Energy Reports 15 (2026) 108941

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

"One Out of Many" consolidating a long-term trend forecast for investing in energy commodities

F. Díaz-Rodríguez [a,*], M.-D. Robles [b,c]

[a] *Universidad Complutense de Madrid, Spain*
[b] *Department of Economic Analysis, Universidad Complutense de Madrid, Spain*
[c] *Instituto Complutense de Análisis Económico, Universidad Complutense de Madrid, Spain*

A R T I C L E I N F O

Keywords:
Forecasts combination
Trend estimation
Crude oil
Long-term forecasting
Trading strategies

A B S T R A C T

Accurately predicting the unobservable long-term trend of energy commodity prices is vital for economic and financial stability, yet single forecasts are often unreliable. This study proposes and evaluates a combined forecasting approach to consolidate a single, robust trend estimate for Crude Oil (CL) prices. First, we generate five individual trend estimates using diverse methods: a Cubic Polynomial, Symmetric Moving Average, the Hodrick-Prescott Filter, the Maximum Overlap Discrete Wavelet Transform (MODWT), and the Kalman filter estimation of the Schwartz and Smith (2000) model. Second, we combine these estimates using seven distinct algorithms: simple average, median, weighted average, Multi-Linear Regression (MLR), Principal Components Analysis (PCA), a Feed-Forward Neural Network (FFNN), and a novel modified Multi-Ensemble Time-Scale (METS 2.0) algorithm designed to filter high-frequency noise. We find that combined forecasts consistently outperform individual estimates, with METS 2.0 algorithm demonstrating the best performance by minimizing errors and remaining stable across volatile market conditions. This is validated by long-term investment strategies, where trend-guided approaches significantly outperform those based on price forecasts. This key result confirms the value of isolating the underlying trend for long-term decisions.

"*One Out of Many*" *refers to* "*E Pluribus Unum,*" *the motto from the first Congress of the United States (August 20, 1776), and is dated before Christ as* "*E Pluribus Unus*" *in Virgil's poem* "*Moretum*" *and* "*Unus fiat ex Pluribus*" *in Cicero's treatise* "*De Officiis*"*. Here, we build* "*One*" *from* "*Out of Many*" *different forecasts.*

1. Introduction

Understanding and monitoring energy commodity price trends are essential due to the pivotal role of commodity prices in shaping economic dynamics, financial markets, and policy decisions. The trend, the main underlying component of prices capturing the prevailing price tendency that will persist in the long term, is essentially *unobservable.* This condition makes trend forecasting extremely challenging, as well as selecting the best among several plausible candidates.

Despite the many alternatives available for energy sources, Crude Oil (CL) is still a key driver for the economy, industry, and geopolitics (Provornaya et al., 2020; Ding et al., 2022; Wei et al., 2023); thus, it also impacts public policy decision-making and is a critical reference in the commodities and financial markets. Wei et al. (2023) find that a shock in CL demand is the most important factor impacting the variance of the stock market returns and volatility in the U.S. and China. Hence, the importance of providing a reliable forecast of CL prices but also the challenge of forecasting such a volatile variable (Ding et al., 2022) with a long-term scope.

The main literature agrees on the utility of combining forecasts instead of identifying a single "best" forecast. Combined forecasts or "meta-predictions" (Vilar et al., 2020) provide more accurate, stable, and less biased results, thus diversifying the risk of large errors, a lousy specification, or modelling uncertainty in any individual forecast. These meta-predictions would be especially relevant when the variable to be forecasted is unobservable, as in the case of the trend component of a time series.

This paper presents a combined forecast approach to consolidate one long-term trend estimate for CL prices and assess its performance in trend-following investment strategies. We focus on forecasting the trend

* Corresponding author.
E-mail addresses: fediaz01@ucm.es, fernanda.diaz.rodriguez@gmail.com (F. Díaz-Rodríguez), mdrobles@ucm.es (M.-D. Robles).

https://doi.org/10.1016/j.egyr.2025.108941
Received 13 June 2025; Received in revised form 9 November 2025; Accepted 21 December 2025
Available online 30 January 2026
2352-4847/

instead of the price for two main reasons. First, long-term decisions, such as project investments, will benefit from price movements that persist over time. Public policy could also benefit from the trend, as it provides longer-term information. Second, by focusing on the trend, we avoid the effect of short-term noise contained in prices, thus reducing the impact of volatility and temporary market conditions on forecasting accuracy. In this paper, we combine trend estimates to provide a long-term forecast, an approach that is original to the best of our knowledge.

The ultimate purpose is to determine the most appropriate method for estimating the trend or, in other words, forecasting the price in the long run. With this objective in mind, we hypothesise that combinations of forecasts provide better results than individual ones. To test this, we compare the results of investment strategies based on the predicted trends from: (1) a set of individual trend estimation methods: Cubic Polynomial, Symmetric Moving Average (MA), Hodrick-Prescott Filter (HP), Maximum Overlap Wavelet Discrete Transform (MODWT) and the Kalman filter estimation of the Schwartz and Smith model (SS2000); and (2) from seven forecasts combination methods: two straightforward methods (the simple average and the median); two standard statistical methods (weighted average and Multi-Linear Regression, MLR), two machine-learning methods (Principal Components Analysis, PCA, and Feed Forward Neural Network, FFNN); and lastly, we also propose to apply a modified version of the Multi-Ensemble Time-Scale algorithm (METS 2.0), which is a novel contribution to the forecasts combination literature (see Table A1 in the Appendix for a list of acronyms and the names of the trends estimated with each method).

As expected, the individual forecasts are highly correlated, but they exhibit low correlations across their short-term deviations from prices. This indicates that each collects different features of the actual trend. Hence, forecast combinations may benefit from such complementarity. Thus, we assess all twelve forecasted trends (five individual and seven combined) using two approaches: the first involves analyzing how well they fit prices, and the second determines which forecasts perform better in trend-following trading strategies across several horizons.

The rest of the paper is organised as follows: Section 2 reviews the literature on forecast combinations. Section 3 presents the data and the individual forecasts. Section 4 outlines the methods for forecasting combinations. Section 5 presents the main results of individual and combined trend forecasts, examining how they perform when used as indicators in trend-following trading strategies. Section 6 discusses the implications of these results, and we provide our concluding remarks in Section 7.

2. Literature review

Since forecasts combination was first proposed (Bates and Granger, 1969), there has been increased support for the general improvement in accuracy it can provide over individual time series forecasting methods (Lemke and Gabrys, 2010; Cang and Yu, 2014; Hsiao and Wan, 2014; Blanc and Setzer, 2016; Chan and Pauwels, 2018; Wang et al., 2018, 2022; Herawati and Djunaidy, 2020; Kourentzes et al., 2019; Vilar et al., 2020; Ding et al., 2022). The discussion revolves mainly around (1) the pre-forecast data processing with the purpose of either denoising the time series (Shabri and Samsudin, 2014; Wang et al., 2022), breaking it down into different scales (Wang et al., 2018; Herawati and Djunaidy, 2020), or both (Saghi and Jahangoshai Rezaee, 2021); (2) the selection of forecasts to be combined (Cang and Yu, 2014; Kourentzes et al., 2019); and (3) the "forecast combination puzzle" (Hsiao and Wan, 2014; Chan and Pauwels, 2018; Blanc and Setzer, 2016; Denk and Löffler, 2024).

2.1. Data processing and hybrid models

According to Lu et al. (2021), forecast and data-cleaning methods have been a *research hotspot* in the past decade. They observe data-cleaning methods like wavelet transform (WT) or other signal processing techniques help improve energy price prediction accuracy. Wang et al. (2018), Herawati and Djunaidy (2020), and Saghi and Jahangoshai Rezaee, (2021) also find an improvement in accuracy when combining signal-processing and forecasting methods. Herawati and Djunaidy (2020) accommodate the nonlinearity and non-stationarity of CL prices in an Empirical Mode Decomposition (EMD) Ensemble that overcomes EMD weakness of scale-separation and Intrinsic Mode Functions (IMF) mixing and feeding the resulting components into a feed-forward neural network to obtain forecasting models for each of them. Salamai (2023) also performs a mode decomposition of the original CL price before applying a predictive deep-learning framework, including artificial neural networks (ANN).[1] Wang et al. (2018) conclude that hybrid models [decomposing & forecasting] capture changes in trends with better precision than single models and find that wavelet transforms produce the best predictions among other decomposition methods, including EMD. Saghi and Jahangoshai Rezaee, (2021) apply WT and fuzzy transform to denoise and decompose the original series of OPEC CL prices before predicting with a multilayer perceptron of the FFNN sort, among other ANN arrangements. They obtain the best results with the WT-only max-denoised series and FFNN forecasting. Díaz-Rodríguez and Robles (2026) compare the EMD and MODWT methods for the specific purpose of estimating a low-frequency trend for fossil fuels. Even after solving EMD problems, the MODWT approximation is found to track prices better and more consistently across different data frequencies. Wang et al. (2022) use signal processing methods to denoise commodity prices before using various forecasting techniques (individual and combined), resulting in an error reduction and a certain degree of diversity in the forecast models to be combined.

In this study, we combine five individual estimates/forecasts with seven different methods: in one of them, we feed an FFNN with the five estimates, including the MODWT approximation, and in another, we perform a hybrid version of the METS algorithm by Percival and Senior (2012). In addition to the original conception of METS, we WT-denoise the differences or residuals of the estimated trends before performing their MODWT decomposition and observe an improvement in METS estimate. However, contrary to the reviewed literature, we do not find any significant improvement when denoising prices with WT before estimating the long-term trend.

2.2. Forecasts Selection

There is consensus on the role of combinations in improving forecasting accuracy (e.g., Cang and Yu, 2014; Wang et al., 2018; and Wang et al., 2023). The improvement achieved depends on the number of combined forecasts, their precision, the diversity among the methods generating them, the uncertainty of the forecasted variable, the presence of structural changes, and the length of the prediction horizon.

The selection of forecasts to combine is a crucial point. Chan and Pauwels (2018) find that selecting the individual forecast to combine by comparing their mean squared forecast error (MSFE) can lead to a biased selection. Cang and Yu (2014) analyse the choice of individual forecasting models to combine using information theory and find that using the optimal set of methods outperforms the combination of all available techniques. Earlier results of Lemke and Gabrys (2010) demonstrate that a ranking-based combination of methods outperforms model selection approaches. Kourentzes et al. (2019) propose a heuristic to formulate the appropriate pool of forecasts for the combination. The procedure of Wang et al. (2018) eliminates redundant or inferior elements in the selection. Armstrong (2001) recommends having at least five elements

[1] Khormali et al. (2025) highlight the effectiveness of ANNs in a complex reservoir engineering application. They find a high accuracy stemmimg from the ANN's ability to learn the "intricate relationships between input and output variables."

in the combination, while Cang and Yu (2014) consider that the optimal subset is between two and five. Chan and Pauwels (2018) find that the MSFE decreases as the number of forecast elements increases. Jose and Winkler (2008) also find the same result when using simple average combinations, but errors diminish gradually.

In our case, the variable to be forecasted is the unobservable trend component of commodity prices, making it challenging to select the best forecasting methods using standard selection metrics. We follow Armstrong (2001) and Cang and Yu (2014) using five different methods. We select a polynomial method, a moving average, a parametric filter, a signal processing technique, and a commodity pricing model. The procedures vary in nature, providing a diverse set that helps reduce overfitting while preventing the generalization of individual errors. According to Lemke and Gabrys (2010), diversity is beneficial for combinations, but only up to a certain point due to the trade-off between diversity and the accuracy of individual methods.[2]

2.3. Forecast combination puzzle[3]

Chan and Pauwels (2018) are puzzled by how the simple average of forecasts often outperforms individual forecasts and how more complex weighting schemes do not always perform better than simple averages. Vilar et al., 2020 and their references, particularly Claeskens et al. (2016), also arrive at the same conclusion. According to Timmermann (2006), for the simple average to be the optimal combination, the variance of the individual forecasting errors must be the same, as well as their (weak) correlation (e.g., Blanc and Setzer, 2016). Hsiao and Wan (2014) indicate that even if those conditions are not satisfied, the equally weighted combination can still minimise the forecasting error for finite samples if corrected for possible bias of any of the combined elements (Palm and Zellner, 1992) and scale-corrected for potential structural changes. However, Wang et al. (2018) indicate that weighting schemes can achieve a better performance than simple averaging when combining individual estimates with different strengths, and Chan and Pauwels (2018) show that this also happens when the individual forecast has a lower variance than the simple average. Thus, as the number of forecasts increases, their average will eventually outperform all the individual estimates.

According to Blanc and Setzer (2016), weights can fit quite well with the training data and perform bad out-of-sample, mainly when the training sample is small. This made the combination results highly sensitive to individual results when there is significant asymmetry among weights, biasing results even when combining unbiased forecasts. Additionally, structural changes will impact optimal weighting schemes more than equally weighted combinations. Additionally, Chan and Pauwels (2018) note that using MSFE to determine the optimal weights in a combination may lead to inconsistencies, as does the assumption that weights are fixed. Percival and Senior (2012) recognise that there is no optimal set of weights across all time scales and propose the METS weighting scheme. Recently, Ding et al. (2022) perform an equally weighted combination of three machine-learning methods (decision trees) and improve their forecasting accuracy, but do not compare with another combination method. While Denk and Löffler (2024) reappraise the forecasting improvement that more complex combinations may have over a simple mean as it may not be sustained out-of-sample over a long period because of changes in the correlation of the individual forecast errors.

Following the preceding, we explore different methods for combinations that include the simple average and the median but also different weighted schemes based on how individual estimated trends rank in fitting prices and others built by an MLR or METS.

2.4. Forecasting commodity prices

Most of the literature on forecast combinations in commodity markets aims at improving the accuracy of forecasting highly uncertain yet observable variables like commodity prices (Herawati and Djunaidy, 2020; Lu et al., 2021; Saghi and Jahangoshai Rezaee, 2021; Salamai, 2023; Shabri and Samsudin, 2014; Szarek et al., 2020; Tsai et al., 2017; Wang et al., 2018, 2023; Zhang and Wang, 2023) or lithium batteries' state of charge (e.g., Almaita et al., 2022).

In this paper, we deal with the forecasting of an unobservable variable, the time trend of energy commodity prices. We assume that employing forecast combination methods can significantly enhance accuracy, especially in situations with higher uncertainty, as is our case. There is limited research on the effectiveness of combining forecasts when the variable being forecasted is unobservable. We find few studies recognizing explicitly that there are unobserved components (Harvey, 2006), unobserved forecasting errors (Timmermann, 2006), or unobservable predictors (Huang and Lee, 2010). Moreover, our study adds to the forecasting combination literature by addressing this particular case in which the unobservable variable to forecast is not any component but the main component, and not an explanatory but an independent variable.

The recent paper by Conlon et al. (2024) is closely related to ours as they aim to forecast the price of CL. They use four linear combinations of predictors from industry fundamentals, commodity markets, and monetary and economic variables. In our case, we forecast the long-term trend of CL prices by combining five estimates of this unobservable variable assumed to be a predictor of prices in the long term. We use seven combination methods (see *subtable 1.B* of Table 1), including some similar to those used by Conlon et al. (2024), but also a multi-linear regression, a wavelets-based ensemble, and two machine-learning methods.

Lu et al. (2021) find that the vast majority of 171 publications (from 2010 to 2020) on energy price forecasts focus on short-term horizons and only 4 (2 %) have a yearly (long-term) horizon. Most of the studies seem to deliver one-step-ahead forecasts pairing the data frequency with the forecast horizon; that is, if the research uses monthly data, the forecast horizon is 1-month ahead, as in Wang et al. (2018), or daily data for daily forecasts as in Wang et al. (2022). The limited forecast horizon may be due to data-frequency-dependent forecasting methods used in these publications. Another constraint may be the connection between the data frequency and the data sample size, as there seems to be a trade-off between selecting low-frequency data to attain a low-frequency forecast at the cost of robustness due to a smaller sample, which may be the case when using a few annual observations.

Our study adds to the long-term forecast research, stretching the horizon at least 12 periods ahead. In the financial markets, the long-term starts in a one-year horizon, which is the standard as the mid-term is quite discretionary, and that is the purposely 12-month horizon for which the individual trends are estimated. Our trend estimation methods are not data-frequency-dependent, except for SS2000-Kalman, as confirmed by the frequency study of Díaz-Rodríguez and Robles (2026). However, we do not use annual data, but rather monthly data, to obtain more observations in the data training sample and, hence, achieve robust results. Moreover, our forecasting approach using trend estimates instead of prices eludes the usual forecasting power deterioration when stretching the horizon more than one step ahead. In fact, our trend forecasts generally perform better in longer horizons and are pretty insensitive to different market conditions. This is of high importance as the log returns of the CL market are more uncertain or dispersed in longer horizons (see *subfigure A.1.b* of Annex *Figure A.1*). In Section 5, the forecasting capacity of the trend trading signals reduces such uncertainty.

[2] Lemke and Gabrys (2010) describe how to measure this trade-off and other diversity features to choose the best individual method over a combination or vice versa.

[3] According to Chan and Pauwels (2018), the name was coined by Stock and Watson (2004), but the puzzle was first noted by Clemen (1989).

Table 1
Formulae of Forecasting Methods (Trend Estimation and Forecasts Combination) and Statistics Description (Fitting, Forecasting, and Performance).

