MEASURING THE COMPETITIVENESS OF COMMODITY MARKETS USING PRICE SIGNALS AND INFORMATION THEORY

Anis Hoayek[1] and Hassan Hamie[2*]

[1] Institut Henri Fayol, École nationale supérieure des mines de Saint-Étienne, Saint-Etienne, France
[2] Energy Economics Group, Vienna University of Technology, Vienna, Austria

Original scientific paper

Abstract: *Technological advancements, abrupt changes in market conditions, and political reforms, among other things, necessitate strong regulatory oversight, and accurate measurement of performance related indicators. The more accurate, information rich, and transparent these measurements/signals, the lower the level of uncertainty felt by value chain participants, who are thus able to recognize and observe whether the market's state is efficient. Its lack, may lead to indecisiveness, translating into false interpretations that could lead to wrong policy directions. This paper provides an ex-post evaluation tool intending to deliver additional insights or quality information that would aid the regulator in assessing the state of the market. The tool is applied to the UK wholesale natural gas market for the period between 2011 and 2020, assessing and testing the market's weak-form efficiency. It claims that today's gas prices reflect a specific type of information, primarily past gas prices, and that only new information can help predict future prices. In this manuscript, based solely on a limited and available untapped dataset (day-ahead price time series), and working under the assumption that gas prices are the result of market processes, a variety of information metrics (gas price randomness, distribution of extreme prices, ability to predict prices - based on historical sets) is extracted with the use of suitable mathematical statistical models. A weighted entropy index is then computed, and measures the state of the commodity market. The results indicate that the analysis has helped gain information, thus reducing uncertainty (relative to a pre-analysis) by 86.5 %. Additionally, there is sufficient evidence that the UK natural gas prices are weak-form efficient.*

Key words: *Markets efficiency, price signals, information theory, commodity markets.*

* Corresponding author.
E-mail addresses: anis.hoayek@emse.fr (A. Hoayek), hasan.hamieh@hotmail.com (H. Hamie)

1. Introduction

Trading goods is an exchange between the owner and another entity for money. This can occur immediately or at a certain time in the future (future contracts). The latter form of contract is gaining grounds, having as aim, to hedge or make speculative profits. International institutions, such as the World Bank, produce commodity market outlooks that provide market analysis for major commodity groups (energy, metals, and others). These reports can forecast prices for up to 46 commodities, including energy. Energy regulators at the national level conduct additional market analysis as they continuously work to ensure fair treatment of all market participants, thereby enabling competition, to drive down prices, protect consumer' interests, and conduct surveys to measure the performance of value chain participants and others. These regulator's actions result in reports on the following dimensions: competition, customer affordability, costs of balancing systems, and so on. For example, the report on the state of the energy market, produced by Great Britain's independent energy regulator, OFGEM, is an important annual report that contains the following: information on supplier profit margins, number of consumers who report not having to change suppliers, average wholesale day–ahead prices, and year-on-year maximum demand during winter, among others.

Such metrics analytics can be used to understand how difficult it is for countries to stabilize wholesale and retail commodity prices (at the national level), the existence of international cartels (at the international level), and the complex relationship between governments of exporting and importing countries (in terms of contract agreements, quota agreements, etc.). These reports and outlooks are aimed at most commodity value chain participants, including futures and speculation traders. As a result, regulators conduct quantitative and qualitative means market analysis. These measures are extensive and frequently based on data that includes prices, quantities (supply and demand), and tariffs. The quantitative analyses range from the reading and analysis of descriptive statistics in the data to the development of in–house statistical methods for forecasting commodity prices under various scenarios, and identifying the risks associated with such forecasts. The methods they use for their quantitative approach can be universally known, accepted, and applied everywhere by most stakeholders, or they can be market specific.

An example of a quantity that competition regulators commonly use is the Herfindahl-Hirschman Index (HHI); it is used to measure market concentration. This method is internationally accepted as an indicator of the competition level. The information required for calculating the HHI is normally kept in-house and specific to the needs of the regulators and is not publishable information reflecting the state of the market, especially if the information contains commercial sensitive information.

The suggested methodology in this paper complements the work done by commodity regulators and provides an ex-post evaluation tool to deliver additional insights into the functioning of markets. It feeds into the activity of measuring performance and ensures that regulators meet their objectives. The Authors will analyze the country-specific individual wholesale market signals by suggesting an in-depth analysis of price changes using mathematical and statistical methods complementing each other. The aim is to extract information relevant to market participants and test the validity of the theory of efficient markets.

There is a broad definition of efficient markets in the literature. It primarily refers to liquid trade, or transparent prices (most importantly, exchange houses, clearing agencies, hubs, online platforms, etc., all show full information of the trades witnessed in a certain time period). It also refers to places where customers and traders have a

free choice of selection (anyone, person or entity, can buy/ sell from/ to anyone) and traders or suppliers have a fair rate, premium, or profits over their transactions/ deals. Moreover, the people or entities participating in the market trades tend to be short and long-term consumers of the commodity or hedgers in some cases, and less and less speculative in nature (Nick, 2016).

The theory of efficient markets states, from a conceptual and rather theoretical standpoint, that all past information in a given market, whether public or private, is discounted in current gas prices. There are numerous variations on this definition in the literature. The weak form of market efficiency will be tested in this manuscript, and it claims that a specific type of information, primarily pas gas prices, is reflected in today's gas prices. Considering historical prices, no additional technical analysis can provide traders with information about future prices. Only new information can assist in predicting future prices. This leaves no room for any value chain participant to exploit and predict future prices, to profit with minimal risk. This definition is similar to the random walk theory, which states that no past movement of gas prices can be used to predict future price movement; thus, there is an equal chance that gas prices will rise or fall in the future.

There is no such thing as weak-efficient market for a particular commodity. In both space and time, the degree of efficiency is not absolute. Some markets are more efficient than others, and are distinguished by commodity liquidity, among other characteristics that make information symmetric to some extent. Even if the flow of information becomes asymmetric at times due to the presence of liquidity, trading options, such as arbitrage, emerge, and restore efficiency to acceptable levels.

