV	return	SR	TP
precious metals	14.18	3.09	73.21	75.50	10.67	0.60	18.61
regime 1: $z_{t-k} \leq \gamma$	11.54	2.87	84.76	72.10	16.79	1.08	
regime 2: $z_{t-k} > \gamma$	18.58	3.50	53.97	73.00	0.46	0.03	
	TDVAR(1)	TDVAR(2)	TDVAR(3)	TVAR(3)	TDVAR(1)	TDVAR(1)	TVAR(3)
							CLI
	REER	REER	REER				REER
		stock					
				stu			
	GSCI	GSCI			GSCI	GSCI	GSCI
threshold variable z_t	oil_{t-10}	Δoil_{t-11}	oil_{t-10}	CLI_{t-11}	oil_{t-10}	oil_{t-10}	oil_{t-2}

Table 17: Summary of forecast performance of best models for the GSCI agriculture index. stu represents the US wheat stock-to-us ratio.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
agriculture	3.70	0.25	66.07	73.30	22.17	0.34	33.13
regime 1: $z_{t-k} \leq \gamma$	3.69	0.28	70.83	68.32	23.19	0.35	
regime 2: $z_{t-k} > \gamma$	3.70	0.22	64.17	73.41	20.39	0.32	
	TDVAR(1)	TDVAR(1)	TVAR(2)	TDVAR(1)	TDVAR(1)	TDVAR(1)	TVAR(2)
		REER	REER				REER
		stock	stock	stock	stock	stock	stock
			stu				stu
	GSCI	GSCI		GSCI	GSCI	GSCI	
threshold variable z_t	ΔIR_{t-11}	ΔIR_{t-11}	ΔCCI_{t-11}	ΔIR_{t-11}	ΔIR_{t-11}	ΔIR_{t-11}	ΔCCI_{t-11}
3-months horizon	MAE	MSE	DA	DV	return	SR	TP
agriculture	7.77	1.06	64.29	72.28	14.29	0.34	28.57
regime 1: $z_{t-k} \leq \gamma$	12.64	2.40	55.56	91.50	57.60	0.96	
regime 2: $z_{t-k} > \gamma$	7.11	0.87	66.67	67.05	9.21	0.24	
	TVAR(2)	TVAR(2)	TVAR(1)	TVAR(2)	TVAR(2)	TVAR(2)	TVAR(3)
	CLI	CLI		CLI	CLI	CLI	CLI
			REER	REER	REER	REER	
	stock	stock		stock	stock	stock	
			stu				
			GSCI	GSCI			
threshold variable z_t	spread_{t-12}	spread_{t-12}	CCI_{t-11}	spread_{t-12}	spread_{t-12}	spread_{t-12}	Δ spread_{t-11}
6-months horizon	MAE	MSE	DA	DV	return	SR	TP
agriculture	10.81	2.08	67.86	75.56	12.36	0.41	28.07
regime 1: $z_{t-k} \leq \gamma$	11.31	4.92	55.26	80.33	26.66	0.32	
regime 2: $z_{t-k} > \gamma$	10.38	1.69	71.54	62.39	8.61	0.69	
	TVAR(2)	TVAR(2)	TDVAR(1)	TVAR(1)	TVAR(1)	TVAR(1)	TDVAR(2)
		CLI					
			REER				
	stock	stock		stock	stock	stock	
	stu			stu	stu	stu	stu
			GSCI	GSCI	GSCI	GSCI	GSCI
threshold variable z_t	ΔCOR_{t-8}	spread_{t-12}	CCI_{t-12}	CCI_{t-12}	CCI_{t-12}	CCI_{t-12}	CLI_{t-12}
12-months horizon	MAE	MSE	DA	DV	return	SR	TP
agriculture	15.57	3.87	75.00	79.12	9.71	0.49	29.55
regime 1: $z_{t-k} \leq \gamma$	19.55	3.02					
regime 2: $z_{t-k} > \gamma$	13.63	4.06					
	TDVAR(2)	TAR(2)	VEC(3,1)	VEC(3,1)	VEC(3,1)	VEC(3,1)	TVAR(2)
			CLI	CLI	CLI	CLI	
							stock
	stu		stu	stu	stu	stu	
			GSCI	GSCI	GSCI	GSCI	GSCI
threshold variable z_t	ΔIR_{t-11}	COR_{t-1}					Δoil_{t-12}

Table 18: Summary of forecast performance of best models for the GSCI livestock index. stu represents the US meat stock-to-us ratio.