1.A Trend Estimation Methods

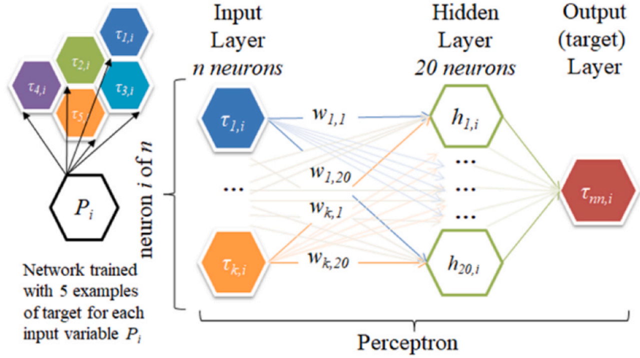
Class	Subclass	Name	Formula	Reference
Statistical	Regression Model	Cubic Polynomial	$\tau_{cubic,i} = f(\ln P_i) = \beta_0 + \beta_1 \ln P_i^3 + \beta_2 \ln P_i^2 + \beta_3 \ln P_i$	
	Moving Average	Symmetric or Centred	$\tau_{symMA,i} = \frac{[0.5(\ln P_{i-6}) + \ln P_{i-5} + \ldots + \ln P_i + \ldots + \ln P_{i+5} + 0.5(\ln P_{i+6})]}{12}$	(Nilsson, 2015).
	Parametric Filter	Hodrick-Prescott	$\tau_{HP,i} =$ $Min_{\{\tau_i\}_{i=1}^{n}} \left\{ \overbrace{\sum_{i=1}^{n} (\ln P_i - \tau_i)^2}^{\text{penalizes the variance of } d_i} + \lambda \overbrace{\sum_{i=2}^{n-1} \underbrace{[(\tau_{i+1} - \tau_i) - (\tau_i - \tau_{i-1})]^2}_{\nabla^2(\tau_{i+1})^2}}^{\text{penalizes the variance of } \tau_i} \right\}$	(Hodrick and Prescott,1997; Maravall and Del Rio, 2001; and Kaiser and Maravall Herrero, 2002).
Signal Processing	Wavelet Transform (WT)	Maximal Overlap Discrete WT (MODWT)	$denoised\{\ln P_{i,j-1}\} = Approx_{i,J} + \sum_{j=1}^{J} D_i$ $\tau_{MODWT,i} = Approx_{i,J}$	(Ramsey & Zhang, 1995; Ramsey & Lampart, 1998b; Connor & Rossiter, 2005; Gençay & Fan, 2006; Crowley, 2007; and Wang et al., 2018).
	Kalman filter	Structural state-space models: SS2000 Two-factor Model	$\ln P_i = \xi_i + \chi_i$ $d\xi_i = \mu_\xi dt + \sigma_\xi dz_\xi.$ $d\chi_i = -\kappa\chi_i dt + \sigma_\chi dz_\chi \bullet$ $\tau_{SS2000,i} = \xi_i$	(Kalman,1960a; Kalman, 1960b, 1960c; Harvey, 1989; and Schwartz and Smith, 2000).

1.B Forecasts Combination Methods

Class	Subclass	Name	Formula	Reference
Statistical Linear Ensemble	Simple (point forecast comb.)	Mean (equally weighted)	$\tau_{mean,i} = \frac{\sum_{i=1}^{n} \sum_{\tau=1}^{k} T}{k}$	(Palm and Zellner, 1992; Timmermann, 2006; Jose and Winkler, 2008; Hsiao and Wan, 2014; Claeskens et al., 2016; Blanc and Setzer, 2016; Chan and Pauwels, 2018; Vilar et al., 2020; and Conlon et al., 2024).
		Median	$\tau_{median,i} = \text{the term} T_i \text{in the} \left\lvert \frac{k+1}{2} \right\rvert th \text{position}$ where $T_i = [\tau_{i,1}, \ldots, \tau_{i,k}]$ and $\tau_{i,1} < \ldots < \tau_{i,k}$	(Vilar et al., 2020; and Conlon et al., 2024)
	Weighted	Performance-based: Minimal Deviations Frequency	$\tau_{weighted,i} = \sum_{i=1}^{n} \sum_{\tau=1}^{k} w_{k,i} \cdot T_{k,i}$ where $w_{T_k,i} = \frac{\sum_{i=1}^{n} \min(\lvert \ln P_i - T_{i,k} \rvert)}{n}$	
		Regression-based: Multi-Linear	$lnP_i = \beta_{0,i} + \beta_{T,i} T_i + \varepsilon_i$ $\tau_{MLR,i} = \beta_{0,i} + \beta_{T,i} T_i$	(Shabri and Samsudin, 2014).
		Optimal w/ signal processing: Multiscale Ensemble Time-Scale (METS)	$d_{METS,i} = Approx_{\tau^{*}_{i,J}} + \sum_{i=1}^{n} \sum_{j=1}^{J} D_{\tau^{*}_{i,j}}$ $\tau_{METS,i} = \ln P_i - d_{METS,i}$ where $\tau^{*}_{i,j} = \min[\text{var}\{T_i\}]_{j=1}^{J}$	(Percival and Senior, 2012).
Machine-learning	Linear	Principal Component Analysis (PCA)	$\begin{bmatrix} \text{var}(\tau_1) & \cdots & \text{cov}(\tau_1, \tau_k) \\ \vdots & \ddots & \vdots \\ \text{cov}(\tau_1, \tau_k) & \cdots & \text{var}(\tau_k) \end{bmatrix}$ where $\rho_{(\tau_1,\tau_k)} = \frac{\text{cov}(\tau_1, \tau_k)}{\sigma_1 \sigma_k}$ and if τ_1 and τ_k are standarized, $\rho_{(\tau_1,\tau_k)} = \text{cov}(\tau_1, \tau_k)$ and $\text{var}(\tau_1) = \text{var}(\tau_k) = 1$ $pc_1 = w_{1,1} \cdot \tau_1 + \ldots + w_{1,k} \cdot \tau_k ; pc_k = w_{k,1} \cdot \tau_1 + \ldots + w_{k,k} \cdot \tau_k PC = T \cdot W \Leftrightarrow T = PC \cdot W^T$ where $W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,k} \\ \vdots & \vdots & \vdots \\ w_{k,1} & \cdots & w_{k,k} \end{bmatrix}$, W is the eigenvectors orthogonal matrix of $T^T \cdot T$ and is a $k \times k$ matrixin which each eigenvector (W_k) is perpendicular to each other andwhose eigenvalues $(w_{k,k})$ are the coordinates of the pc_p and carry the variance weight of each pc_k.Note that there are as many pc as $\tau. \tau_{PCA,i} = PC_{1,i} \cdot W^T_{1,i}$	(Zhang and Wang, 2023).
	Non-linear	Artificial Neural Network: static model feed-forward	$input\ data_i = \{(\ln P_i, T_{k,i})\}$	(Saghi and Jahangoshai Rezaee, 2021).

(*continued on next page*)

Table 1 (*continued*)



$$\text{where}\quad \left.\begin{array}{c} h_{1,i} = w_{1,1}\cdot\tau_{1,i} + \ldots + w_{k,1}\cdot\tau_{k,i} \\ \vdots \\ h_{20,i} = w_{1,20}\cdot\tau_{1,i} + \ldots + w_{k,20}\cdot\tau_{k,i} \end{array}\right\} h_{1,i} + \ldots + h_{20,i} = \tau_{network,i}$$

$$T_i\cdot\begin{bmatrix} w_{1,1} & \cdots & w_{k,1} \\ \vdots & \vdots & \vdots \\ w_{1,20} & \cdots & w_{k,20} \end{bmatrix} = \begin{bmatrix} w_{1,1}\cdot T_{1,i} + \ldots + w_{k,1}\cdot T_{k,i} \\ \vdots \\ w_{1,20}\cdot T_{1,i} + \ldots + w_{k,20}\cdot T_{k,i} \end{bmatrix} = \begin{bmatrix} h_{1,i} \\ \vdots \\ h_{20,i} \end{bmatrix}$$

where: i refers to each of the n month-end dates available in the total sample; $T_i = \{\tau_{cubic,i}, \tau_{symMA,i}, \tau_{HP,i}, \tau_{MODWT,i}, \tau_{SS2000,i}\}$; k is the total number of forecasts in T; j are the levels of decomposition; and J is the lowest level.

1.C Fitting Statistics

Name	Metric	Description
Dstat	Direction	$Dstat = \frac{1}{n}\sum_{i=1}^{n} A_{i,k}$ while $A_{i,k} = 1$ only when $(\ln P_{i+1} - \ln P_i)(\tau_{i+1,k} - \tau_{i,k}) \geq 0$, and $A_{i,k} = 0$ otherwise.
MAE	Mean Absolute Error	$MAE_k = \frac{1}{n}\sum_{i=1}^{n} \lvert \ln P_i - \tau_{i,k}\rvert$
MAPE	Mean Absolute Percentage Error	$MAPE_k = \frac{1}{n}\sum_{i=1}^{n} \left\lvert \frac{\ln P_i - \tau_{i,k}}{\ln P_i}\right\rvert$
MSE	Mean Squared Error	$MSE_k = \frac{1}{n}\sum_{i=1}^{n} (\ln P_i - \tau_{i,k})^2$
RMSE	Root Mean Squared Error	$RMSE_k = \sqrt{\frac{\sum_{i=1}^{n} (\ln P_i - \tau_{i,k})^2}{n}}$
SNR	Signal-to-Noise Ratio	For our particular purpose, we modified the calculus of the SNR to: $-20\log_{10}\cdot norm(\tau_k - \ln P)/norm(\tau_k)$ where $norm(x) = \sqrt{\text{sum}(\text{abs}(x)^2)}$

1.D Forecasting Statistics

Name	Metric	Description
fc	Forecast Capacity	$fc_k = (Ulim_{bmk}^{0.05} - Llim_{bmk}^{0.05})/(6\sigma_k)$
FDstat	Direction	$ForecastDstat = \frac{1}{n}\sum_{i=1}^{n} A_{i,k}$ while $A_{i,k} = 1$ only when $(\ln P_{i+1} - \ln P_i)(\tau_{i,k} - \tau_{i-1,k}) \geq 0$, and $A_{i,k} = 0$ otherwise.
MSFE	Mean Squared Forecasting Error	$MSFE_k = \frac{1}{n}\sum_{i=1}^{n} (\ln P_{i+h} - \tau_{i,k})^2$
R-squared	Out-of-Sample Coefficient of Determination	$R^2_{OS_{k,h}} = 1 - \frac{MSFE\tau_{k,h}}{MSFE\tau_{mean,h}}$

1.E Performance Statistics

Name	Metric	Description
Sharpe Ratio	Risk-adjusted Return	$sr = \frac{(\bar{r}_{\tau_k} - r_f)}{\sigma_{r_{\tau_k}}}$ where rf is the risk-free rate

P = price; τ = trend estimate; k = estimation method; h = forecast horizon; i = observation number; n = total number of observations

3. Data and selected individual forecasting methods

We use a monthly price dataset of first-front-month (Contract 1) NYMEX future CL contract (Cushing, OK CL light sweet grade) sourced by the U.S. Energy Information Agency (EIA.gov) from April 1983 to December 2021. The long time frame of our time series ensures that the price series captures various market conditions, including the odd circumstances of the COVID-19 pandemic and the first negative price in April 2020 (diluted by the monthly average); and the substantial number of observations (465) allows for robustness in the analysis. The spot prices of the West Texas Intermediate CL (WTI light sweet grade, Cushing, OK, FOB) from the same source though perfectly and positively correlated with the NYMEX future, have a shorter history starting in January 1986. The Brent CL spot price can be considered the benchmark for the global CL price due to its delivery location. However, the Brent series available in the EIA is also shorter than the NYMEX future price series and both series are 99.6 % correlated. Moreover, future prices possess characteristics and advantages that make them suitable for the purpose of this study, as noted in Díaz-Rodríguez and Robles (2026). We log-transform the price series to stabilise the variance and improve their statistical properties.[4] Annex *subfigure A.1.a* of *Figure A.1* shows the evolution of the price data series used in both log levels and a USD scale.

In contrast to the literature on commodities estimating the forward curve, modelling the market volatility, or forecasting returns, this is a purely price-based study with price levels as the only market data input in the computations as our targeted variable is the trend, the price long-term level. Our forecasting approach is endogenous; assuming the information to forecast the trend is contained in the prices. We recognize, however, that there are exogenous variables that may not only affect prices in the short-term but change the expected or forecasted course of the trend, their assumed persistent, but not deterministic, component in the long-term. Nonetheless, we take prices as given and do not intend to explain the forecasted trend level, its course, or change of direction by any other market, economic, or geopolitical factor. Results in Section 5 show that some trend forecasts perform well in the long term, even in the presence of external shocks.

We do not rely on the forward curve to forecast long-term prices either. We only use first-front-month (Contract 1) future prices for the reasons mentioned above. But we do not consider longer-term contracts, for example, third-front-month (Contract 3) prices to forecast prices 3-months ahead. The first-front-month (Contract 1) prices pair with spot prices. However, longer-term contracts imply higher basis risk affected by exogenous factors, thus, decoupling the spot and futures series providing spread calendar strategies an opportunity for profit.

Conlon et al. (2024) document that the monthly average is the top choice by most CL price forecasters because it gives more accurate combined forecasts than end of the month data (EoM) data. From our perspective, this may be because monthly average data is already estimating short-term trends. However, the authors cautioned on data series influencing the economic significance of CL price forecasts, specifically monthly average and EoM prices. Díaz-Rodríguez and Robles (2026) also warn about the period covered by each datum across different data frequencies when analysing three energy commodity prices (including CL). A monthly average synthesises daily data within a month, thus continuously covering the same period; hence, the former is a lower-frequency reflection of the latter. We employ a similar series to those used by Conlon et al. (2024) and find that, even in the case of back-testing with EoM returns, our best forecasts remain unchanged, with a performance very similar to that obtained with monthly average data.[5] Thus, we make a complementary note to that of Conlon et al. (2024) on the choice of the combination method and trading strategy, which also influences the results of CL price forecasting.

We deliberately chose a set of diverse methods to estimate trends, providing a mix of complex and parsimonious benchmarks. This comprehensive approach allows us to capture different data features. We follow Díaz-Rodríguez and Robles (2026) in selecting a set of procedures from five families to estimate a continuous, smooth and price-tracking trend of three energy commodities, CL included. We use three statistical methods—Cubic Detrending (polynomial regression), Symmetric (Centred) MA, and the HP filter—, a signal processing method—MODWT approximation—, and the structural model for pricing commodities by Schwartz and Smith (2000).[6]

Subtable 1.A of Table 1 presents these methods according to the classification of Wang et al., 2023. Annex A.1 contains a brief explanation of the methods.[7]

Fig. 1 shows the estimated trends on top and the short-term deviations from the trend on the bottom. The Cubic detrending (τ_{cubic}), the HP filter $\tau_{HP,}$ and the Wavelets (τ_{MODWT}) methods provide the smoothest trends. We observe two big cycles of fifteen to twenty years, each separated by a breakpoint of the market conditions in 2004[8] when CL prices began an expansion race above 40 USD per barrel. The short-term deviations from these trends (lower subfigure in Fig. 1) move around zero synchronically but in different magnitudes. The deviations of the τ_{SS2000} are mainly positive and sizable before the breakpoint and get closer to zero afterwards.

Fig. 2 describes the distribution of the estimated trends. The boxplots in the upper left subfigure and the Kernel distributions in the lower right subfigure show that the estimates provided by the five trend-expert methods present similarities but are not redundant. The lower subfigures show a clear bimodal distribution, describing two poles around which prices move: a low-volatility period characterized by low and stable price levels before the breakpoint and a high-volatility period with higher and more volatile prices afterwards. In Fig. 1, the trend estimated with the symmetric MA (τ_{symMA}) follows prices much closer, but it still moves smoothly in contrast to the one estimated with SS2000 (τ_{SS2000}), which looks relatively rough. Before the breakpoint, it moves quite far below prices and shows a z-saw shape, although, at that point, it softens and moves closer to prices.

Table 2 contains the descriptive statistics of the estimated trend distributions, showing that the SS2000 estimate (τ_{SS2000}) differs most from the other four estimates. While the standard deviation and coefficient of variation of these four estimated trends is lower than the actual prices, as expected, given that trends are not contaminated by short-term noise, the τ_{SS2000} is more volatile than prices.

Table 3 shows the cross-correlation among the log levels (*sub tables 3.1.a* and *3.1.b*) and first difference (*sub tables 3.2.a* and *3.2.b*) of the estimated trends (*sub tables 3.1.a* and *3.2.a*) and the log prices (in between *sub tables 3.1* and *3.2*) and among short-term deviations (*sub tables*

[4] The following references to prices are generic, though we operate and perform computations using log prices (*lnP*). We use the specific *lnP* reference only when needed for clarification, as in the formulae of Table 1, for example.

[5] Results can be available upon request.

[6] Note that two of our individual estimators can be considered as parsimonious, ARIMA-based benchmarks: (1) The Symmetric Moving Average is a classic filter that is a restricted form of an ARIMA model, and (2) the Schwartz and Smith (2000) model's trend, a continuous-time Brownian motion, is discretized as a random walk.

[7] We use "estimate" and "forecast" interchangeably. The difference between the two is very subtle. An estimate is an *a posteriori* optimal approximation of current data with information available today (in-sample). A forecast is an *a priori* approximation of future data when information has not been generated yet (out-of-sample) but is not randomly generated; in such cases, the term "prediction" is more common.