Given assumption that prices are the result of market processes, the authors believe that there is a wealth of information available that is currently untapped. Our contribution is to investigate the information contained inherent in prices by measuring the following signals, for a given market, with a focus on the UK wholesale natural gas hub. The following signals will be tested:

- The Shannon entropy signal enables capturing the level of uncertainty (randomness) of gas prices time series without imposing any constraints on its theoretical probability distribution.
- Efficient markets assume commodity prices normalcy, so there is minimal probability of an extreme event occurring above a certain level. Most energy markets, are inefficient in terms of extreme events and are thus heavily influenced by unpredictable extreme prices, leaving their price distribution vulnerable to fat tail risks. Thus, the record theory will be used to understand the behavior of tail risk, and the distribution of extreme prices will be evaluated. The authors specifically want to determine whether the underlying observations of a records series are independent and identically distributed, unlike fat tails.

The Shannon and records theories are relevant to our analysis and regulators, especially in markets where available data on supply and demand is scarce. By scarce, it is not meant absence, nor that any actor does not observe the information. However, it is meant that the flow of information is not symmetric and unavailable for all value chain participants at all times. Such data are critical to the observation and analysis as they produce market equilibrium prices. This document will show that robust results can be obtained without requiring full information about the process used to generate equilibrium gas prices.

- The predictability of daily gas prices will be evaluated using parametric models based on the history of a gas price return time series.

- Nonparametric deep learning models will be used to assess the predictability of daily gas prices based on the history of a gas price return time series.

The weighted measure of added information will be computed after the market signals have been extracted and measured. This information will provide the regulator with more evidence about the state of the market, allowing it to better perform its market oversight duties.

This paper provides an ex-post evaluation tool to the gas value chain participants, with the aim to deliver additional insights into the functioning of markets. The tool is essentially composed of several mathematical and statistical methods complementing each other, and the result of which is an index-value (Mutual Information, MI), ranging from zero to one. In case the value approaches 1, this strongly suggests that the gas market under study exhibit prices that are weak-form efficient. In the other extreme case (MI equals to zero), implies non-efficiency in the gas market. To test the tool, an empirical illustration is performed on the UK market, more specifically in the period that spans from 2011 to end of 2020. In an attempt to test the performance of the tool and compare our results to the existent evidence on the state of the UK gas market found in recent literature (briefly summarized in section three). The remainder of this paper is structured as follows. Section 2 defines the materials and methods and presents related literature. Section 3 describes the case, presents the data and results, and interprets them. Section 4 concludes the paper by making recommendations for future research.

2. Methods

Main Information was extracted from different resources investigating the data. These resources are called signals and are denoted by $S_i, i = 1, \ldots, I$, where S_i is the ith signal and I is the total number of considered signals. To analyze the performance of each signal, one must explore a family of models denoted by F_i = "the set of possible models that can be applied in order to extract a measure of signal S_i". The selection of the elements of each family of models is based on three precise criteria: literature, data availability/structure, and expert/modeler opinion. Subsequently, the best suitable model is selected based on mathematical and statistical (e.g., statistical tests) analysis, and this optimal model within family F_i is denoted by M_i, $i = 1, \ldots, I$. Once the best model is selected, the result is either conclusive or nonconclusive. Certainly, the user can choose to have a non-binary value and measure the degree of conclusiveness; however, to the best of the author's knowledge and understanding, this adds another level of subjectivity and complexity.

The above finding is denoted by the following expression:

$$C_i = \left\{ \begin{array}{l} 1 \; if \; M_i \; is \; conclusive \\ 0 \; otherwise \end{array} \right. \tag{1}$$

In the case of $C_i = 1$, one should reflect on the selection of M_i: was it well justified in terms of closed mathematical and statistical arguments or was the optimal model selected partially based on subjective assumptions? To assess this, the following binary variable is introduced:

$$J_i = \left\{ \begin{array}{l} 1 \; if \; M_i \; is \; justified \\ 0 \; otherwise \end{array} \right. \tag{2}$$

Now, to measure the efficiency of the market given signal S_i, the random variable describing the efficiency of the market is described by the following probability distribution (See Eq. (3). and Eq. (4).):

$$(\mathbb{P}\left[Efficient/Signal_i\right], \mathbb{P}\left[not - Efficient/Signal_i\right]) \tag{3}$$

with:

$$\mathbb{P}\left[Efficient/Signal_i\right] = 1 - \mathbb{P}\left[not - Efficient/Signal_i\right] \tag{4}$$

The data gathered on the decision variables thus far will result in the estimation of the probabilities in all possible scenarios for signal S_i, as presented in Table 1.

Table 1. $\mathbb{P}\left[Efficient/Signal_i\right]$ (or $\mathbb{P}\left[not - Efficient/Signal_i\right]$) in all possible scenarios.

		C_i	
		0	1
J_i	0	$\dfrac{1}{2}$	$\dfrac{3}{4}$ $(or\ \dfrac{1}{4})$
	1	Not available	$\dfrac{4}{4}$ $(or\ 0)$

The authors deal with the classical discrete distribution; hence, the classical Shannon distribution can be used to extract the amount of information related to signal S_i by computing the corresponding entropy. In the case of conclusive results $C_i = 1$, regardless of how justified the model is, there is additional information, and the regulator can measure such an improvement based on significant (more discriminatory) probability of efficiency $\mathbb{P}\left[Efficient/Signal_i\right] \neq 0.5$. In other words, in a conclusive context the information about how efficient the market is, brought by the corresponding signals is far from the uniform (less informative) case by a significant probability distance, i.e. (See Eq. (5).),

$$|\mathbb{P}\left[Efficient/Signal_i\right] - 0.5| \geq 0.25 \tag{5}$$

The case of $C_i = 0$, and $J_i = 0$ is inconclusive, and the selection of the optimal model is not justified and does not lead to concrete results. In this case, signal S_i does not add information about the efficiency of the market and $\mathbb{P}\left[Efficient/Signal_i\right] = 0.5$. Such a signal shall be eliminated from the analysis, or another theoretical aspect revealing more information regarding the efficiency of the market is to be investigated and exploited. Finally, the case of $C_i = 0$ and $J_i = 1$ cannot occur because a mathematically/statistically justified model is conclusive.