1-month horizon	MAE	MSE	DA	DV	return	SR	TP
livestock	2.43	0.10	66.67	74.38	15.77	0.39	33.56
regime 1: $z_{t-k} \leq \gamma$	2.72	0.10	69.23	67.62	9.33	0.20	
regime 2: $z_{t-k} > \gamma$	2.37	0.10	66.20	71.99	16.98	0.43	
	TDVAR(1)	TDVAR(1)	TVAR(2)	TDVAR(3)	TDVAR(3)	TDVAR(3)	TDVAR(3)
	CLI	CLI		CLI	CLI	CLI	CLI
	REER	REER	REER				REER
							stock
	stu	stu					
	GSCI	GSCI	GSCI	GSCI	GSCI	GSCI	GSCI
threshold variable z_t	Δspread$_{t-8}$	Δspread$_{t-8}$	Δspread$_{t-8}$	Δspread$_{t-8}$	Δspread$_{t-8}$	Δspread$_{t-8}$	ΔVOLA$_{t-5}$
3-months horizon	MAE	MSE	DA	DV	return	SR	TP
livestock	5.28	0.45	63.69	70.07	8.29	0.32	29.35
regime 1: $z_{t-k} \leq \gamma$	6.64	0.46		64.73	8.17	0.30	
regime 2: $z_{t-k} > \gamma$	5.07	0.45		56.39	8.46	0.33	
	TDVAR(3)	TDVAR(1)	VEC(3,1)	TVAR(3)	TVAR(3)	TVAR(3)	TDVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	
		REER	REER				REER
				stock	stock	stock	
							stu
		GSCI					
threshold variable z_t	oil$_{t-9}$	Δspread$_{t-8}$		COR$_{t-7}$	COR$_{t-7}$	COR$_{t-7}$	IR$_{t-1}$
6-months horizon	MAE	MSE	DA	DV	return	SR	TP
livestock	6.88	0.85	63.69	68.86	5.76	0.33	25.00
regime 1: $z_{t-k} \leq \gamma$	7.25	0.96	59.60	62.16	6.61	0.36	
regime 2: $z_{t-k} > \gamma$	6.83	0.84	69.57	56.14	4.55	0.27	
	TDVAR(3)	TDVAR(3)	TVAR(3)	TVAR(3)	TVAR(3)	TVAR(3)	TDVAR(3)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
							REER
			stock	stock	stock	stock	
							stu
			GSCI				
threshold variable z_t	oil$_{t-9}$	oil$_{t-9}$	COR$_{t-7}$	COR$_{t-7}$	COR$_{t-7}$	COR$_{t-7}$	Δspread$_{t-8}$
12-months horizon	MAE	MSE	DA	DV	return	SR	TP
livestock	10.25	1.79	67.26	72.78	5.17	0.41	31.58
regime 1: $z_{t-k} \leq \gamma$	9.39	2.27	55.00	67.89	4.85	0.35	
regime 2: $z_{t-k} > \gamma$	10.38	1.09	68.92	63.80	5.62	0.55	
	TDVAR(3)	TVAR(3)	TVAR(2)	TVAR(3)	TVAR(3)	TVAR(3)	TVAR(2)
	CLI	CLI	CLI	CLI	CLI	CLI	CLI
							REER
		stock	stock	stock	stock	stock	stock
			stu				stu
			GSCI				
threshold variable z_t	oil$_{t-9}$	COR$_{t-7}$	ΔCLI$_{t-8}$	COR$_{t-7}$	COR$_{t-7}$	COR$_{t-7}$	oil$_{t-8}$

References

[1] H. Ahumada, M. Cornejo. 2015. Explaining commodity prices by a cointegrated time series-cross section model. *Empirical Economics*, 48, 1667–1690.

[2] H. Ahumada, M. Cornejo. 2016. Forecasting food prices: The case of corn, soybeans and wheat. *International Journal of Forecasting*, 32, 838–848.

[3] J.T. Bernard, L. Khalaf, M. Kichian, S. McMahon. 2008. Forecasting commodity prices: GARCH, jumps, and mean reversion. *Journal of Forecasting*, 27, 279–291.

[4] Y. Chen, K. Rogoff, B. Rossi. 2010. Can exchange rates forecast commodity prices? *Quarterly Journal of Economics*, 125, 1145–1194.

[5] M. Costantini, J. Crespo Cuaresma, J. Hlouskova. 2016. Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *Journal of Forecasting*, 35, 652–668.

[6] J. Crespo Cuaresma, I. Fortin, J. Hlouskova. 2018. Exchange rate forecasting and the performance of currency portfolios. *Journal of Forecasting*, 37, 519–540.

[7] T. Dangl, M. Helling. 2012. Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106, 157–181.

[8] S. Degiannakis, G. Filis, T. Klein, T. Walther. 2020. Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*, forthcoming.

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

[10] R. Giacomini, B. Rossi. 2010. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25, 595–620.

[11] J.J.J. Groen, P.A. Pesenti. 2011. Commodity prices, commodity currencies, and global economic developments. In T. Ito, A. Rose (Eds.). *Commodity prices and markets, NBER East Asia seminar on economics: Vol. 20*, Chicago: Chicago University Press, 15–42.

[12] M. Guidolin, A. Timmermann. 2005. Economic implications of bull and bear regimes in UK stock and bond returns. *Economic Journal*, 115, 111–143.

[13] M. Guidolin, A. Timmermann. 2009. Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150, 297–319.

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

[15] B. Jacobsen, B.R. Marshall, N. Visaltanachoti. 2019. Stock market predictability and industrial metal returns. *Management Science*, 65, 3026–3042.

[16] R.E. Just, G.C. Rausser. 1981. Commodity price forecasting with large-scale econometric models and the futures market. *American Journal of Agricultural Economics*, 63, 197–208.

[17] O.A. Ramirez, M. Fadiga. 2003. Forecasting agricultural commodity prices with asymmetric-error GARCH models. *Journal of Agricultural and Resource Economics*, 28, 71–85.

[18] S&P Dow Jones, 2019. S&P GSCI Methodology. S&P Dow Jones Indices: Index Methodology.

[19] S&P Dow Jones, 2018. Continued Petroleum Sector Strength as S&P Dow Jones Indices Announces 2019 S&P GSCI Weights, press release, 1 November 2018, New York.

[20] X. Xu. 2017. Short-run price forecast performance of individual and composite models for 496 corn cash markets. *Journal of Applied Statistics*, 44, 2593–2620.

[21] X. Xu. 2018. Using local information to improve short-run corn price forecasts. *Journal of Agricultural & Food Industrial Organization*, 16, 1–15.

[22] X. Xu. 2020. Corn cash price forecasting. *American Journal of Agricultural Economics*, 102, 1297–1320.