[8] We determine the breakpoint for CL price series in July 2004 using the Cashin and McDermott (2002) method (see *subfigure A.1.c* of Annex *Figure A.1*). This is between April 2006, when the CL's standardized price crossed for the first time the boom threshold according to Jacks (2019) metrics (as shown in *subfigure A.1.d* of Annex *Figure A.1*), and 2003, when the SS2000 parameters estimated by Díaz-Rodríguez and Robles, (2026) indicate a clear breakpoint.

3.1.b and *3.2.b*).

In the shaded upper middle panel, the correlation between prices and the estimates of their main component, the trend, is quite high and significant, as is the correlation among them in the *subtable 3.1.a*, which averages 0.9485 with a significantly low standard deviation of 0.031. According to the *Dstat* statistic (Wang et al., 2022) described in *subtable 1.C* of Table 1, the estimated trends move 98.71 % of the time in the same direction as prices; in fact, always synchronously in the period with high volatility and higher prices. However, in *subtable 3.2.b*, the cross-correlations among the first differences of the trends are quite low, so as the ones between these and those of *lnP* in the shaded lower middle panel. The lowest correlation is between τ_{cubic} and τ_{SS2000} (and theirs with *lnP*) and is insignificant.[9] This is a sign that though trend estimates might move in the same direction most of the time, they do it in quite different magnitude than prices.

In *subtable 3.1.b*, the correlations among the short-term deviations are lower than those among the trend estimates and differ substantially among them, meaning our trend-expert methods are not making the same mistakes, but also pointing to the diversity of the information the estimates contain. The lowest, yet positive, correlation is between τ_{symMA} and τ_{SS2000} deviations and is insignificant. In *subtable 3.2.b*, the first differences of all estimates' deviations (residuals) are highly and significantly intercorrelated, with the exception of those of the SS2000 estimate, meaning that τ_{SS2000} provides most of the diversity to the group of selected trend estimates.

The information conveyed by all the individual trend estimates and their residuals at both levels differ enough to leave room for improving trend forecasting by combining them and seeing what our experts are leaving out individually.

4. Forecasts combination

The "best model" in a forecasting comparison should not be conflated with the "true data-generating process," which remains unobservable. Moreover, when dealing with an unobservable variable, such as in this study, we assume perpetual model incompleteness. Thus, we are agnostic on what model is the best, which is the reason to include the SS2000 model and not a more advanced one, for example. Instead of selecting one from our individual expert models or looking for other experts, we seek to obtain the best forecast by combining all the individual estimated trends with different procedures and letting each decide which forecast to give more weight. We use two simple linear pooling methods (an equally weighted Mean and a Median), three optimal pooling methods (a historical-performance Weighted Average, a Multilinear Regression, MLR, and the novel wavelet-based algorithm Multi-Ensemble Time-Scale, METS), and two machine-learning methods (Principal Components Analysis, PCA, and Artificial Neural Network, ANN). These methods are presented below, organized in a manner that highlights their features or origin. *Subtable 1.B* of Table 1, however, includes a brief description of these methods classified according to Wang et al., 2023.yet ob

4.1. Simple combination methods: mean, median, and weighted average

The individual estimated trends are similar and highly correlated (Tables 2 and 3), so it seems reasonable to consolidate one estimate by their *Mean* with the advantage of stable and predictable equal weights. As we have mentioned previously, some estimates are more similar than others. Thus, we also use the *Median* to combine them, allowing for weighting changes over time in the case of high uncertainty. These measures of central tendency are widely used, as evidenced by Vilar et al., 2020.

We also use a *Weighted Average* setting the weights (w_j) of each of the five trend estimates (τ_j) according to their historical performance, fitting prices. Instead of the common arrangement that depends inversely on their MSE, our scoring system depends directly on the frequency with which each estimate has the minimal absolute difference or deviation from log prices (lnP) calculated at each month-end i across the n observations (see Fig. 3). The trend estimate with the greatest weight will be the one with the best fit to the prices (*lnP*) trajectory; in this case, τ_{symMA} is the top scorer with a frequency of 231 observations, that is, almost 50 % of the time, showing the minimal deviation from prices.

4.2. Adjustment method: MLR

To estimate the "latent trend" from the set of different individual estimates, we combine our trend forecasts using a multi-linear regression as Lu et al. (2021) and Shabri and Samsudin (2014) do. The MLR method aims to estimate the coefficients associated with individual trends (regressors) to achieve the optimal individual trend combination, that is, the one that best fits prices (endogenous variable). We have four similar performers and one quite different (τ_{SS2000}), which provides complementarity more than mere diversity, as τ_{SS2000} would perform better when the other four perform worse.[10] The bottom subfigure in *Annex Figure A.3* shows the MLR trend estimate (τ_{MLR}) in which τ_{symMA} and τ_{SS2000} have the biggest weight in the MLR trend.[11] The first (τ_{symMA}) moves smoothly close to prices throughout the whole sample, and the second (τ_{SS2000}) seems affected by the short-term variations but does not move as close to prices before the breakpoint mentioned in the Data section. The other three estimates (τ_{cubic}, τ_{HP}, and τ_{MODWT}) have a similar evolution (see Fig. 1), capturing only the big cycles; their weight in the regression model is close to zero and reassigned to a constant level instead. The MLR extracts two main features of the estimated trends, which are the components of the τ_{MLR}: a base ground level represented by the constant in the model and the main trajectory of prices recovered by the coefficients of τ_{symMA} (along with some soft mid-term swings) and τ_{SS2000} (plus some shorter-lived swings inherited from the roughness of τ_{SS2000}).

4.3. Machine-learning-based combination methods: PCA and neural networks

We use two ML-based ensemble methods to consider the possibility of nonlinearity and complex interactions across individual trend estimates. First, we employ Principal Component Analysis (PCA) to calculate the latent trend. PCA helps to identify the significant co-movements among multiple variables, thereby eliminating much of the individual idiosyncratic noise and protecting against over-fitting in-sample (Zhang and Wang, 2023). The PCA trend estimate (τ_{PCA}) arises implicitly as the linear combination of the five distinct trend estimates that captures the maximum variance among all possible linear combinations. This is the first principal component and represents the overall shape of the individual trends. In our case, τ_{PCA} explains over 99 % of the system variance, i.e., it keeps most of the information of the five individual estimates.[12] All the loads ($w_{1,k}$) are similar, as is usual in highly correlated systems. *Subfigure A.4.b* of Annex *Figure A.4* depicts the computed principal components (PC_p), the loadings ($w_{p,k}$), and the eigenvectors

[9] We omit including p-values for simplicity but they are available upon request.

[10] This balance proves to be good, as we observe an improvement in goodness of fit and error variability in the MLR combined estimate when adding the fifth estimate (τ_{SS2000}).

[11] The high correlation among similar performers makes the interpretation of weights in the MLR combination difficult. The weight might change substantially with different data frequencies (see *Annex A.3*).

[12] The PCA is computed from the covariance matrix as the individual estimated trends have the same nature, their mean, and variance being on the same scale.

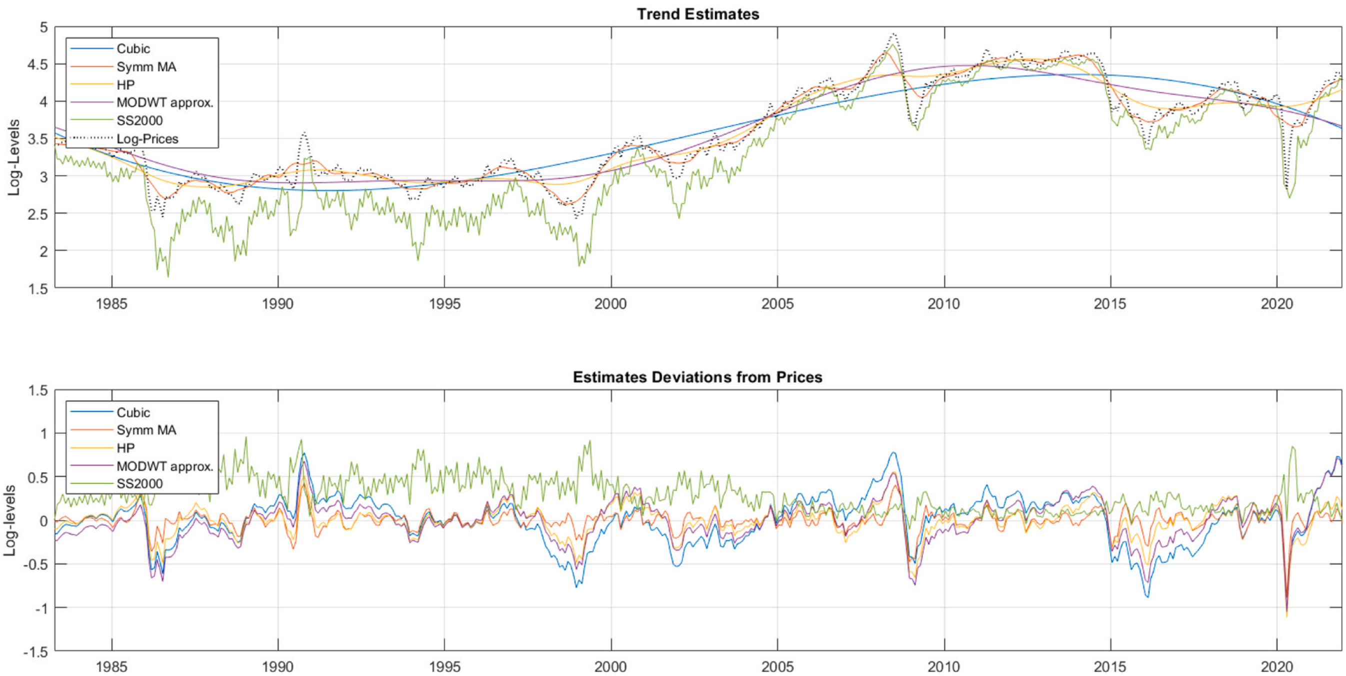


Fig. 1. CL Price Estimated trends and their Deviations from Prices.

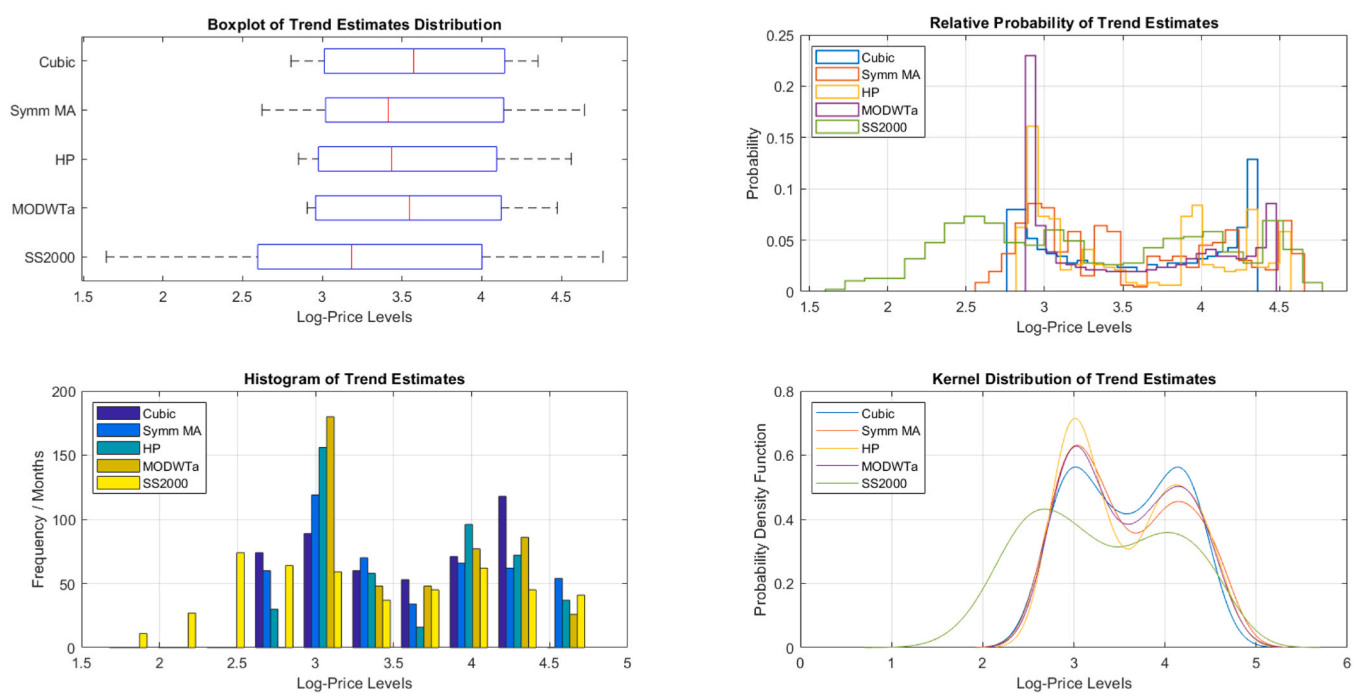


Fig. 2. Distribution of CL Price Estimated Trends.

(W_p).

Secondly, we apply a Feed-Forward Neural Network trained to learn how to estimate the trend from the five estimated trends given *lnP* at each month-end *i*. The FFNN, which resembles a multi-layer perceptron, uses a single hidden layer with 20 neurons (as specified in Table 1.B). The architecture consists of an input layer with five examples (one for each of the five individual trend estimates) for each input neuron and an output layer with one neuron (which produces the final combined forecast).

To train the network and, most importantly, to prevent overfitting, the full dataset was partitioned into three independent subsamples: 70 % for training (0.70n), 15 % for validation (0.15n), and 15 % for testing (0.15n).

We employed the standard early stopping technique. The model was trained iteratively on the training set, while its performance (Mean Squared Error) was simultaneously monitored on the separate validation set. The training process was automatically halted when the error on the validation set failed to improve, and the 'least-error weights' from that best-performing epoch were retained as the final model. This ensures the model generalizes well to new data. The testing subsample was held out

Table 2
Descriptive Statistics of the CL Estimated Trends.

		Maximum	Minimum	Mean	Median	StdDev	Skewness+	Kurtosis++	Coeff, Var.	Variability+++
2.a Log-Prices										
		4.8980	2.4257	3.5781	3.4243	0.6268	0.2333	1.7775	17.52 %	
2.b Trend Estimates										
	Cubic	4.3540	2.8025	3.5781	3.5738	0.5583	0.0007	1.4577	15.60 %	45.5 %
	symmMA	4.6456	2.6210	3.5780	3.4131	0.6077	0.2353	1.6433	16.98 %	18.9 %
	HP	4.5630	2.8509	3.5781	3.4349	0.5896	0.2310	1.4783	16.48 %	29.3 %
	MODWT	4.4745	2.9050	3.5850	3.5474	0.5719	0.1682	1.4501	15.95 %	38.5 %
	SS2000	4.7607	1.6441	3.2912	3.1832	0.7854	0.0874	1.7725	23.86 %	32.8 %
2.c Deviations (residuals)										
	Cubic	0.7797	-1.1051	0.0000	0.0366	0.2850	-0.3244	3.7193		
	symmMA	0.4259	-0.8878	0.0002	0.0076	0.1182	-1.1673	11.7098		
	HP	0.5520	-1.1134	0.0000	0.0046	0.1835	-0.8442	6.7736		
	MODWT	0.7105	-1.0496	-0.0068	0.0018	0.2414	-0.3176	4.2489		
	SS2000	1.0599	-0.1058	0.2869	0.2465	0.2055	0.9051	3.5592		

+ Positive values of obliquity (skewness) indicate that the distribution function has a longer tail to the right of the mean. Negative values indicate that the distribution function has a longer tail to the left of the mean. A Normal distribution has a 0 asymmetry, that is, it is symmetrical. + + Matlab function does not use the −3 Fischer convention, so a Normal distribution shows a Pearson kurtosis of 3. Kurtosis < 3 indicates that the distribution function has a flatter top than the normal distribution. A > 3 kurtosis indicates that the peak of the distribution function is sharper than it is for the normal distribution. The kurtosis of the original price series is very close to Normal, it is only a bit flatter. But the log-series is definitely flatter than Normal. The kurtosis of a sample may differ from that of the whole population. Thus, the kurtosis is biased by a systematic amount based on the sample size and Matlab can correct the bias. + ++Price variability accounted by the trend is the quotient of the standard deviation of the original and the detrended series (Lemke and Gabrys, 2010).

Table 3
CL Price Estimated Trends Correlation.