Moving forward in the analysis, a fair question arises: how to attribute and allocate weights to each signal? The latter is defined by the variable $W_i, i = 1, \ldots, I$, and can be defined by a multitude of approaches. In this context, the weights will be allocated based on the analytic hierarchy process (AHP) model.

Finally, to show whether and how these different signals improve the quality of the information regarding efficiency, the difference between information less entropy (based on a uniform prior distribution), denoted by $\mathcal{H}_{prior}$, and the obtained aggregated (weighted) post entropy $\mathcal{H}_{posterior}$ is calculated. This quantity is called the mutual information (MI) and translates into measuring the reduced amount of uncertainty; it is also defined as shown in Eq. (6).

$$MI = \mathcal{H}_{prior} - \mathcal{H}_{posterior} \tag{6}$$

The classical approach of the decision theory, AHP, and the MI, will be defined and explained in the next sections.

2.1. Measure of Shannon entropy (Signal 1)

Shannon entropy as a signal studies the stability of the underlying time series in terms of probabilistic distribution by providing information on the amount of hidden uncertainty in a given random variable (Lesne, 2014). The higher the uncertainty in the short-term, the more efficient the market, with the corresponding time series, tends to be. First, the prices' rates of returns are calculated, and a binary random variable r_i is generated, and expressed in Eq. (7).

$$r_i = \begin{cases} 1 \ if \ return \ is \ positive \\ 0 \ if \ return \ is \ nonpositive \end{cases} \tag{7}$$

The original commodity prices should not be used due the following reasons. Prices must be non-negative, whereas returns or log-returns can be any value, making them easier to model. Furthermore, prices typically have a unit root, whereas returns are more likely to be stationary. It is well known that stationary time series have many convenient properties for econometric analysis. A non-stationary, time series means that the moments will change over time. For example, in the case of prices, the mean and variance would both be affected by the previous period's price. Taking the percent change (or log difference) removes this effect often. This is a common measure, in which the rates of return are calculated and the logarithmic function is applied (Meucci, 2011). It informs the user that there is a stable trend and that the returns are close to zero. This could imply that the behavior is stationary and stable.

The random variable under consideration is simply the time series of price returns denoted by X in Eq. (8). The random variable X's entropy is defined as follows:

$$H(X) = -\frac{1}{log_2(n)} \sum_{i=1}^{n} p_i * log_2(p_i) \tag{8}$$

where each outcome i (in our case return labels) has a probability p_i of appearance in the series, with n being the total number of outcomes. This is commonly used to describe a discrete set of data. When entropy approaches 1, the underlying uncertainty is maximized and the random variable under consideration behaves like a uniform distribution. However, when entropy is close to 0, it almost certainly implies a context where a Dirac distribution can be considered. It should be noted that Dirac is a probabilistic distribution that assigns full probability to one event in a space of events. In other words, there exists an event A for a Dirac distribution such that $P[A] = 1$. This means that event A will occur, and there is no fuzziness in explaining the random variable's behavior. In this case, the regulator will use this information to determine whether the market is efficient, which is critical for the latter to plan accordingly.

The authors have also considered different sequences $Seq_i = \{1, 2, 3\}$ of negative and positive returns r_i over one, two, and three consecutive days/observations. In the first case ($S_{eqi} = 1, one \ day$), the space of possible outcomes is limited to two (either one positive return or one negative); in the case of two consecutive days occurring at time t and $t + 1$ ($Seq_i = 2$), there exist four possible outcomes to analyze ($r_t = 0, r_{t+1} = 0$); ($r_t = 0, r_{t+1} = 1$); ($r_t = 1, r_{t+1} = 0$); ($r_t = 1, r_{t+1} = 1$), thus the space of possible outcomes increases exponentially with the number of consecutive days considered. The interest here lies not in predicting if the return will end up as negative or positive rather in detecting patterns over a wider temporal sequence of returns, thus improving the detection and formation of gas price trends.

Thus, Equation 8 is performed thrice, and for each set of sequences defined as, Seq_i, the efficiency measure is computed. In efficient market theory, the entropy H shall be as close to unity, when the n possible outcomes are equally probable, $p_{outcome} = \frac{1}{n}$.

Alternatively, when one outcome occurs at a higher (relative to other outcomes) probability, a certain pattern exists, and this will attract some key value chain players, such as investors, to conduct technical analysis, find the pattern, and exploit the market, thus leading to nonefficiency.

The rolling time window analysis can be used to examine the temporal evolution of gas market efficiency. Starting with the first 30 observations, the next window is selected by moving forward across time, removing the first observation, and adding a new observation to the most recent observation in the previous window, all while keeping the number of observations in a window to 30. Entropy is computed for each window.

Furthermore, to cross-validate the results, the authors divided the total number of windows into two subsets, training and testing (70%/ 30% split). The nonparametric Kolmogorov Smirnov test is used at each cross-validation step to determine whether the entropy values on the training and testing samples are identically distributed (drawn from the same underlying distribution). Then, if the $p - value$ of the test is greater than 5% (i.e., the H_0 of the Kolmogorov Smirnov test is reasonably accepted, implying that both samples are drawn from the same distribution), one can conclude that there is entropy coherence between the training and testing distributions. In addition, for more robustness checks, the time dependency factor is relaxed, and a random selection of windows is chosen and divided into validation and testing subsets. This logic of testing similarity between random training and testing samples is applied and performed 100 times, and the proportion/ratio of accepted H_0 is calculated across all 100 cross-validation trials. As a result, the higher the ratio, the more certain and robust our entropy measure.

2.2. Measure of the extremes (Signal 2)

The long-term stability of de-trended prices is best analyzed by studying their extremes, i.e., the tail of the distribution. The Models most used are extreme value theory, record theory, long-term growth model, ANOVA, a measure of mean reversion business cycle, etc. The authors think that the extreme value and record theories are the most useful because they produce robust results without requiring any knowledge of the process used to generate the underlying data (which is also unavailable for our application). By focusing solely on the extreme observations, it is assumed that both approaches reduce the bias caused by an entire distribution. Furthermore, these approaches take the effect of heavy tails into account, which is not explicitly considered in other statistical approaches. The record theory has the advantage of producing exact (Arnold et al., 1998) and non-asymptotic results, (Lindström & Regland, 2012). Furthermore, the results are distribution-free, so the record theory does not require the choice of a specific underlying distribution ex ante.