3.1 L o g - l e v e l s.											
3.1.a Trend Estimates.						*Log-Prices*	*3.1.b Deviations (residuals).*				
	Cubic	*symmMA*	*HP*	*MODWT*	*SS2000*		*Cubic*	*symmMA*	*HP*	*MODWT*	*SS2000*
Cubic	1	0.9181	0.9470	0.9708	0.8908	0.8907	1	0.5522	0.7554	0.8768	-0.2045
symmMA		1	0.9827	0.9509	0.9766	0.9822		1	0.8047	0.6455	0.0470
HP			1	0.9802	0.9489	0.9563			1	0.8837	-0.1296
MODWT				1	0.9194	0.9229				1	-0.1766
SS2000					1	0.9827					1
Log-Prices						1					
				*Mean (Div3)**	0.9485				*Mean (Div3)**	*Mean*	0.4055
				*STD (Div4)**	0.0309				*STD (Div4)**	*STD*	0.4640
3.2 F i r s t D i f f e r e n c e.**											
3.2.a Trend Estimates.						*Log-Prices*	*3.2.b Deviations (residuals).*				
	Cubic	*symmMA*	*HP*	*MODWT*	*SS2000*		*Cubic*	*symmMA*	*HP*	*MODWT*	*SS2000*
Cubic	1	0.1399	0.4607	0.8720	0.0342	0.0398	1	0.9598	0.9948	0.9991	0.3660
symmMA		1	0.6011	0.2405	0.1835	0.2469		1	0.9720	0.9618	0.3881
HP			1	0.6754	0.1398	0.1828			1	0.9966	0.3741
MODWT				1	0.0522	0.0682				1	0.3671
SS2000					1	0.1893					1
Log-Prices						1					
				*Mean (Div3)**	0.3399					*Mean (Div3)**	0.7379
				*STD (Div4)**	0.2923					*STD (Div4)**	0.3137

* Diversity indicators in the slected pool of forecasts (Lemke and Gabrys, 2010).

and used only after this entire process to provide an unbiased evaluation of the final, selected model's out-of-sample performance.

4.4. *Wavelets combination method: multiscale ensemble time-scale (METS)*

The seventh method used is a novel, wavelet-based ensemble based on the *Multi-Ensemble Time-Scale* (METS) algorithm by Percival and Senior (2012). This approach is fundamentally an ensemble method rather than a simple combination. Whereas combination methods (such as a weighted average) combine the entire output of different models, an ensemble builds a new forecast from the constituent parts of the individual models.

Percival and Senior (2012) originally developed METS to solve a problem in physics: estimating an unobservable "true time" by combining signals from different atomic clocks, each performing differently at various time scales. The utility of METS for this paper is the direct analogy: our "true trend" is unobservable, just like their "true time." We assume no individual trend estimate is always the best; rather, each captures different features of the unobservable trend. Paraphrasing Percival and Senior (2012), we can approximate the unobservable trend by "combining measured [trend] differences… to form a reference [trend] that is more stable than any of the [individual estimates]."

The METS procedure is non-parametric and works by first defining the input signals not as the trend estimates (τ_k) themselves, but as their deviations from the observed price ($d_k = lnP - \tau_k$). Each deviation series d_k is then decomposed into multiple time scales, or frequencies, using the *Maximum Overlap Discrete Wavelet Transform* (MODWT). Following decomposition, an optimal ensemble is created at each scale, where the algorithm "weights the signals in each time scale only by their variances" to build a new set of optimally stable composite wavelet coefficients. Finally, these composite coefficients are reassembled using an *Inverse Wavelet Transform* (IWT) to create a single, unified signal that is more stable than any of the individual inputs.

The original METS algorithm combines all time scales, which is not ideal for our purpose. The central objective of this paper is to isolate the long-term trend and explicitly remove high-frequency noise. Therefore, we introduce a modified procedure, METS 2.0, which incorporates a

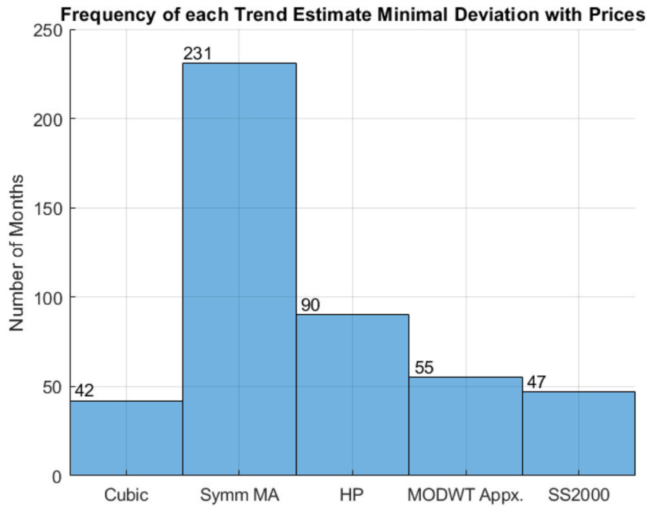


Fig. 3. Minimal Deviations of Trend Estimates from CL Prices.

denoising step prior to the ensemble creation. This approach is inspired by other hybrid models that use signal processing for data cleaning (Shabri and Samsudin, 2014; Wang et al., 2022).

The METS 2.0 procedure begins with denoising, where the individual deviation series (d_k) are first processed with a wavelet filter to remove the high-frequency, irregular signals. These components are incompatible with the persistent, low-frequency trend we aim to capture (see Figure A.5.a). The METS combination is then applied only to the remaining, more persistent cyclical and smooth components. Next, each denoised d_k series is decomposed using the MODWT into different scale levels (from 7 to 1). Following this, a new ensemble signal is created by selecting the component with the least variance at each respective scale level.

Our results indicate the lowest variance is found in the d_{symMA} (deviations from the *Symmetric Moving Average*) for almost all scales, except for the second and third highest/finest scales, where d_{SS2000} (deviations from the Schwartz & Smith model) is selected. Finally, the consolidated deviation signal is reconstructed from the IWT of these selected, least-variance components. This modified process ensures the final forecast is built exclusively from the most stable and persistent dynamics shared by the individual models, preventing short-term noise from contaminating the long-term trend estimate.

The final combined trend forecast from our modified algorithm, τ_{METS}, is obtained by subtracting this consolidated deviation signal from the observed log-price series. A detailed technical description of the complete METS 2.0 procedure is provided in Annex A.5 (see Figures A.5.a to A.5.d). Fig. 4 presents a visual comparison of the trend forecasts generated by all seven combination methods. A preliminary inspection suggests that the τ_{METS} and τ_{MLR} forecasts track the price evolution most closely. In contrast, other estimates, such as τ_{median} and $\tau_{network}$, appear to diverge more significantly, particularly during periods of high volatility marked by sharp price fluctuations.

5. Tests and results

To evaluate the differences among the estimated trends and select the best one, we must face the problem of forecasting an unobservable variable, which implies that the forecasting error remains uncertain even in the case of a potential perfect forecast.[13] Hence, we use two procedures to evaluate the ability of the estimates to capture the "true trend." First, we employ a battery of statistical procedures and select the one that best fits the prices (the observed signal about the trend) in both in-sample and out-of-sample settings. The second consists of evaluating their performance in trend-following trading strategies and selecting the best forecast for profitability, efficiency, and capacity, which we explain later in this section.

5.1. Fitting and forecasting

We compare the descriptive statistics of the combined forecasts and their deviations from actual prices (Table 4) with those of the individual forecasts (Table 2). The coefficient of variation of combined forecasts fits the price volatility better in most cases. Exceptions are the Median and Neural Network combinations, which show 1 % below and above prices' variation coefficient, respectively. As shown in Annex *Figures A.6.a, A.6.b,* and *A.8.b,* $\tau_{network}$ is dominated by τ_{SS2000}, which is the most different among the individual estimates and shows a 6 % higher coefficient and a significantly lower minimum than prices. τ_{symMA} is the only individual estimate whose variation is similar to that of prices (see Table 2).

Table 5 depicts the results of the Signal-to-Noise Ratio, SNR (see description in *subtable 1.C* of Table 1) that tests how strong the signal of each estimated trend is in comparison with prices (see Timmermann, 2018 and Zhang and Wang, 2023). The weakest predictor is τ_{SS2000}, and τ_{METS} is the strongest. The second and third positions are τ_{MLR} and τ_{symMA}. This confirms previous results in Table 4 that point to τ_{METS} as the trend that best fits prices in terms of standard deviation.

We now compare the trends regarding the frequency at which they attain the minimum deviation from prices (Fig. 5). τ_{METS} shows the highest frequency of minimal deviations, followed by τ_{MLR}. The scoring stands in different market conditions: τ_{METS} still has the largest proportion of observations with minimal deviations from prices in both low and high volatility periods and τ_{MLR} follows second, as shown in the *subfigure 6.a* of Fig. 6. The *subfigure 6.b* of Fig. 6 shows that these two estimates are the first (τ_{MLR}) and third (τ_{METS}) most insensitive to market changes. It is important to mention that the most sensitive estimates in this regard are τ_{SS2000} and τ_{PCA}, in that order, while τ_{median} and τ_{cubic} are tied in third.

Fig. 7 shows the deviations of the individual estimates (*subfigure 7.a*) from those of the combined ones (*subfigure 7.b*); the latter are visibly more stable (and their average standard deviation smaller) than the former. Moreover, when the individual estimates diverge substantially from prices, the combined estimates often reduce the difference.

We also assess the precision of the estimated trends by analysing the average size of their deviation from prices with four mean error metrics (Mean Square Error, MSE; Root MSE, RMSE; Mean Absolute Error, MAE; and Mean Absolute Percent Error, MAPE) described in *subtable 1.C* of Table 1 and usually used in energy price forecasting (see Lu et al., 2021). Results reported in Fig. 8 indicate that the combined estimates generally show smaller mean errors than the individual estimates across the four metrics. τ_{METS} and τ_{MLR} have the lowest mean error across all metrics and are relatively insensitive to changes in market volatility conditions. This confirms that combining methods improves the flaws of individual estimated trends, providing more precision and stability across different market conditions. We now compare the estimated trends performance out-of-sample. To do so, we compute them recursively from an initial subsample of 116 observations[14] and forecast them 1, 3, 6, and 12

[13] For this reason, we refer to deviations from prices instead of forecasting errors, as we may never know if it is an error in forecasting the unobservable long-term trend level or a short-term deviation from prices.

[14] Lu et al. (2021) find that only 17 % of publications divide the raw data into three data sets: training, validation, and testing, and most consider training and testing only. We divide the total data sample into 349 rolling-over subsamples to estimate (train) and test. The first is our training subsample, set at 25 % of the total sample size to attain fair estimates but enough out-of-sample test data to see how effective our trend estimates are in long-term trading strategies when facing turbulent market conditions. This sampling arrangement allows us to observe whether the estimation methods and combinations are sensitive to small samples.

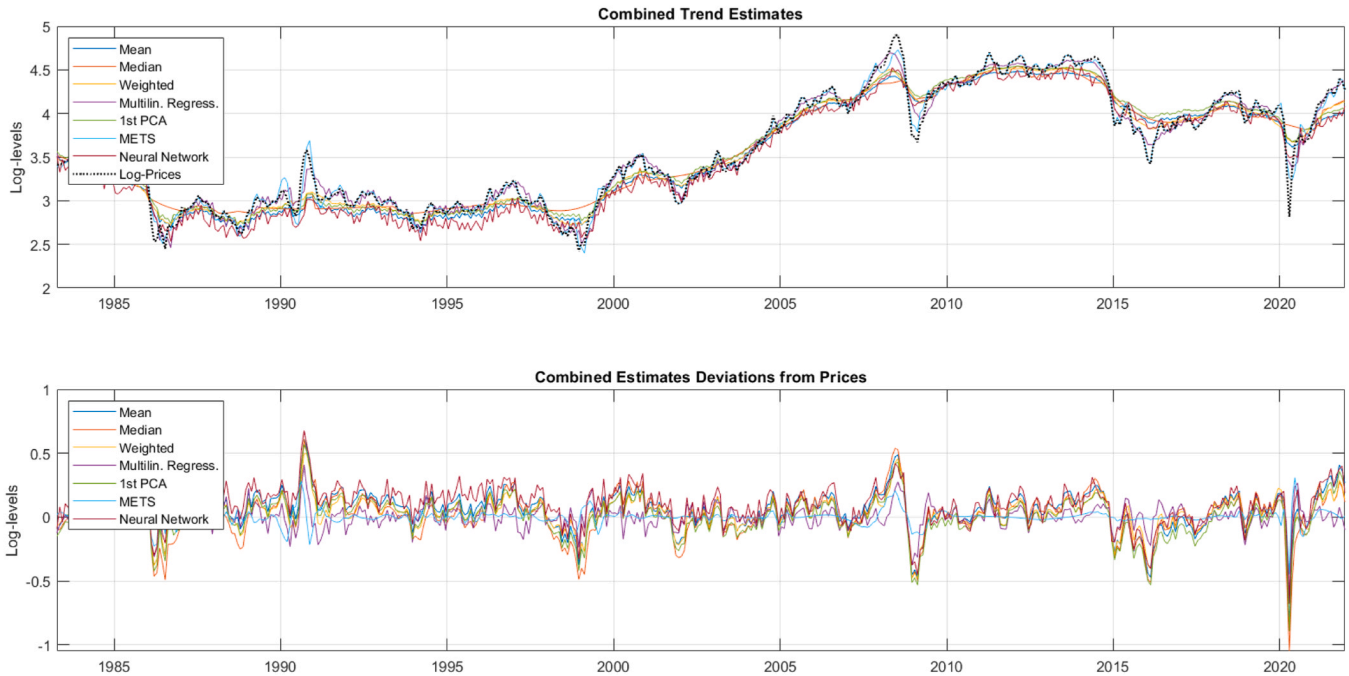


Fig. 4. Combinations of CL Individual Estimated Trends.

Table 4
Descriptive Statistics of Combined Estimated Trends.

	Maximum	Minimum	Mean	Median	StdDev	Skewness+	Kurtosis++	Coeff. Var.	Variability+++
4.a Log-Prices									
	4.8980	2.4257	3.5781	3.4243	0.6268	0.2333	1.7775	17.52 %	
4.b Individual Trend Estimates									
Cubic	4.3540	2.8025	3.5781	3.5738	0.5583	0.0007	1.4577	15.60 %	45.5 %
symmMA	4.6456	2.6210	3.5780	3.4131	0.6077	0.2353	1.6433	16.98 %	18.9 %
HP	4.5630	2.8509	3.5781	3.4349	0.5896	0.2310	1.4783	16.48 %	29.3 %
MODWT	4.4745	2.9050	3.5850	3.5474	0.5719	0.1682	1.4501	15.95 %	38.5 %
SS2000	4.7607	1.6441	3.2912	3.1832	0.7854	0.0874	1.7725	23.86 %	32.8 %
4.c Combined Trend Estimates									
Mean	4.4849	2.6701	3.5221	3.4125	0.6094	0.1559	1.4717	17.30 %	24.03 %
Median	4.5429	2.8472	3.5579	3.4349	0.5856	0.1932	1.4498	16.46 %	27.42 %
Weighted	4.5179	2.7203	3.5499	3.4219	0.6048	0.1973	1.5254	17.04 %	21.26 %
Multilin.Regress.	4.6952	2.4602	3.5781	3.4356	0.6196	0.1884	1.7155	17.32 %	15.16 %
PCA	4.5442	2.7178	3.5729	3.4658	0.6154	0.1562	1.4685	17.22 %	24.34 %
METS	4.7314	2.3981	3.5778	3.4249	0.6213	0.2073	1.7247	17.36 %	9.29 %
Neural Network	4.5729	2.5313	3.4823	3.3677	0.6377	0.1535	1.5161	18.31 %	23.69 %

$^{+}$ Positive values of obliquity (skewness) indicate that the distribution function has a longer tail to the right of the mean. Negative values indicate that the distribution function has a longer tail to the left of the mean. A Normal distribution has a 0 asymmetry, that is, it is symmetrical. $++$ Matlab function does not use the -3 Fischer convention, so a Normal distribution shows a Pearson kurtosis of 3. Kurtosis < 3 indicates that the distribution function has a flatter top than the normal distribution. A > 3 kurtosis indicates that the peak of the distribution function is sharper than it is for the normal distribution. The kurtosis of the original price series is very close to Normal, it is only a bit flatter. But the log-series is definitely flatter than Normal. The kurtosis of a sample may differ from that of the whole population. Thus, the kurtosis is biased by a systematic amount based on the sample size and Matlab can correct the bias. $+ ++$Price variability accounted by the trend is the quotient of the standard deviation of the original and the detrended series (Lemke and Gabrys, 2010).

months ahead. Then, we increase the sample by 1 observation each time to make the forecasts; that is, we compute 349, 347, 344, and 338 forecasts 1, 3, 6, and 12 months ahead, respectively. Notice that we feed the models with "real-time data" each time step[15] (see Annex A.7 for the details of the dynamic forecasting exercise).

Results in Table 6 indicate that τ_{METS} and τ_{MLR} (see *subtable 6.b*) are the best scorers out-of-sample as they present the lowest and most stable MSFE (see description in *subtable 1.D* of Table 1) in 1- to 6-month horizons, and $\tau_{weighted}$ shows the least MSFE and standard deviation in the 12-month horizon tied by $\tau_{METS,}$ which is the best in all seven tests.

Interestingly, though the MSFE of all estimated trends (aside from $\tau_{network}$ in the period of high volatility) increases monotonically with the forecast horizon, the forecasting power of the estimated trends does not decline substantially as the standard deviation of the MSFE across horizons is much less than the one of the MSFE in each horizon. Moreover, the sensitivity of the MSFE to changes in the market volatility decreases in longer forecast horizons for all estimated trends (see Annex A.8 for details on the MSFE in different volatility periods). Zhang and Wang, (2023) find the opposite, a significant deterioration of individual

15 "Time steps" refer to discrete points in time within a sequential forecasting (or estimation) process where calculations are performed at regular time intervals. This term is commonly used in numerical simulations, control systems, machine learning and time series analysis. While "vintages" refer to the state of a dataset at a particular point in time, recognizing that data is often revised over time as more information becomes available. This term is particularly used in economics and statistics.

Table 5
Signal-to-Noise Ratio of CL Estimated Trends.