To capture the variability, the model started by calculating the absolute value of the first difference in the price observations. The number of upper records in the time series was then computed. The latter is defined as a higher observation (commodity price) than its proceeding observations, which is equivalent to saying the maximum observation up to this point. Subsequently, the authors need to determine whether the underlying observations were independent and identically distributed ($i.i.d$) and, if so, select the appropriate record model.

The statistic to test the null hypothesis that the data come from a sequence of $i.i.d$ random variables is computed, as per Eq. (9).

$$N = (N_T - \log \text{ T})/\sqrt{\log T} \tag{9}$$

where T denotes the length of the time series, and N_T is the number of records. It has been demonstrated that under the null hypothesis, N converges to a standard normal distribution denoted by $N(0, 1)$ (Hamie et al., 2018). Then, the null hypothesis cannot be rejected if N is less than the theoretical $(1 - \alpha)^{th}$ quantile of the standard normal distribution (α is the -confidence level of the statistic N generally fixed to 5%). Else, other record models more adapted to $non - i.i.d.$ observations shall be used. The probability that a certain observation in the future, at a certain time t, would qualify to be a record for the classic $i.i.d.$ model, and is denoted by Eq. (10).

$$P_t = \frac{1}{t} \tag{10}$$

This implies that in the presence of a commodity daily price time series that consists of 100 observations (3 months' interval) and aim to forecast the probability of having a record on the 101st observation, the probability would be equal to 0.0099. The probability of a new record in a classic model will thus converge to 0 in the long term, and records would be concentrated among the first observations. In line with the reasoning followed in the Shannon signal, a record-based statistical test is conducted on sliding windows of data with a fixed step. Each window w, where the $i.i.d.$ assumption is accepted, is labeled by 1 and 0 otherwise. Therefore, a decision variable similar to the Bernoulli random variable logic is created and expressed in Eq. (11).

$$D(w) = \begin{cases} 1 & \text{if } iid \text{ Hypothesis is accepted on window } w \\ 0 & \text{Otherwise} \end{cases} \tag{11}$$

The windows are then divided into training and testing sets following a cross-validation process. This is performed a hundred times, and on each of the cross-validation runs, another statistical nonparametric test is conducted. It checks if the random variable D has the same behavior on the training and testing window sets (test to see if the proportion of $D = 1$ is the same in both sets).

The test is shown in Eq. (12).

$$H_0: p_1 = p_2 \text{ versus } H_1: p_1 \neq p_2 \tag{12}$$

where p_1 denotes the proportion of $D = 1$ among the windows in the training sample, and p_2 is the proportion of $D = 1$ among the windows in the testing sample. The test statistic used to accept or reject the null hypothesis is as follows (See Eq. (13).)

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{13}$$

where n_1 and n_2 denote the number of windows in the training and testing sets, respectively, with

$$p = \frac{n_1 p + n_2 p_2}{n_1 + n_2} \tag{14}$$

Under H_0 one can show that z follows a standard normal distribution. Then, if $|z| > 1.96$, one may reject H_0 with an error risk of 5% (i.e., the behavior of record observations is not the same in the training and testing samples). Otherwise, when H_0 is accepted, there is coherence between the training and testing of the Bernoulli distributions. This logic of testing similarity between the training and testing samples is applied on all cross-validation runs. Then, the proportion of accepted H_0 among the cross-validation trials is calculated. Hence, a proportion close to 1, implies consistency in results, thus one can now compute the proportion of accepted $i.i.d.$ behavior (using the number of records on each window) over all windows.

2.3. Measurement of predictability using parametric models (Signal 3)

Classical learning methods are employed to estimate the coefficients of a regression model. These are commonly used to analyze high-frequency time series data, and the estimation is traditionally based on the optimization of a loss function. The analysis is then either constrained or governed by statistical assumptions, or solely based on machine learning and numerical techniques that are nonparametric. Two broad categories are thus used in the literature of commodity price prediction: parametric univariate/ multivariate and nonparametric models (Hamie et al., 2020).

The main continuous dataset measured and collected by regulators or competition authorities worldwide and for any commodity is composed of price time series in most application cases. As a result, the data are univariate, and the potential applicable models for price prediction are autoregressive in nature (AR, MA, ARMA, ARIMA, GARCH, etc.). The energy commodity price time series, in general, exhibit random work behavior (Narayan & Popp, 2010), making the task of predicting future values based solely on the series itself complex (Lee & Strazicich, 2003) and uncertain (Mishra & Smyth, 2016). Commodity price predictability implies that prices should follow a random walk process and exclude predictability based solely on past price movements (Bohl et al., 2021). To overcome this complexity, the return price series is considered, and a series of tests on the price returns is conducted.

The plot of the price return prices is visually examined first, and then two statistical tests [Augmented Dickey Fuller (ADF) and Kwiatkowski Phillips Schmidt Shin (KPSS)] are employed to check for stationarity. Accordingly, the modeler can choose which of the autoregressive models to use for return prediction. The authors can now investigate whether and how well future returns can be predicted based on previously calculated price returns. The data sample is divided into two subsamples for this purpose: a training set to train the model, and a testing set to measure the model's performance.

2.4. Measurement of predictability using nonparametric models (Signal 4)

Moving into the nonparametric analysis for autoregressive problems, the most popular model that outperforms almost all other autoregressive models is the long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997). This model has a deep neural network structure with a particular architecture that consider short and long-term impacts of observations on future value prediction. Like neural networks, and similar to other deep learning models, the LSTM coefficients are derived based on a gradient retro propagation process to minimize a given loss function (Pachón-Suescún et al., 2020). In addition, LSTM is a distribution-free model; this means there is no need to make any assumption about the probabilistic distribution of the underlying variables or of the errors.