5.a Individual Trend Estimates					*5.b Combined Trend Estimates*						
Cubic	*symmMA*	*HP*	*MODWT*	*SS2000*	*Mean*	*Median*	*Weighted*	*MLR*	*PCA*	*METS*	*Network*
22.09	29.76	25.93	23.55	19.64	26.95	26.39	28.45	31.66	27.52	**35.90**	26.04

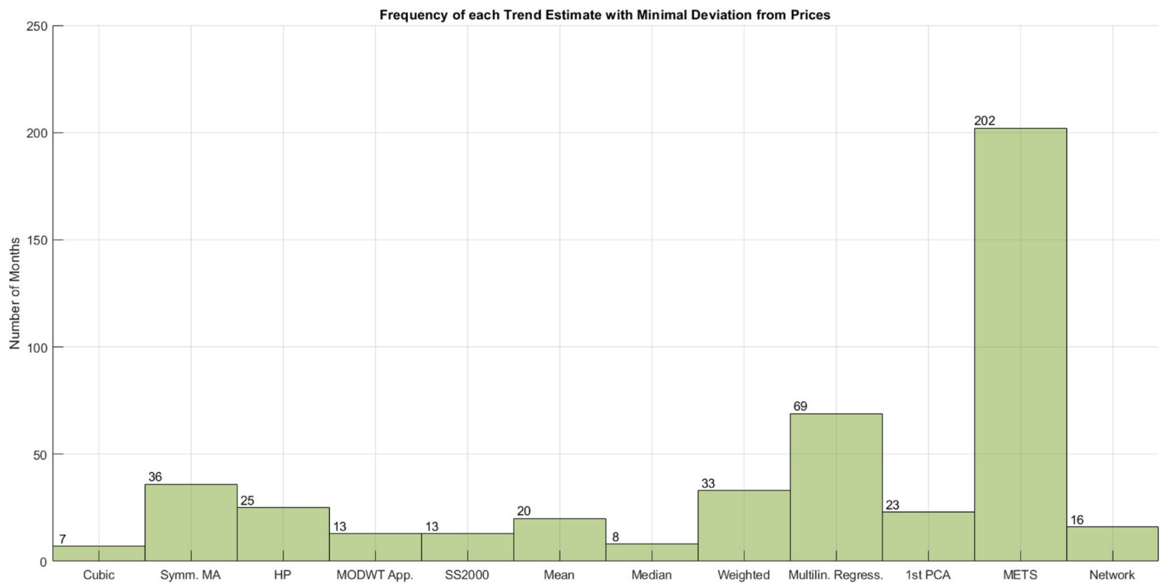


Fig. 5. Minimal Deviations of Individual and Combined Estimates from CL Prices.

predictors in longer horizons that they attribute to their lack of a long-run approach. In contrast, our estimated trends have a long-term nature.

Additionally, the out-of-sample R^2_{os} statistic (see description in *subtable 1.D* of Table 1) in Table 6 used by several authors referred by Zhang and Wang, 2023 like Campbell and Thompson (2008) indicates that practically all the trend estimates forecast CL prices are more accurate than the forecast benchmark in the literature (τ_{mean}) except τ_{SS2000} and $\tau_{network}$ which is greatly influenced by τ_{SS2000} that shows the largest MSFE across all horizons.

We also use a modified form of the time-aligned direction statistic (Wang et al., 2022), FDstat (see description in *subtable 1.D* of Table 1) and find no differences in the way all the estimated trends anticipate prices' direction: 93.97 %, 95.38 %, 94.75 %, and 93.77 % of the time in 1-, 3-, 6-, and 12-months horizons, respectively.

Finally, we formally test if these performance differences are statistically significant. We compute the Diebold-Mariano (DM) test, with the results also presented in Table 6. This test provides critical validation, confirming that the superior performance of τ_{METS} and τ_{MLR} is not due to random chance. The DM test p-values show that both τ_{METS} and τ_{MLR} are statistically significantly more accurate (at the h=1, h=2, and h=3 horizons) than the simpler combination methods, including τ_{Mean}, τ_{Median}, and τ_{PCA}. Notably, the test also indicates that the difference between τ_{METS} and τ_{MLR} is not statistically significant, placing them in a top tier of their own.

5.2. *Trend-following trading strategies*

We now assess the comparative performance of all trend forecasts in trend-following trading strategies with the sole purpose of stretching the conditions of the previous statistical evaluation, rather than the merit of the strategies themselves. These strategies aim to capitalize either on prices reverting to the trend level in the long term or on the momentum of commodity market trends by identifying and following the direction of short-term movements to decide whether to open positions or not.

We define a strategy rule for different investment horizons using the forecasts provided by the individual and combined methods, and compare the results in terms of profitability, efficiency, and risk in the long term to choose the best trend forecast.[16] The trend-following strategy focuses on making trading decisions on a month-end date, considering the estimated trend calculated using all available information to that date. Thus, we use the forecasts computed in the out-of-sample exercise in subsection 5.1 to determine the positions in the trading strategy for 1, 3, 6, and 12 months ahead (investment horizons). Positions are initiated at the end of each month from November 1992 to November 2021, which implies 349, 347, 344, and 338 time steps, respectively. On any given month-end (time step), positions that were opened in the past and have not yet reached the intended horizon may coexist with new positions being initiated. The exercise involves the following trading rules:

– Positions can be open long (buy), short (sell), or not open. Once opened, the position will be held until the end of the horizon.[17]
– The decision for each horizon is based solely on the trading indicator within the trading system (i.e., long and short positions at different horizons can be opened at the same time step). Note that the

[16] For traders employing a multi-pronged strategy, it is important to clarify that while the seven combined trend forecasts involve combining multiple individual forecasts, the resulting trading signals are generated independently by each forecast, thus avoiding mixed buy and sell signals.

[17] Termination dates do not align with the actual CL futures calendar, but are purposely aligned with the chosen horizons of this study. Holding CL future contracts until expiration requires the physical delivery of CL contracts (both long and short) only possible for participants with delivery access at Cushing, OK. Such access is typically limited to large commercial energy firms, not retail investors.

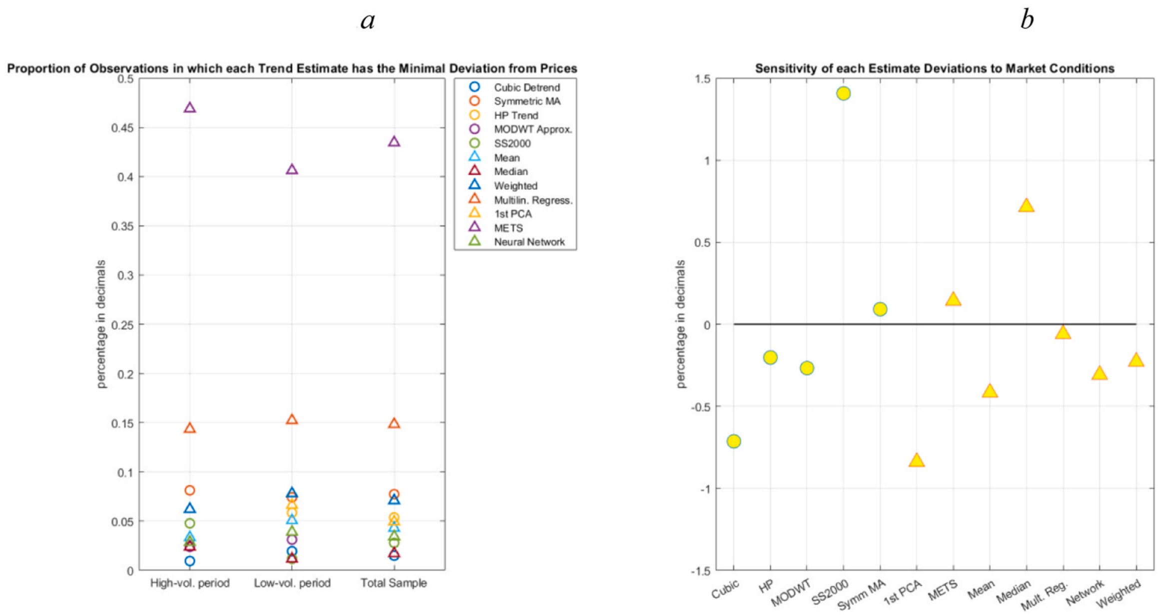


Fig. 6. Deviations of CL Individual vs Combined Estimated Trends across Market Conditions. In Subfigure 6.b, if positive, the Individual (circle) or Combined (triangle) Estimate has more minimal deviations in a market with high volatility and higher prices (Expansion) than in one with low volatility and lower prices (Recession), and vice versa if negative. Closer to zero, the frequency of minimal deviations is less sensitive to changes in the market.

portfolio held at each month-end is not intended as a spread calendar strategy, and each position is closed at the end of its designated horizon.

– Trading decisions are made at the end of the month. If a decision is made to open a position, it is assumed to be executed contemporaneously in after-hours trading at the corresponding monthly average closing price.
– Positions are never escalated or switched. The strategy involves trading one unit of the commodity (one barrel of CL). Trade fees and carrying costs are adjusted accordingly, but other transaction costs (margins, taxation, liquidity, close-out, etc.) are not considered.
– Strategy results are evaluated at the end of the horizon, using the average nominal log return between the opening price and the last monthly average at that date.[18, 19]

These rules may differ from real trading. For example, we assume financial resources to be unlimited and independent of past trading results or a limited initial endowment. We also assume the risk-free interest rate and other costs remain constant to isolate the impact of our trend forecasts on trading results (*ceteris paribus*). This aligns with our endogenous approach, which is based solely on the commodity price. It is also important to mention that the trading examples provided here are for illustrative purposes only, do not reflect actual trades, and do not constitute investment advice. The past performance of our forecasts and strategies is not indicative of future results. This exercise is not intended to be implemented in real trading and produce actual returns, but to compare the trend forecasts' ability across different conditions imposed by the trading systems and historical scenarios.

5.2.1. *Trading systems*

a. Reversion to the Trend.

At time i and horizon h, if the trend estimate at that horizon is above the price, then buy (+1); if below, sell (-1); if equal, do nothing (0). We constrain the decision by restricting it to at least exceeding a break-even point equal to the risk-free asset (rf), assuming a flat, constant annual rate of 3 %; otherwise, a position is not opened. The trading decision is evaluated individually at the end of the horizon, $i+h$, with the actual log return of each term horizon ($r_{h,i+h}$).

$$position_{T,h,i} = \begin{cases} +1 & if \dfrac{T_i}{\ln P_i} > 1 + rf_h \Rightarrow Buy \\ -1 & if \dfrac{T_i}{\ln P_i} < 1 - rf_h \Rightarrow Sell \\ 0 & otherwise \Rightarrow do\ nothing \end{cases}$$
$$where \quad i = sampstpt : n - h$$

$$r_{h,i+h} = \ln P_{i+h} - \ln P_i$$
$$P\&L_{T,h,i} = position_{h,i} \times r_{h,i+h}$$
$$where \quad i = sampstpt \quad : \quad n - h$$

To evaluate the results of this trading strategy with two price indicators: (1) an always buying Roll-over strategy assuming the current price is the best forecast of the future price as it summarizes all the past price data according to the Martingale property and (2) Nilsson's h-sliding-windows moving average (MA) of monthly returns ($r_{1,i}$) across the same term horizons.[20]

$$MA_{h,i} = \tfrac{1}{h}\sum_{t=1}^{h} r_{1,i}$$
$$where \quad i = sampstpt \quad : \quad n - h$$
$$MA_{1,i} = r_{1,i}$$
$$r_{1,i} = \ln P_i - \ln P_{i-1}$$

[18] While concerns exist regarding the feasibility for investors to buy or sell at the monthly average price by the end of the month, particularly in futures markets (Conlon et al., 2024), they might utilize over-the-counter (OTC) assets, such as Cliquet (or Ratchet) Asian-Chooser options, to secure this monthly average price for consecutive end-of-month transactions. Additionally, two corporations seeking to mitigate volatility risk could establish a private agreement to exchange an asset at the month's average price by the end of the month.

[19] Besides the monthly average nominal price log-return, Zhang and Wang (2023) consider other two kinds of oil returns: the closing EoM price return (above the risk-free rate) and the real returns computed by deflating the oil average nominal price return by the CPI of the USA.

[20] There are other trading rules providing buy and sell signals depending on prices moving above or below a certain MA (i.e see Narayan et al., 2015).

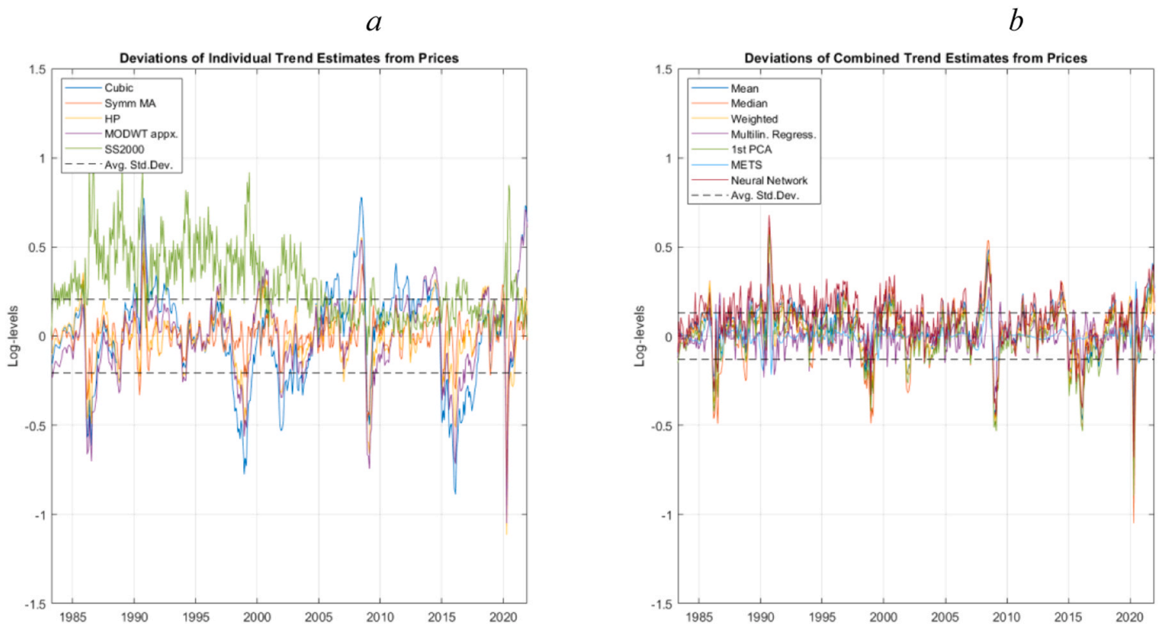


Fig. 7. Deviations of CL Individual vs Combined Estimated Trends.

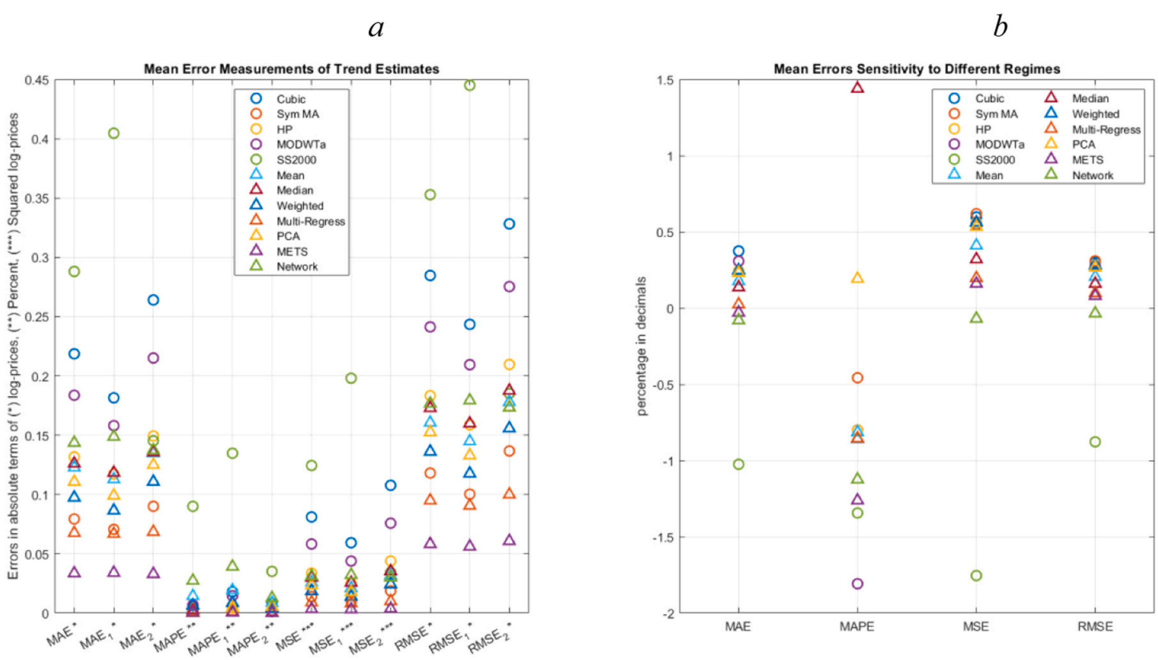


Fig. 8. Mean Error of CL Individual and Combined Estimated Trends. Subfigure 8.a shows four mean error measurements using the total sample, as well as the subsamples of the low-volatility with lower prices period and the high-volatility with higher prices period. In Subfigure 8.b, if positive, the Individual or Combined Estimate has a larger average error in the high-volatility with higher prices period than in the low-volatility with lower prices period, and vice versa if negative. Closer to zero, the average error measure is less sensitive to changes in the conditions of market volatility.

$$position_{MA,h,i} = \begin{cases} +1 & \text{if } MA_{h,i} > 0 + rf_h \Rightarrow Buy \\ -1 & \text{if } MA_{h,i} < 0 - rf_h \Rightarrow Sell \\ 0 & \text{otherwise} \Rightarrow \text{do nothing}. \end{cases}$$
$$where \quad i = sampstpt \quad : \quad n - h$$

$$P\&L_{MA,h,i} = position_{MA,h,i} \times r_{h,i+h}$$
$$where \quad i = sampstpt \quad : \quad n - h$$

We do not consider the alternative of investing in the risk-free rate. We isolate the profit and loss (P&L) of CL trading only and consider the risk-free rate as a transaction cost only when deciding to trade the commodity. Although this is not an accurate reflection of the actual cost, we include it to recognise that it is non-zero and impose a break-even condition to the trend forecasts to further their assessment. However, we do not include any cost in the P&L as we do not perform a net return calculation.

b. Momentum.