For computational efficiency-related reasons, it is common in deep learning models to normalize/standardize underlying input and output variables. The Appendix (Figure A1) depicts a general representation of the LSTM model's architecture. In fact, the return price at time t, X_t, is combined with the output at time $t-1$, h_{t-1} (showing the short-term aspect of the model), as well as another input showing long-term information (price return selected up to a certain fixed window of time in the past) denoted by C_{t-1}, to generate the model's output h_t. It should be noted that the input and output are linked via a special nonlinear activation function and some mathematical operations (e.g., convolution and cross-correlation between underlying variables).

The selection of optimal values for the numerous hyperparameters is the LSTM model's main complexity (e.g., number of layers, number of neurons in each layer, and activation function, etc.). To address this complexity, some hyperparameters are computed using a grid search process (Claesen & De Moor, 2015), whereas others are selected manually and subjectively based on evidence found in literature and the author's knowledge and experience. The hyperparameters for the LSTM model are listed in Table 2.

Table 2. Hyperparameters of the LSTM model

Hyperparameters	Selected value
Number of hidden layers	Two hidden layers of the LSTM type
Number of neurons in the first hidden layer	2
Number of neurons in the second hidden layer	1
Optimizer	Stochastic gradient descent
Loss function	Mean absolute error
Batch size	40
Epochs	30
Window length	30

A word of caution needs to be mentioned as well, and that is to avoid the risk of overfitting that is faced in prediction models, more specifically deep machine learning models, with many tuning hyperparameters features. In fact, in machine learning, a very complex model can be as inaccurate (overfitting) as an overly simple one (underfitting). Then the complexity of the architecture of an LSTM model should be selected in a way to make a tradeoff between the over and underfitting phenomena. In conclusion, and to be consistent, the model with the highest R2 and lowest MSE on the testing dataset was chosen. Based on the prediction performance of the optimized LSTM model, the modeler can analyze the output of the signal, and conclusion regarding market efficiency can be deduced. Consistent with previous signals, and using the same dataset, the goal is to investigate the possibility of predicting future returns using the LSTM model. The data sample is again divided into two subsamples.

2.5. Assigning weights to signals: The analytic hierarchy process

Having established different signals and their measurement, what follows is the possibility to allocate different weights before aggregating the findings. At this stage, the regulator or authority has already collected various data from market participants (mainly commodity prices, and quantities whenever possible). The following step is to prioritize signal results by allocating the desired weight for each signal analysis.

An accurate approach to quantify and measure these weights is the AHP. Here, signals are evaluated pairwise and compared and sorted by the degree of relative importance. AHP is the most frequent method used in multicriteria decision-making (Saaty, 2004). Signals that are conclusive and justified are based on models with a minimum level of statistical errors and at the same time point in a clear direction. Hence, in the AHP matrix such a signal is assigned a higher weight relative to other signals. It can simply be categorized as more important in terms of information.

Using the AHP method, a pairwise comparison matrix is created, and it shows the relative importance of the different criteria (signals) concerning each other. The interpretation ranges from "less important" (0.5), "equally important" (1), "Highly

important" (2). The elements of the pairwise comparison matrix is denoted by r_{ij}, with $i = 1, \ldots, 4$ and $j = 1, \ldots, 4$ which designate the pair of signals that are compared. In other words, when $r_{ij} = x$, it means that signal i is x times more important/significant in concluding the efficiency of the underlying time series than signal j. Linear algebra and normalization operations are applied to the matrix previously defined to identify the weight of each criterion. Mathematically, this is done by calculating $r_{ij}^* = \frac{r_{ij}}{\sum_{k=1}^4 r_{kj}}$ and $w_i = \frac{\sum_{k=1}^4 r_{ik}^*}{4}$, where w_i denotes the weight that should be assigned to each of the four signals. These weights are then normalized and interpreted as a probability distribution, $w_i^* = \frac{w_i}{\sum_{k=1}^4 w_k}$. Thus, after computing the normalized weights, signals that are conclusive and justified tend to have stronger weights than those that are not. To verify whether the results are statistically significant or not, the consistency is calculated. The statistical test is based on linear algebra operations combining the matrix and weights computed above. In what follows, the main mathematical equations used to compute the consistency ratio (CR), which is a comparison between the consistency of the considered pairwise matrix and a random pairwise matrix, are defined.

The new pairwise matrix is defined by the transformation expressed in Eq. (15).

$$r_{ij}^{**} = r_{ij} * w_j, with\ i = 1, \ldots, 4\ and\ j = 1, \ldots, 4 \tag{15}$$

Then $\lambda_i = \frac{\sum_{k=1}^4 r_{ik}^{**}}{w_i}$ and $\lambda_{max} = \frac{\sum_{i=1}^4 \lambda_i}{4}$ are defined. λ_{max} denotes the highest Eigenvalue of the new pairwise matrix and should be compared with $\lambda_{max} = 4$ (dimension of the matrix) which is the largest Eigenvalue in a perfectly consistent context (Franek & Kresta, 2014). The Consistency index (CI) is computed and defined as follows: $CI = \frac{\lambda_{max} - 4}{4 - 1}$, which is a measure of consistency. Finally, the consistency ratio (CR) is estimated, and it measures the proportion of inconsistency in the matrix, i.e., how close the model results are to a complete random case; it is also defined as follows: $CR = CI/0.9$. If the value of CR is smaller or equal to 10%, the consistency is acceptable; otherwise, the pairwise comparison matrix (computed in step 1) should be revised.

2.6. Mutual information

To evaluate the individual contribution of each signal to our proposed measurement of entropy, the concept of MI is used. Before the reception of any information through signals, prior entropy $\mathcal{H}_{prior}$ is computed under the assumption of a uniform discrete distribution, see Eq. (16).

$$\mathbb{P}\,[Efficient] = \mathbb{P}\,[non - Efficient] = 0.5 \tag{16}$$

Then

$$\mathcal{H}_{prior} = -[\mathbb{P}\,[Efficient] * log_2 \mathbb{P}\,[Efficient] + \mathbb{P}\,[not - Efficient] * log_2 \mathbb{P}\,[not - Efficient]] = 1 \tag{17}$$

Receiving additional information through signals $k = 1, \ldots, 4$ impacts entropy by considering conditional probabilities. Hence, the expression for conditional entropy on signal k becomes:

$$\mathcal{H}_k = -[\mathbb{P}\,[Efficient|signal_k] * log_2 \mathbb{P}\,[Efficient|signal_k] + \mathbb{P}\,[not - Efficient|signal_k] * log_2 \mathbb{P}\,[not - Efficient|signal_k]] \tag{18}$$

The aggregated impact of all signals is the posterior entropy $\mathcal{H}_{posterior}$, utilizing the weights determined using the AHP method, is expressed in Eq. (19).

$$\mathcal{H}_{posterior} = \sum_{k=1}^{4} w_k^* * \mathcal{H}_k \qquad (19)$$

The impact of information from the received signals in terms of reduced uncertainty related to stability is measured by the difference in $\mathcal{H}_{prior}$ and $\mathcal{H}_{posterior}$ and is referred to by the previously defined MI. The higher the value of MI, the higher the value of the information received, relative to complete uncertainty; the new information received from the signals has thus contributed to uncertainty reduction.

3. Application to the UK wholesale natural gas market

The authors decided to use the UK wholesale natural gas market as a case study to test the four signals approach, AHP weight attribution, and MI concept. Initially, the British Gas Corporation, dominated the natural gas market in the UK, purchasing all indigenous production and securing external supplies. Furthermore, it was a monopolist in providing the commodity to UK end users. The formulation of upstream prices was complicated as each contract's terms were individually negotiated over several months. Downstream prices were calculated using the weighted average cost of natural gas plus a margin to cover the remaining costs (transport and distribution).

The Gas Act of 1986 removed the monopoly by allowing third-party access through its assets (pipelines). Simultaneously, the first natural gas regulator (OFGAS) was established. During the liberalization process that followed, the most important acts from Parliament emerged in 1995 and 1996 to fully liberalize the market. It allowed the establishment of a new licensing system composed of pipeline operators, wholesalers/shippers, retailers (also known as suppliers) and network codes (Helm & Jenkinson, 1997). A direct result is the entry of new market participants along the entire value chain. Thus, the market shares of British Gas fell from a complete monopoly to less than 50% for each customer type (small/large firms, power stations, others).

Additional regulatory and infrastructure developments included the merger of OFGAS into OFGEM, opening of the interconnector (IUK), unidirectional BBL pipelines between continental Europe and the UK, addition of LNG receiving terminals, establishment of a virtual hub [national balancing point (NBP)] to allow the system balancing and as a trading point, and establishment of a virtual hub (NBP) to allow system balancing and trading.

This directly impacted market liquidity and the fact that UK customers could choose their supplier regardless of their size. Since its establishment, the NBP has become a cornerstone for both the British over the counter trades and those on the exchange (ICE futures). It is now a dynamic and active trading market, with all value chain participants, including traders, having confidence in buying and selling natural gas on a standardized basis.

Natural gas accounts for slightly more than one-third of primary energy in the UK. In 2019, 39.6 billion cubic meters (bcm) of natural gas were produced, including all onshore and offshore active fields. Over the last decade, the rate of production growth has been steadily declining. This is due primarily to decreasing production rates in mature large fields and a slow rate of new discoveries. Natural gas consumption in the UK (78.8bcm) is a combination of domestic and imported sources. Imports primarily come from Norway followed by supplies from continental Europe (natural gas exchange between the island and the continent, such as Russia and the Netherlands,

or quantities redirected and originally sourced from Norway, and LNG) through the IUK and BBL pipelines, as well as direct LNG imports. The commodity is traded on the NBP in a fully liberalized fashion, that is regional and global prices affect overall prices, but - more importantly - market fundamentals have a higher impact.

The participants in such a trade are classified into several categories: banks, customers, local producers, traders, OTC brokers, and external producers. Additionally, other participants are transmission operators (who must trade to balance the network), investors such as insurance companies, private investors for speculation, and some local commodity traders. In summary, a value chain participant would want to either buy or sell natural gas to balance a portfolio, hedge against market risks, and/or for speculation purposes. The evolution of the maturity and efficiency of a certain trading hub is based on several elements, among which, the following can be highlighted:

The number of market participants (in particular active traders), the variety of traded products available that attract not only physical but also financial participants (looking for option-traded products), and the churn ratio, defined as the traded volumes divided by the overall size of the market (Heather, 2010).

Natural gas prices and traded volumes are reported at the broker level, wire services (Reuters, Bloomberg, etc.), trade press/ national press, regulators, and others because market transparency is an important feature of commodity trading (Heather, 2012). The results of the evaluation of the maturity and transparency of the NBP market are encouraging: As of 2018, one can conclude that the NBP market is mature, transparent, and liquid (Heather, 2019).

In what follows, a brief section summarizing and highlighting the most recent findings found in the literature regarding the state of the UK naturals gas market is presented to our readers. In the period between 2005 and 2018, the NBP hub was the dominant hub in terms of its influence on price changes and spillovers on its neighboring European gas hubs (Broadstock et al., 2020).

In a more recent period (2011-2020), the British NBP gas market still holds its leading place, and is considered to be well established, a sign of a liquid functioning natural gas market, that is subject to few price bubbles, relative to other gas markets (Austrian VTP gas hub for instance) that are located in the same geographical area and that uses the same pricing mechanism (Akcora & Kocaaslan, 2023). The main reason for such as difference is attributed to the hub development level of the latter market which lags the former (NBP).

Currently, it is observed that the position it used to hold has weakened (Papież et al., 2022). Instead of being the emitter of the shocks, it has become a receiver. The Dutch TTF and German NCG are now the main transmitters of prices and volatility shocks, albeit the fact that more recently, the dominance of all three gas hubs have been weakened, due to recent crisis that affected gas supply and demand. This shift is attributed to Brexit in the case of NBP, gas supply crisis in the eastern part of Europe, decrease dependence on Russian gas and the increase in the supply of LNG.