We employ a simple Momentum strategy based on the slope between two points h periods apart. However, traders may have other technical tools available in real-time, like the Relative Strength Index, a more sophisticated momentum oscillator. In our case, at time i, if the slope between two points of the trend estimate (T_i) is positive, we buy (open a

Table 6
Forecasting Error of Individual and Combined Trend Estimates.

Forecasting Method	MSFE*				MSFE STD across h	MSFE STD**				Diebold-Mariano Test***				R^2_os****			
	h= 1	h= 3	h= 6	h= 12		h= 1	h= 3	h= 6	h= 12	h= 1	h= 3	h= 6	h= 12	h= 1	h= 3	h= 6	h= 12
6.a Individual Trend Estimates																	
Cubic	0.099	0.113	0.131	0.162	0.027	0.315	0.336	0.363	0.403	**0.000**	**0.000**	**0.030**	0.282	0.721	0.699	0.675	0.639
symmMA	0.033	0.055	0.084	0.131	0.042	0.182	0.234	0.289	0.359	**0.012**	0.216	0.282	0.446	0.907	0.853	0.791	0.707
HP	0.049	0.070	0.097	0.136	0.038	0.221	0.264	0.311	0.368	**0.000**	**0.007**	0.054	0.406	0.863	0.814	0.760	0.695
MODWT	0.145	0.155	0.169	0.191	0.020	0.346	0.357	0.373	0.395	**0.000**	**0.000**	**0.002**	0.074	0.592	0.588	0.582	0.574
SS2000	3.246	3.277	3.321	3.403	0.068	1.234	1.240	1.252	1.264	**0.000**	**0.000**	**0.000**	**0.000**	-8.143	-7.734	-7.222	-6.602
6.b Combined Trend Estimates																	
Mean	0.355	0.375	0.404	0.448	0.040	0.432	0.449	0.474	0.505	**0.000**	**0.000**	**0.000**	**0.000**	0.000	0.000	0.000	0.000
Median	0.044	0.064	0.090	0.132	0.038	0.205	0.248	0.295	0.357	**0.000**	**0.016**	0.106	0.424	0.876	0.829	0.777	0.704
Weighted	0.035	0.055	0.082	**0.122**	0.038	0.185	0.232	0.283	**0.344**	**0.000**	0.091	0.531	0.919	0.900	0.852	0.797	**0.727**
Mult.Regress.	0.024	0.048	0.079	0.128	0.045	0.155	0.219	0.280	0.357	0.839	0.894	0.851	0.567	0.932	0.873	0.805	0.714
PCA	0.074	0.093	0.119	0.157	0.036	0.242	0.275	0.315	0.365	**0.000**	**0.000**	**0.001**	**0.028**	0.791	0.752	0.706	0.649
METS	**0.024**	**0.047**	**0.078**	0.123	0.043	**0.150**	**0.213**	**0.274**	0.344					**0.934**	**0.874**	**0.808**	0.725
Network	0.631	0.642	0.677	0.703	0.033	0.656	0.661	0.682	0.693	**0.000**	**0.000**	**0.000**	**0.000**	-0.777	-0.710	-0.676	-0.570

* Mean Squared Forecasting Error. Minimal MSFE in each horizon is highlighted in bold blue font.
**The minimal STD in each horizon is highlighted in bold blue font.
***The Diebold Mariano test (1995) loss function is equivalent to the difference of MSEs between models. The results highlighted in blue and bold font reject the null hypothesis (H0) of equal predictive accuracy ($MSFE_{METS}=MSFE_{\{cubic,\ldots,network\}}$) with a p-value of 0.05 in favor of the alternative (H1) of statistically significant different predictive accuracy. In most of those cases, METS is a significantly more accurate predictor than any of the alternatives in terms of a lower MSFE. The Multi-linear Regression comes closely behind METS in second, and the Weighted and Symmetrical MA come in third.
****A positive out-of-sample R^2 statistic indicates that the Forecasting Method is a more accurate forecast (in terms of MSFE) than the Mean method considered benchmark in the forecasts combination literature. The maximal statistic in each horizon is highlighted in bold blue font.

long $position_i$ +1), and if negative, we sell (open a short $position_i$ −1). We restrict this trading system to at least breaking even, which is equal to the costs of carry (the storage of one unit of the commodity for the length of the *h* period)[21] and the transaction fee per trade,[22] both expressed in log levels; otherwise, a position is not opened. While these figures do not accurately reflect the actual costs, estimating their market price is not our objective. Nevertheless, we include them in our trading decisions to acknowledge their non-zero value and to test the robustness of our trend forecasts under various conditions and constraints.

$$position_{T,h,i}=\begin{cases} +1 & \text{if} \quad T_i - T_{i-h} > 0 + fee + carry_h \Rightarrow Buy \\ -1 & \text{if} \quad T_i - T_{i-h} < 0 - fee - carry_h \Rightarrow Sell \\ 0 & \text{otherwise} \Rightarrow \text{do nothing.} \end{cases} \quad where \quad i = sampstpt : n-h$$

In order to contrast the trading signals indicated by the trend forecasts in this trading strategy, we consider also in the trading system the slope between actual log-prices at *i* and *i-h*, ($\ln P_i$ - $\ln P_{i-h}$) following the same decision logic as shown above.

The results of the Momentum strategy are evaluated just as the ones of the Reversion strategy, delivering the gross P&L of CL trading only.

5.2.2. Trading results

Tables 7 and 8 summarise the Reversion and Momentum trading exercise results with the total cumulative return, the forecasting efficiency and a risk-adjusted metric (Sharpe ratio described in *subtable 1.E* of Table 1). Annex *Figures A.9.b* and *A.9.e* present the evolution of the cumulative returns of all the trading indicators across all trading horizons.

We do not annualise the returns with a horizon lower than 12 months as we compare the cumulative returns of all the forecasts within each horizon. However, we can still compare returns across horizons in those tables. Though we depict gross returns, we compare efficiency across forecasts in each horizon as shorter horizons may *caeteris paribus* entail more trades than longer ones; hence, they are more expensive due to trading fees. Thus, we calculate a Forecasting Efficiency ratio as the total cumulative return by the trading ratio.

Annex *Tables A.9.a* and *A.9.b* provide more details of the trading results for both strategies. The first two vertical panels of these tables show the ratio of trading signals provided by the forecasts and their tendency, if any, to short or long positions across the four horizons considered. The total and mean returns, along with their efficiency, are displayed in the middle vertical panels. The last two vertical panels of the tables show the returns volatility (market price risk) and the Sharpe ratio. However, we do not consider risk aversion in the trading decisions.

a. Reversion to the Trend.

Table 7 shows that only for a 1-month horizon, the price indicator (Returns MA) cumulative returns exceed those of the trend forecasts; however, τ_{MODWT} follows closely and becomes the most profitable in a 3-month term and produces positive and increasing returns for all the term horizons with a high and stable number of trades (see Annex *Table A.9. a*). In longer than 1-month terms, other individual (τ_{cubic}) and combined ($\tau_{weighted}$) trend forecasts can provide competitive profits. For 6- and 12-

[21] Storage has a dynamic cost related with production levels and other variables that we are not reflecting here. We consider a constant storage cost of $0.50 USD per barrel per month. We refer to data from 2015 showing monthly storage cost per barrel in salt cavern ($0.25), tank ($0.50–0.75), and ship ($0.75 - $1.40). We also observe CME Crude Oil Storage futures in the first quarter of 2023 showing a monthly average cost per barrel of $0.25. This contract has the same size of the Crude Oil future contract (1000 barrels) and provides the right to storage for a calendar month.

[22] The transaction cost of the bid-ask spread may not be substantial as the CL market is highly liquid. However, transaction fees may affect the P&L of shorter-term strategies as they change position more frequently than long-term strategies (roll-over costs). We consider a constant fee of $1.25 USD per trade of 1000 barrels. We refer to the 2023 data on NYMEX exchange fees.

months, all the forecasted trends outrun the Returns MA indicator. τ_{METS} seems to have stable profits. Thus, it may be suitable for conservative strategies or to bind a floor or a cap. τ_{HP} only seems to work in the 12-month horizon, but not in the shorter terms; this is not surprising given the HP construction aimed at detecting business cycles that last more than one year. However, surprisingly, the best individual contender in the previous section (τ_{symMA}) performs modestly in this second assessment and only in a 12-month horizon.

Annex *Table A.9.a* shows that the Returns MA has a slight but persistent bias to long positions. This may be due to price bubble-prone components (see Annex Fig. 9.*a* and *9.b* and Emekter et al., 2012 reference on rational speculative bubbles in commodity markets). In contrast, the estimated trends are generally somewhat biased to short positions; this is consistent with their assumed cautious nature following long-lasting market fundamentals and the “golden rule of forecasting”: forecasts have to be conservative (Armstrong et al., 2015). τ_{cubic} is the only one providing almost the same amount of buying and selling signals; *it* also maintains a quite high trading ratio in all term horizons, while most indicators produce fewer trading signals as the term horizon increases.

Table 7 shows that τ_{cubic} has the highest cumulative total return in the 12-month horizon. τ_{SS2000} produces trading signals half the time with a strong bias to short positions in all term horizons. It keeps its mean return close to zero yet positive in the 1-month horizon, possibly due to its high-frequency irregularities, which is the same reason why it delivers negative returns in longer terms.

In general, the trend reversion strategy achieves better results in the long term, with almost all the forecasts. In our study, even the rollover strategy (trading ratio of 1 and position bias of $+1$) behaves similarly to the trend forecasts across horizons because the price series we use is a monthly average series, hence a series of monthly trends.

Regarding efficiency, the Returns MA forecast is the most efficient in a 1-month horizon, but the trend forecasts become more efficient in longer terms. The rollover price forecast becomes more efficient as the horizon is longer and is the most efficient in 6 months. We attribute this also to our price series being an actually short-term trends series, although not long-term as we aim in this study. However, always rolling over is a costly or trading-intensive strategy. Thus, although always rolling-over and trading $\tau_{weighted}$ signals have very similar cumulative returns in a 12-month horizon, the first has the highest trading ratio, reaching the maximum number of possible trades in the period, while the second traded less than half of the time, hence, being significantly more efficient.

The top performing forecasts in terms of total returns, forecasting efficiency, and Sharpe ratio in 1-, 3-, and 6-month horizons are Returns MA, τ_{MODWT}, and Roll-over, respectively. However, in a 12-month horizon, there is no one top performer in all three aspects; τ_{cubic} is the top scorer according to the risk-adjusted ratio and accumulates the highest profit with 73 % of the time trading but $\tau_{weighted}$ shows the maximal efficiency, delivering a comparatively fair return with only 43 % of the possible trades (see Annex *Table A.9.a*).

b. Momentum.

Table 8 shows that some trend forecasts have positive returns in 1- and 3-month horizons, respectively. However, only τ_{METS} competes with the price forecast, which is the most profitable in 1 month, and τ_{METS} becomes the most profitable in 3 months. Both indicators have a high number of trades and a long bias in those horizons (see Annex *Table A.9.b*). τ_{METS} also provides positive returns in a 6-months strategy before 2008, but after 2014, it is τ_{MODWT} the one providing profits; the period in-between is of high distress in the global economy, including the energy commodity markets. With the exception of τ_{MODWT} producing positive cumulative returns in all term horizons with a high number of short-biased trades (see Annex *Table A.9.b*), all the other indicators provide poor results in 6- and 12-month horizons. For example, τ_{SS2000} opens positions almost all the times in all term horizons but only produces a positive return in a 1-month horizon. At the same time, τ_{HP} provides the worst momentum trading signals in all the term horizons.

In contrast to what is observed in the Reversion strategy in Table 7, Table 8 indicates that Momentum is a strategy more suitable for the short-term as the best results are generally obtained in horizons less than 12 months with the exception of τ_{HP} forecast showing poor results across all horizons and τ_{MODWT} forecast delivering positive results in 6- and 12-months horizons. Thus, despite this particular trading strategy may not suit our long-term trend forecasting approach, in horizons longer than 1-month terms, still a trend forecast may work better than a price forecast.

In Table 8, the forecasts that accumulate the highest return are also the most efficient and show the highest Sharpe ratio in each horizon: the price and τ_{METS} forecasts in 1- and 3-months, respectively, and τ_{MODWT} in both 6- and 12-months.

The characteristics of Momentum influence results as it is an acceleration- more than a drift-driven strategy, as mentioned by Nilsson (2015). Hence, it is meant to capture short-term trends. The comparison of the four graphs in *Figure A.9.e* evidence that Momentum is more a short- than a long-horizon trend-following trading strategy. Chaves and Viswanathan (2016) also find that the positive performance of the momentum strategy in commodities does not continue beyond 2 or 3 months.[23] Thus, in this particular strategy, several trading signals perform fairly in a 1-month horizon but poorly as the term increases, except for the τ_{MODWT}, which accumulates positive returns in all horizons as it also does in the reversion strategy. In fact, τ_{MODWT} can produce positive log returns when all other indicators provide misleading signals. In a 12-month horizon, τ_{MODWT} can produce positive log returns in periods characterised by either global economic or market-specific shocks around 1995, 2000, and 2004, while other indicators cannot (see the lower right subfigure of Annex Figure A.9.d).

We have not taken into consideration in the decision systems the CL futures market term structure, which may influence the profitability of the momentum strategy in commodities trading (Rallis, 2010) linked to the “normal backwardation” rationale envisioned by Keynes in 1930 and Hicks in 1939, and reviewed by Houthakker (2017). Narayan et al., (2015) find that momentum-based strategies' profitability in commodities trading is sensitive to subsamples before and after the global financial crisis starting in 2007[24] and is also data-frequency dependent.

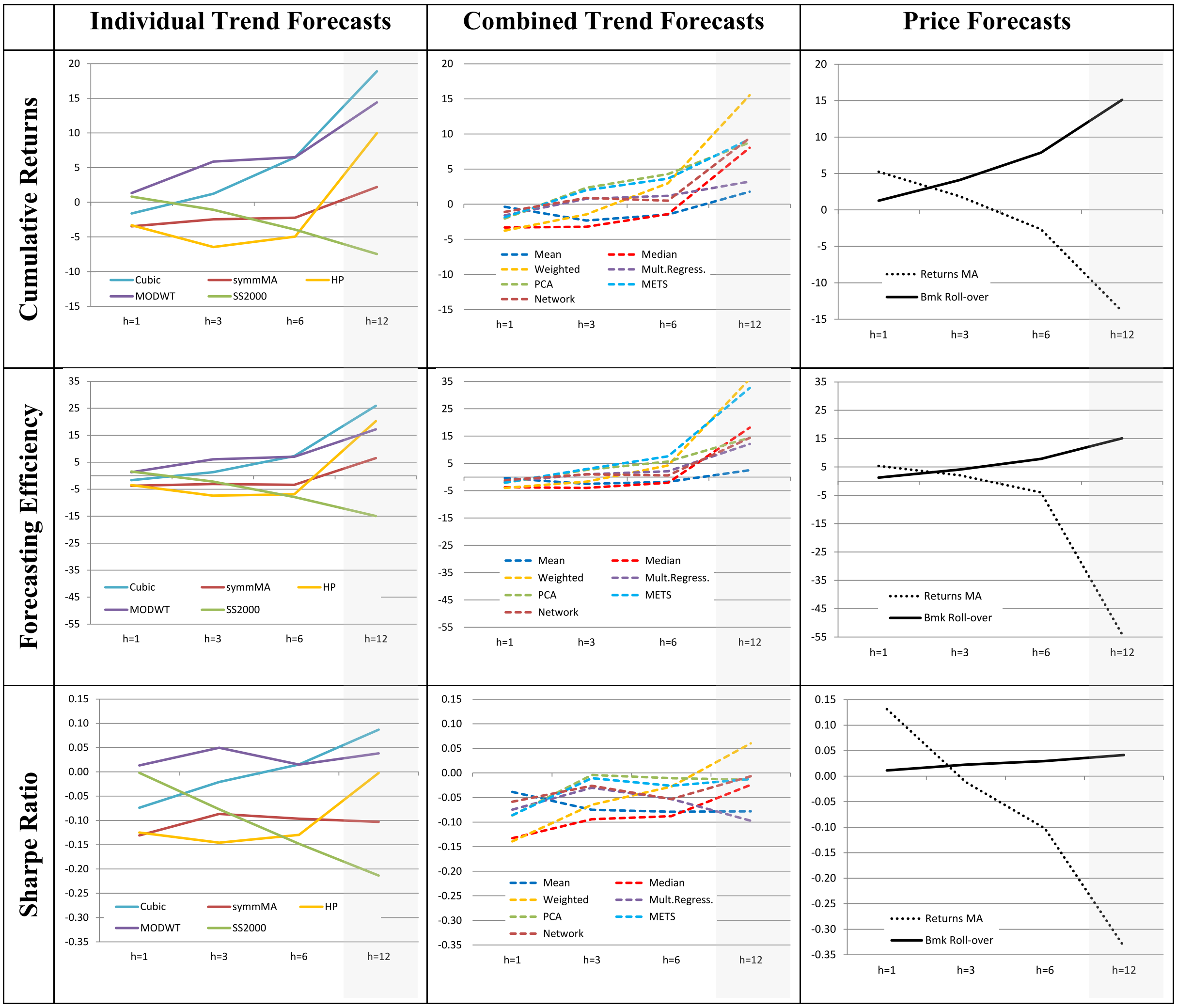
5.2.3. Forecasting capacity

Now, we focus on the Trend Reversion strategy as it is the one from the two presented that is more suitable for a long-term approach, as confirmed by its results. As a measure to compare the forecasting ability of the forecasts, we set the benchmark at the 95 % confidence critical returns of the rollover strategy, which are the log-returns of the price series itself, and evaluate the forecasts within those limits. Table 9 shows that all the forecasts from 3- to 12-month horizons consistently augment the capacity within the 95 % limits; that is, all increase the probability of returns happening within those limits and reduce the benchmark 5 % probability of extreme results. τ_{SS2000} forecasting capacity within limits is the highest in 1- and 3-month horizons and is the one reducing most of the probability of extreme negative returns in those horizons. While τ_{METS} concentrates the highest forecasting capacity in 6- and 12-month horizons and has the least probability of extreme negative returns in those horizons. In both cases, there is a close-to-neutral bias in absolute terms between extreme positive and negative returns. This is particularly interesting to investors of long-term projects who can plan over a wide price range but are highly averse to extreme negative scenarios;

[23] The authors find no evidence of momentum in spot prices beyond a 3-month horizon and the short-lived momentum is attributed to seasonality. On the contrary, they find a strong performance of momentum strategies in futures markets supported on a positive basis, not on a sustained trend of spot prices.