Although, the liquidity in NBP natural gas futures has witnessed a downward slope recently, due to the above-mentioned events, however, a reasonable number of future gas contracts are still being purchased, especially in the winter (for hedging purposes). In that regards, there is still strong evidence on the validity of the theory of gas storage in the UK (Martínez & Torró, 2023). This concept is important to any stakeholder participant in a market; regulators for instance, need to plan important value chain facilities installation, such as storage; traders on the other side, need to decide whether to import LNG, or increase their storage inventory. The latter theory dictates, that when spot supplies are tight, the volatility of future contracts nearing maturity,

as well as spot day-ahead gas prices increases, more so to that of the more distant futures contracts. The theory of gas storage in the UK is proven empirically, as no clear evidence has been found between future gas price volatility and commodity scarcity, especially in the case where future contracts are not nearing maturity.

Still in the context of gas futures, European gas market integration as a whole has been affected (Chen et al., 2022). Analyzing pre-and post-COVID gas prices, the empirical findings suggest that recent extreme events have led to a supply and demand imbalance in the market, and consequently European gas market integration has been negatively impacted.

The level of distress recently witnessed, not only affects the gas supply and demand imbalances and European gas market integration, but also the electricity market is affected, as both goods are substitutes and complements (Uribe et al., 2022). It is shown that the effect of natural gas prices on the higher quantiles of electricity prices is much larger, in Finland, Denmark, Germany, than the effect of gas prices in lower quantile of electricity prices, which implies that the gas price volatility is transmitted to electricity prices. Thus, further gas market integration is proposed to increase resilience in European electricity markets, which by itself, and as mentioned previously, a cause of concern, in times of distress.

As a summary from most recent literature, the NBP, is still considered as a gas hub benchmark in Europe (despite recent events, such as Brexit and COVID), contributing to the strong gas market integration, and that the theory of storage in UK is strong. The above serves as evidence of an efficient and stable market, implying a low level of uncertainty that could be felt by value chain participants. The authors recognize, the recent events (Brexit, gas supply shocks in the east of Europe, more reliance on LNG sport market, COVID).

The flowchart of the methodology is summarized in the below explained four steps, as well as Figure 1.

Working under the assumption that gas prices are the result of market processes, the authors believe that there is a wealth of information available that is currently untapped. The contribution is to investigate the information contained inherent in gas prices.

Step 1: Define signals relevant to the market being studied. The choice is based on three main factors: availability/ type of data at hand, literature review, and expert's opinion.

To analyze the performance of the four signals defined earlier, one needs to explore a family of models denoted by F_i that can be applied with the aim to extract the measure of each signal. Subsequently the selection of the most suited model is founded on mathematical and statistical (e.g., statistical tests) analysis, and this optimal model is defined by M_i. In the UK case study, the following models are used:

- The Shannon entropy signal enables capturing the level of uncertainty (randomness) of gas prices time series $\sim$ Model M_1
- The Record theory will be used to understand the behavior of tail risk, and the distribution of extreme prices will be evaluated $\sim$ Model M_2
- The predictability of daily gas prices will be evaluated using parametric ($\sim$ Model M_3) and non-parametric deep learning model $\sim$ Model M_4 based on the history of a gas price return time series.

Step one shall be concluded by measuring the following properties of each of the signal/ model results: first, if the model/signal results are conclusive, and second, whether the model choice is theoretically justified.

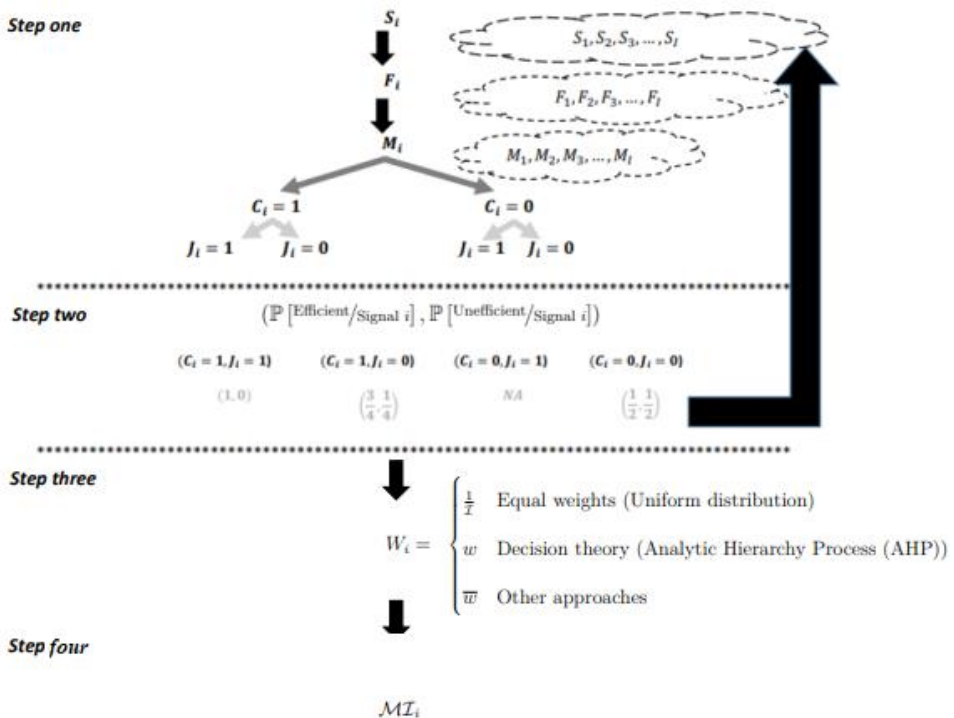Step 2: Measuring market efficiency specific to each signal

Figure 1. Methodology flowchart

Using the results obtained at the end of Step 1, one should be able to compute for each of the signals an efficiency value, defined previously as $\mathbb{P}\left[Efficient_{signal_i}\right]$

Before collecting the data and undergoing the methodology, the regulator is indifferent in its judgment about whether the market is efficient or not. The probability of either states is thus uniformly distributed, equivalent to say that, in the post-methodology scenario, results will lead to a nonconclusive and non-justified state, i.e. $(C_i = 0, J_i = 0)$, thus indifference in the state of the market, where $\mathbb{P}\left[Efficient_{signal_i}\right] = \mathbb{P}\left[non-Efficient_{signal_i}\right] = 0.5$, as defined earlier in equation 16. The remaining possible outcomes are defined previously in section 2.