[24] In *subfigure A.9.f.3* of *Annex Figure A.9.f*, cumulative returns where pretty tight in a 1-month horizon Momentum strategy before the Great Recession which followed the 2007/9 crisis.

Table 7
Overall Performance of Forecasts in the Reversion Strategy.



they may even be willing to sacrifice extraordinary profit to avoid unexpected loss.[25]

In general, all the forecasts increase their profit probability as the horizon increases, settling all above 50 % in the 12-month horizon with the exception of the τ_{SS2000} and the MA Returns forecasts, which perform better in shorter terms. τ_{MODWT} has the largest profit probability in 1- and 3-month horizons, while the Roll-over benchmark and τcubic excel in 6- and 12-month horizons, respectively.

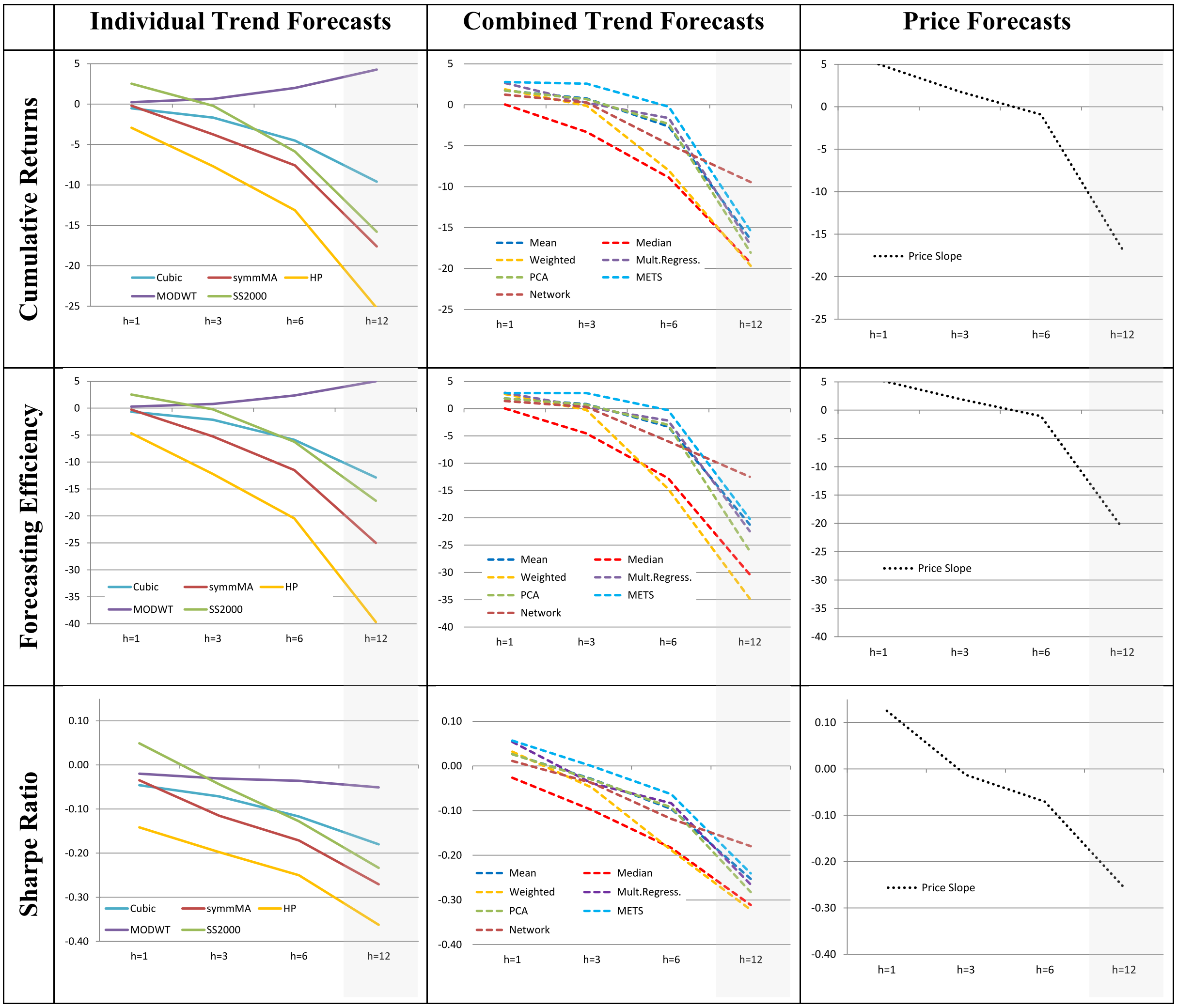
While the content of Table 9 is expressed in terms of probability across different horizons, Fig. 9 is in terms of density, focusing only on a 12-months ahead horizon, showing graphically how forecasts stretch the odds within the benchmark returns limits, and Table 10 shows the forecasting capacity of the trading signals defined as the ratio of the width within the returns limits specified by the benchmark divided by the width of the returns dispersion in the trading process controlled by each forecast (see description in *subtable 1.D* of Table 1).

Fig. 9 shows the distribution of the 12-month benchmark returns, and the centred yellow-shaded area within limits contains 95 % of the observed returns in the market. The critical return values of the forecasts at the 2.5 %-lower and 2.5 %-upper limits are marked with circles for the individual trends, triangles for trend combinations, and a black dot for the price forecast. All the markers are within the yellow-shaded area, meaning that in the long term (12 periods ahead), all the forecasts stretch the 95 %-confidence limits specified by the market price log returns or benchmark. All the forecasts exceed the benchmark forecasting capacity across all horizons, mainly in the longest one (Table 10), and almost all the individual and combined trend forecasts stretch the limits more than the Price Returns MA forecast in the long-term horizon (Fig. 9). That means those trend trading processes have fewer observed returns outside the 95 %-limits set by the market, thus having a higher forecasting capacity. In the long-term, τ_{METS} has the best forecasting ability (Table 10), followed by $\tau_{MLR,}$ and their return densities are outlined by dashed lines in Fig. 9.

CL log-returns are more uncertain as the term increases (see *subfigure A.1.b* of Annex *Figure A.1*). Thus, these results are of great importance because as the term increases, the trend forecasts are capable of stretching the probability of returns occurring in a less dispersed range.

[25] On the contrary, if investors target short-term extraordinary profits instead, a price forecast like the MA Returns may be useful as it has the largest positive extreme returns bias among all forecasts with a one-month horizon. However, it is also riskier as it has less than a 50 % probability of positive returns.

Table 8
Overall Performance of Forecasts in the Momentum Strategy.



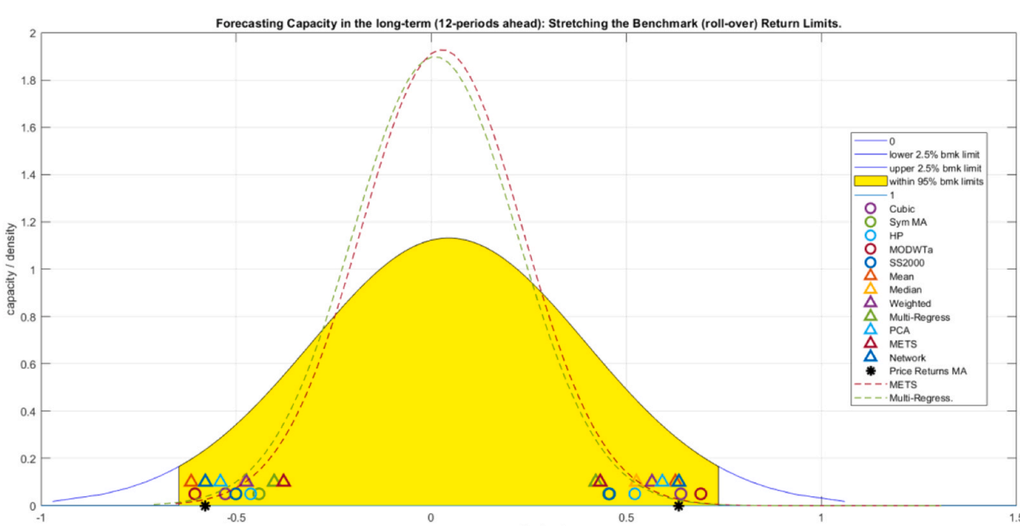


Fig. 9. Stretching the Limits in the Long-term Trend Reversion Strategy.

6. Discussion

The use of monthly average prices in this study instead of EoM or daily prices provides the advantage of a first approximation of price trends more suitable for the approach of this study. Additionally, Díaz-Rodríguez and Robles (2026) conclude that the main characteristics of the long-term trend of a daily price series are preserved when using price series with a lower frequency.

Moreover, given the short-term nature of noise, which is present in higher frequencies, pre-forecast data denoising of hybrid models is unnecessary when using low-frequency data and/or long-term forecasting methods, such as most of the ones used in this study. In contrast, the METS algorithm by Percival and Senior (2012), introduced in this paper as a combination method, deals with estimation residuals, errors, or differences that may contain such noise. Hence, we add a previous step to the algorithm: the WT denoising of the differences to be decomposed, which yields better results and adds METS 2.0 to the hybrid forecasting models collection.

According to how the trends fit prices, the best individual trend estimate is τ_{symMA}, confirming the virtue of its inclusion in the selection. However, no method always has the least deviation; even if it did, there is no certainty that it will remain so in the future (Vilar et al., 2020). τ_{symMA} is outrun by at least two combined estimates (τ_{METS} and τ_{MLR}) and even contested by a third ($\tau_{weighted}$). Among the individual estimated trends, τ_{SS2000} and τ_{cubic} are the worst as they have the least number of minimal deviations from prices. Nonetheless, the combination can benefit from their inclusion with other individual estimates because they are the least deviated estimates from prices on fewer occasions.

Timmermann (2006) argues that combining forecasts with different adaptability capacities will improve, at least on average, the behaviour of individual forecasts across periods of different stability, as structural changes are difficult to identify in real-time. In this study, we observe that the combined estimated trend fitness to contemporaneous prices is generally less sensitive to a change of volatility conditions in the market than that of individual estimated trends if mean error measurements are considered (see *subfigure 8.b* of Fig. 8) but considering only the number of periods with minimal deviation from prices, individual and combined estimates are equally sensitive (see *subfigure 6.b* of Fig. 6).

τ_{METS} is the forecast with the smallest mean error measurements and the highest frequency of minimal deviation from prices. It is also quite insensitive to market changes. Moreover, τ_{METS} is also the top scorer in fitting prices in 1- to 6-month horizons. Thus, trying the hybrid version of the METS ensemble (METS 2.0) we propose in this study would be worthwhile using different weighting systems. In Annex *Figure A.5.d*, we compare some alternative arrangements of METS.

Results attest that the accuracy improvement combinations can provide in terms of error measures and frequency of minimal deviations from target, and aligns with Bates and Granger (1969) and Armstrong (2001), among many others, in supporting that linear combinations reduce the average error of the individual estimates. In general, whenever a combination achieves the same precision as an individual method, it prevails as it is less risky (Lemke and Gabrys, 2010).

However, the puzzling generalised optimality of simple averages in the literature is not always the case against other combinations or individual estimates, as we evidenced in this study where a simple average combination forecast (τ_{mean}) performs poorly compared to most of the other combinations used in this study and does not even outperform the individual τ_{symMA} in fitting prices contemporaneously according to four error measurements and SNR, but also underperforms in forecasting prices according to the MSFE.

We follow a second criterion to assess our individual and combined trend forecasts, considering different horizons. The results show that price indicators or forecasts commonly used in trend-trading strategies perform well only for the next period. However, generally, estimated trends (or trend forecasts) perform better as trading indicators for longer-term horizons, particularly the individual forecast τ_{MODWT}, which

Table 9
Forecasting Capacity in the Trend Reversion Strategy.

Trading Indicator	within 95 % return limits*				below lower 2.5 % limit*				above upper 2.5 % limit*				Extreme returns bias**				Profit Probability				
	h= 1	*h*= 3	*h*= 6	*h*= 12	*h*= 1	*h*= 3	*h*= 6	*h*= 12	*h*= 1	*h*= 3	*h*= 6	*h*= 12	*h*= 1	*h*= 3	*h*= 6	*h*= 12	*h*= 1	*h*= 3	*h*= 6	*h*= 12	
9. a Individual Trend Forecasts																					lower
Cubic	0.9496	0.9528	0.9620	0.9796	0.0302	0.0260	0.0198	0.0092	0.0202	0.0212	0.0183	0.0112	-0.0099	-0.0049	-0.0015	**0.0020**	0.4809	0.5074	0.5302	**0.5744**	
symmMA	0.9528	0.9737	0.9785	0.9972	0.0313	0.0168	0.0143	0.0021	0.0159	0.0094	0.0071	0.0007	-0.0154	-0.0074	-0.0072	-0.0014	0.4583	0.4832	0.4884	0.5113	
HP	0.9507	0.9625	0.9751	0.9940	0.0324	0.0262	0.0175	0.0036	0.0170	0.0113	0.0074	0.0024	-0.0154	-0.0149	-0.0101	-0.0011	0.4606	0.4586	0.4747	0.5467	
MODWT	0.9517	0.9532	0.9559	0.9634	0.0241	0.0219	0.0229	0.0186	0.0243	0.0249	0.0212	0.0180	0.0002	**0.0029**	-0.0016	-0.0006	**0.5156**	**0.5355**	0.5294	0.5513	
SS2000	**0.9890**	**0.9934**	0.9954	0.9938	**0.0058**	**0.0043**	0.0035	0.0053	0.0052	0.0022	0.0010	0.0010	-0.0005	-0.0021	-0.0025	-0.0043	0.5126	0.4910	0.4744	0.4639	
9. b Combined Trend Forecasts																					
Mean	0.9599	0.9533	0.9637	0.9698	0.0225	0.0287	0.0230	0.0197	0.0176	0.0180	0.0133	0.0105	-0.0049	-0.0107	-0.0097	-0.0093	0.4953	0.4859	0.4930	0.5066	
Median	0.9628	0.9649	0.9815	0.9929	0.0250	0.0226	0.0122	0.0044	0.0122	0.0125	0.0062	0.0027	-0.0128	-0.0101	-0.0060	-0.0017	0.4581	0.4792	0.4924	0.5371	
Weighted	0.9508	0.9627	0.9763	0.9908	0.0330	0.0227	0.0138	0.0045	0.0162	0.0147	0.0100	0.0046	-0.0168	-0.0080	-0.0038	0.0001	0.4550	0.4907	0.5152	0.5688	
Mult.Regress.	0.9548	0.9698	0.9814	0.9988	0.0272	0.0172	0.0115	0.0009	0.0180	0.0130	0.0071	0.0003	-0.0092	-0.0042	-0.0044	-0.0006	0.4807	0.5051	0.5063	0.5181	
PCA	0.9500	0.9555	0.9711	0.9830	0.0306	0.0237	0.0161	0.0100	0.0193	0.0208	0.0128	0.0070	-0.0113	-0.0029	-0.0032	-0.0030	0.4758	0.5143	0.5211	0.5359	
METS	0.9627	0.9795	**0.9978**	**0.9991**	0.0231	0.0113	**0.0014**	**0.0006**	0.0142	0.0092	0.0008	0.0003	-0.0088	-0.0020	-0.0005	-0.0003	0.4766	0.5142	0.5251	0.5527	
Network	0.9499	0.9556	0.9567	0.9742	0.0291	0.0248	0.0259	0.0147	0.0210	0.0196	0.0173	0.0111	-0.0081	-0.0052	-0.0086	-0.0036	0.4870	0.5055	0.5022	0.5359	
9. c Price Forecasts																					higher
Returns MA	0.9531	0.9567	0.9786	0.9976	0.0171	0.0235	0.0145	0.0022	**0.0298**	0.0199	0.0069	0.0001	**0.0127**	-0.0036	-0.0075	-0.0021	0.4870	0.5114	0.4857	0.4244	
Roll-over	0.9500	0.9500	0.9500	0.9500	0.0250	0.0250	0.0250	0.0250	0.0250	**0.0250**	**0.0250**	**0.0250**	0.0000	0.0000	**0.0000**	0.0000	0.5148	0.5245	**0.5346**	0.5504	

*Limits are set by the Roll-over returns critical values at 95 % level of confidence.
**It is calculated in absolute terms as the difference between the probability above the upper limit and the probability below the lower limit.
Note: The highest values in each *h* horizon are highlighted in bold font with the exception of values below lower limit in which the lowest values are shown in bold font.

Table 10
Forecasting Capacity in the Trend Reversion Strategy.

Trading Indicator	*h*= 1	*h*= 3	*h*= 6	*h*= 12
10.a Individual Trend Forecasts				
Cubic	0.6545	0.6621	0.6916	0.7736
symmMA	0.6681	0.7454	0.7731	1.0081
HP	0.6613	0.7034	0.7575	0.9176
MODWT	0.6581	0.6628	0.6710	0.6967
SS2000	0.8479	0.9112	0.9619	0.9434
10.b Combined Trend Forecasts				
Mean	0.6850	0.6663	0.7022	0.7279
Median	0.7019	0.7073	0.7912	0.8998
Weighted	0.6631	0.6968	0.7552	0.8688
Mult.Regress.	0.6699	0.7235	0.7878	1.0961
PCA	0.6565	0.6702	0.7290	0.7975
METS	0.6974	0.7729	1.0233	1.1132
Network	0.6547	0.6709	0.6761	0.7442
10.c Price Forecasts				
Returns MA	0.6670	0.6739	0.7745	1.0819
Benchmark (roll-over)	**0.6533**	**0.6533**	**0.6533**	**0.6533**

is a top scorer in both trading strategies in terms of returns, efficiency and Sharpe ratio in horizons longer than 1-month (3-months in Trend-Reversion and 6- and 12-months in Momentum); validating its inclusion in the selection. However, there are also other top-performing forecasts: τ_{cubic} and $\tau_{weighted}$ in a 12-month Trend-Reversion strategy, and τ_{METS} in a 3-month Momentum one. Additionally, trend forecasts perform better in periods of market unrest as price forecasts may be subject to bubbles and bursts.

Though no one trend forecast clearly outruns the others in both strategies, across all the horizons in all the aspects of the second assessment, two wavelet-based forecasts, τ_{MODWT} (individual) and τ_{METS} (combined), perform well overall. Also, results do suggest clearly that strategies targeting the long term (position trading, direct investment, and public policy), should be implemented with long-term trend forecasts and not operators on current prices, which may only be effective in the short-term as the first can increase the forecasting capacity of the later in longer periods improving the probability of profits and avoiding extreme negative returns in those periods.

The sophistication of some of the forecasting methods presented in this study may entail an initial cost of implementation, and some others may not have a straightforward interpretation for traders. However, despite its relative complexity, top scorers like METS 2.0 and MODWT follow simple algorithms that are traceable and provide intuitive results that make it worth the work not only for position trading but, most importantly, for other long term decisions like project investment, and policy-making; and when targeting long-term variables (economic, climate, etc.), in particular, unobservable ones. Further research may include a super meta-learning process with combined trend forecasts to improve long-term strategies. We also recognise that energy projects and energy policy entail decades of planning. Then, the 12-month horizon in this study needs to be stretched to 12 years, either testing how far our current forecasts can go or using annual data as we would be looking for an even lower-frequency forecast. Using annual data may have the challenge of a smaller sample. However, given the results of Díaz-Rodríguez and Robles (2026) frequency analysis of energy commodities, there may be good prospects for it as the essential aspects of the trend are retained. Moreover, the notorious seasonality in the price series of Natural Gas and Gasoline would be diluted in our 1-year-horizon trend approach and forecasts may work just like it does with commodities with a much less evident seasonality like CL prices.

Another strand to explore for future research may also be the integration of both long-term and short-term forecasts to cater to different trading strategies.

7. Concluding remarks

This study investigates methods for estimating trends in time series data, focusing on their application in long-term forecasting. Combining individual estimates generally improves accuracy and reduces risk compared to relying on a single method, particularly when forecasting unobservable variables. The combination of forecasts with METS 2.0 emerges as a new promising approach as it turns out to be the most effective in terms of fitting and minimizing errors while remaining stable across different market conditions.

For long-term strategies, relying on trend forecasts outperforms using current prices as indicators. This highlights the limitations of short-term price forecasts and emphasises the value of dedicated trend estimation methods for long-term predictions.

Denoising the original data may not be necessary for all estimation methods, especially when dealing with low-frequency data and/or long-term forecasts. This finding reshapes the recent practice of denoising in hybrid forecasting methods.

The current literature on energy price forecasts heavily focuses on short-term horizons. This study extends the forecast horizon for energy prices 12 months ahead. It emphasizes the need for more research on long-term forecasting methods, which are crucial for informed decision-making in various contexts, such as investment or public policy.

Finally, although we have observed the consistent economic outperformance of the METS 2.0, there are additional extensions of the analysis that can provide valuable insights. First, it would be useful to analyze the model's performance across different market regimes (e.g., pre-2008 and the COVID-19 pandemic) and determine whether there are specific economic or market behaviors that could explain the results. Additionally, testing the models with analysis based on simulated data could strengthen the comparison across trend models and combination methods.

Authors Statement

F. Díaz-Rodríguez: Conceptualization, Methodology, Programming, Validation, Formal analysis, Investigation, Writing - Original Draft, Data. Visualization. **M.-D. Robles:** Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge financial support from the Ministerio de Ciencia, Innovación y Universidades (Project Reference: PID2023-146832OB-I00). We are indebted to the editor and anonymous reviewers for their insightful and constructive comments, which greatly improved the manuscript. We also thank the discussants and participants at CEMA2024, ISF2024, CFE–CMStatistics2024, and the IV Thesis Workshop QFE2024 for their valuable feedback, with particular appreciation to Professors S. Ouzan, P. Six, and C. Frau. Special thanks go to P. Sullivan and H. Till for their helpful remarks, and to A.-M. Fuertes for her support. Any remaining errors are the sole responsibility of the authors.

Appendix

Table A1
Acronyms and trend estimates

List of acronyms		*Trend estimates*	
Acronym	*Name*	*Trend*	*Model*
ANN	Artificial Neural Network	τ_{cubic}	Cubic Polynomial Detrend Method
CL	Crude Oil	τ_{symMA}	Symmetrical or Centred Moving Average
EMD	Empirical Mode Decomposition	τ_{SS2000}	Schwartz and Smith, (2000) long-term component
EoM	End-of-Month	τ_{HP}	Hodrick-Prescott Trend
FFNN	Feed Forward Neural Network	τ_{MODWT}	Maximal Overlap Discrete Wavelet Approximation
HP	Hodrick-Prescott		
IMF	Intrinsic Mode Function	*Trend*	*Forecast combination method*
MA	Moving Average	τ_{mean}	Simple Mean
MAE	Mean Absolute Error	τ_{median}	Median
MAPE	Mean Absolute Percent Error	$\tau_{weighted}$	Weighted by the frequency of minimal deviation from prices
METS	wavelet-based Multiscale Ensemble Time-scale	τ_{MLR}	Multi-linear Regression
MLR	Multi-linear Regression	τ_{PCA}	First component of the Principal Components Analysis
MODWT	Maximal Overlap Discrete Wavelet Transform	τ_{METS}	Wavelet-based Multiscale Ensemble Time-Scale
MSE	Mean Squared Error	$\tau_{network}$	Feed Forward Neural Network
MSFE	Mean Squared Forecast Error		
NYMEX	New York Mercantile Exchange		
PCA	Principal Components Analysis		
P&L	Profit & Loss		
RMSE	Root Mean Squared Error		
SS2000	Schwartz and Smith, (2000) model		
WT	Wavelet Transform		

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.egyr.2025.108941.

Data availability

Data will be made available on request.

References

Almaita, E., Alshkoor, S., Abdelsalam, E., Almomani, F., 2022. State of charge estimation for a group of lithium-ion batteries using long short-term memory neural network. J. Energy Storage 52, 104761.
Armstrong, J.S. (Ed.), 2001. Principles of forecasting: a handbook for researchers and practitioners, 30. Kluwer Academic, Boston, MA.
Armstrong, J.S., Green, K.C., Graefe, A., 2015. Golden rule of forecasting: Be conservative. J. Bus. Res. 68 (8), 1717–1731.
Bates, J.M., Granger, C.W., 1969. The combination of forecasts. J. Oper. Res. Soc. 20 (4), 451–468.
Blanc, S.M., Setzer, T., 2016. When to choose the simple average in forecast combination. J. Bus. Res. 69 (10), 3951–3962.
Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.
Cang, S., Yu, H., 2014. A combination selection algorithm on forecasting. Eur. J. Oper. Res. 234 (1), 127–139.
Cashin, P., McDermott, C.J., 2002. The long-run behavior of commodity prices: small trends and big variability. IMF Staff Pap. 49 (2), 175–199.
Chan, F., Pauwels, L.L., 2018. Some theoretical results on forecast combinations. Int. J. Forecast. 34 (1), 64–74.
Chaves, D.B., Viswanathan, V., 2016. Momentum and mean-reversion in commodity spot and futures markets. J. Commod. Mark. 3 (1), 39–53.
Claeskens, G., Magnus, J.R., Vasnev, A.L., Wang, W., 2016. A simple theoretical explanation of the forecast combination puzzle. Int. J. Forecast. 32 (3), 754–762.
Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. International journal of forecasting 5 (4), 559–583.
Conlon, T., Cotter, J., Eyiah-Donkor, E., 2024. Forecasting the price of oil: A cautionary note. J. Commod. Mark. 33, 100378.
Crowley, P.M., 2007. A guide to wavelets for economists. J. Econ. Surv. 21 (2), 207–267.
Denk, S., Löffler, G., 2024. Predicting the Equity Premium with Combination Forecasts: A Reappraisal. Rev. Asset Pricing Stud. raae009.
Ding, X., Fu, L., Ding, Y., Wang, Y., 2022. A novel hybrid method for oil price forecasting with ensemble thought. Energy Rep. 8, 15365–15376.
Emekter, R., Jirasakuldech, B., Went, P., 2012. Rational speculative bubbles and commodities markets: application of duration dependence test. Appl. Financ. Econ. 22 (7), 581–596.
Harvey, A., 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.
Harvey, A., 2006. Forecasting with unobserved components time series models. Handb. Econ. Forecast. 1, 327–412.
Herawati, S., Djunaidy, A., 2020. Implementing Method of Empirical Mode Decomposition based on Artificial Neural Networks and Genetic Algorithms for Crude Oil Price Forecasting (July). In: In Journal of Physics: Conference Series, 1569. IOP Publishing, 022075 (July).
Hodrick, R.J., Prescott, E.C., 1997. Postwar US business cycles: an empirical investigation. J. Money Credit Bank. 1–16.
Houthakker, H.S., 2017. Normal backwardation. In Value, capital and growth. Routledge, pp. 193–214.
Hsiao, C., Wan, S.K., 2014. Is there an optimal forecast combination? J. Econ. 178, 294–309.
Huang, H., Lee, T.H., 2010. To combine forecasts or to combine information? Econom. Rev. 29 (5-6), 534–570.
Jacks, D.S., 2019. From boom to bust: A typology of real commodity prices in the long run. Cliometrica 13 (2), 201–220.
Jose, V.R.R., Winkler, R.L., 2008. Simple robust averages of forecasts: Some empirical results. Int. J. Forecast. 24 (1), 163–169.
Kaiser, R., Maravall Herrero, A., 2002. A complete model-based interpretation of the Hodrick-Prescott filter: spuriousness reconsidered. Banco de España, Servicio de Estudios.
Kalman, R.E., 1960a. A New Approach to Linear Filtering and Prediction Problems. J. Basic Eng. 82.
Kalman, R.E., 1960b. Contributions to the theory of optimal control. Bol. Soc. Mat. Mex. 5 (2), 102–119.
Kalman, R.E., 1960c. On the general theory of control systems. Proceedings First International Conference on Automatic Control. USSR, *Moscow*, pp. 481–492.
Khormali, A., Ahmadi, S., Aleksandrov, A.N., 2025. Analysis of reservoir rock permeability changes due to solid precipitation during waterflooding using artificial neural network. Journal of Petroleum Exploration and Production Technology 15 (1), 17.
Kourentzes, N., Barrow, D., Petropoulos, F., 2019. Another look at forecast selection and combination: Evidence from forecast pooling. Int. J. Prod. Econ. 209, 226–235.
Lemke, C., Gabrys, B., 2010. Meta-learning for time series forecasting and forecast combination. Neurocomputing 73 (10-12), 2006–2016.
Lu, H., Ma, X., Ma, M., Zhu, S., 2021. Energy price prediction using data-driven models: A decade review. Comput. Sci. Rev. 39, 100356.
Maravall, A., Del Rio, A., 2001. *Time aggregation and the Hodrick-Prescott filter* (No. 0108). Banco De. España.
Díaz-Rodríguez, F., Robles, M.-D. (2026). In Search of a Low-Frequency Trend in Energy Commodity Prices. *Available at SSRN 4392351*.

Narayan, P.K., Ahmed, H.A., Narayan, S., 2015. Do momentum-based trading strategies work in the commodity futures markets? J. Futures Mark. 35 (9), 868–891.
Nilsson, L. (2015). Trend Following-Expected Returns. Available at SSRN 2689861.
Palm, F.C., Zellner, A., 1992. To combine or not to combine? Issues of combining forecasts. J. Forecast. 11 (8), 687–701.
Percival, D.B., Senior, K.L., 2012. A wavelet-based multiscale ensemble time-scale algorithm. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 59 (3), 510–522.
Provornaya, I.V., Filimonova, I.V., Nemov, V.Y., Komarova, A.V., Dzyuba, Y.A., 2020. Features of the petroleum products pricing in Russia, in the USA, and Saudi Arabia. Energy Rep. 6, 514–522.
Rallis, G. (2010). On commodity trading strategies: momentum, term structure, maturity, indexation (Doctoral dissertation, City, University of London).
Saghi, F., Jahangoshai Rezaee, M., 2021. Integrating wavelet decomposition and fuzzy transformation for improving the accuracy of forecasting crude oil price. Comput. Econ. 1–33.
Salamai, A.A., 2023. Deep learning framework for predictive modeling of crude oil price for sustainable management in oil markets. Expert Syst. Appl. 211, 118658.
Schwartz, E., Smith, J.E., 2000. Short-term variations and long-term dynamics in commodity prices. Manag. Sci. 46 (7), 893–911.
Shabri, A., Samsudin, R., 2014. Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model. Math. Probl. Eng. 2014.
Stock, J.H., Watson, M.W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of forecasting 23 (6), 405–430.
Szarek, D., Bielak, Ł., Wyłomańska, A., 2020. Long-term prediction of the metals' prices using non-Gaussian time-inhomogeneous stochastic process. Physica A Statistical Mechanics Applications 555, 124659.
Timmermann, A., 2006. Forecast combinations. Handb. Econ. Forecast. 1, 135–196.
Timmermann, A., 2018. Forecasting methods in finance. Annual Review of Financial Economics 10 (1), 449–479.
Tsai, S.B., Xue, Y., Zhang, J., Chen, Q., Liu, Y., Zhou, J., Dong, W., 2017. Models for forecasting growth trends in renewable energy. Renew. Sustain. Energy Rev. 77, 1169–1178.
Vilar, J., Casal, R.F., Fernández-Lozano, C., 2020. Covid-19 projections for Spain using forecast combinations. *BEIO*. BoletíN. De. Estad. ística e Invest. óN. Oper. 36 (2), 99–125.
Wang, J., Li, X., Hong, T., Wang, S., 2018. A semi-heterogeneous approach to combining crude oil price forecasts. Inf. Sci. 460, 279–292.
Wang, J., Wang, Z., Li, X., Zhou, H., 2022. Artificial bee colony-based combination approach to forecasting agricultural commodity prices. Int. J. Forecast. 38 (1), 21–34.
Wang, X., Hyndman, R.J., Li, F., Kang, Y., 2023. Forecast combinations: an over 50-year review. Int. J. Forecast. 39 (4), 1518–1547.
Wei, Y., Yu, B., Guo, X., Zhang, C., 2023. The impact of oil price shocks on the US and Chinese stock markets: A quantitative structural analysis. Energy Rep. 10, 15–28.
Zhang, Y., Wang, Y., 2023. Forecasting crude oil futures market returns: A principal component analysis combination approach. Int. J. Forecast. 39 (2), 659–673.