Step 3: Attributing and allocating weights for each signal.

Moving forward in the analysis, a fair question will arise, and it is how to attribute and allocate weights for each signal. In this manuscript the AHP weight allocation is chosen and explained in the manuscript. Another possible weight attribution method is simply equal weights among signals/ models.

Step 4: Compute the measure of information.

Once weights are allocated, the user can now compute the weighted entropy for all previously defined signals, and use the information collected from Step 3. The weighted entropy shall be compared with a prior entropy computed under the assumption of a total lack of information (indifference state). In this final step, the user will compute a measure of information gained from the analysis, and that contributed to uncertainty reduction.

3.1. Data

The authors use the NBP daily time series for natural gas prices, except for signal four where monthly average prices are used. Figure 2 presents the monthly NBP wholesale natural gas prices recorded between January 2011 and December 2020 collected from PowerNext.
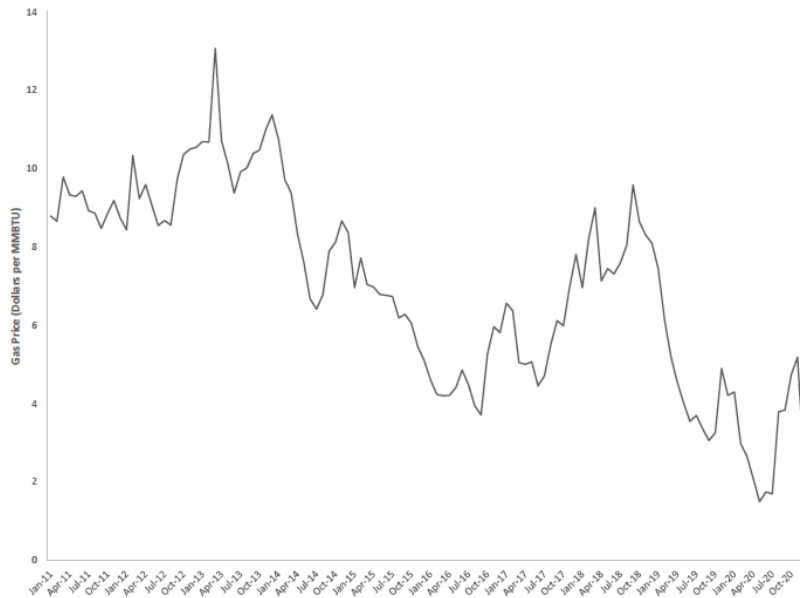

Figure 2. Natural gas prices times series in the UK

3.2. Retrieved signals

This section presents the findings for the individual signals' short- and long-run price stability, price predictability, and market competitiveness. The signals are then aggregated, the weights corresponding to the AHP mechanism identified, and the level of MI within the signals quantified.

3.2.1. Shannon entropy results (Signal 1)

The natural gas price returns are computed and discretized (returns for the training set are plotted in the Appendix, Figure A2). The descriptive statistics of the entropy results for the different sequences, while using the sliding windows analysis, are listed in Table 3. The aim of computing these values is to detect patterns over a wider temporal sequence of returns, thus improving the detection and formation of gas price trends.

Table 3. Entropy results

	One day (Seq$_i$ = 1)	Two days (Seq$_i$ = 2)	Three days (Seq$_i$ = 3)
Min; max	0.555;0.999	0.561; 0.998	0.559; 0.996
Mean	0.971	0.960	0.936
Median	0.993	0.975	0.952
Proportion of Similarity based on Kolmogorov Smirnov test	98%	97%	94%

The median of the entropies for all sequences is just below unity, implying that the price returns are more likely to be modeled as random than having a discernible pattern. Furthermore, even when the minimum value, which is the entropy result for one specific window (1 out of 2,561), is considered, the lowest reported entropy still complies with the preceding statement.

There should be equal probabilities for both outcomes in an efficient or weak market (1-day return sequence; being positive or negative). This will produce entropy close to 1. Conceptually, and consistent with the efficient market theory, the lower the number of outcomes (sequences of days), the higher the chance of having equal probabilities for each outcome. The higher the number of outcomes, the harder it is to conserve equal probabilities of occurrences per sequence. This is proven in Table 3 results.

If the average entropy across all windows is 0.971 (Seq_i = 1), or close to one, one can say that the signal is conclusive (and it announces a strong uncertain behavior of the returns). In other words, it is difficult to conceptualize information that can be used by gas supply chain participants to exploit the market.

The results of random seeding cross-validation, which is used for additional robustness checks and to loosen the time dependency factor, are also reported in same Table 3. It is observed that in 98% of the cross-validation cases, the Kolmogorov-Smirnov test accepts the similarity between the training and testing distributions. Then, it can be inferred that the entropy signal is coherent, and robust and that its use is justified. It is thus hardly possible to conceptually extract information that gas supply chain participants can use to exploit the market. Hence, the observed uncertainty in the underlying data is maximal, and the selected natural gas market can be considered a market with high efficiency.

3.2.2. Extreme results (Signal 2)

Based on record theory, the authors start with the cross-validation process described in Section 3.2.1. Monthly widows (30 days) of data are considered with a 1-day sliding step. The number of runs for cross-validation is considered 100. The comparison of the proportion test above accepts proportions equality between the training and testing decision variable distributions in 94% of the runs. Then, can say that the use of the record signal is justified and coherent. Finally, by considering all the data (all windows), the proportion of windows where the $i.i.d.$ behavior (classical model) of the underlying distribution is accepted is 92.7% which is close to 1 (C_2 = 1). From that, it is inferred that the record signal is conclusive and that prices are efficient in terms of extreme values. The model is assertive, and the probability of witnessing a record or shock in the long run is low, which is additional evidence of the efficiency of the considered natural gas market.

3.2.3. Predictability using parametric models (Signal 3)

A visual inspection of the price returns plot suggests a stationary process with some extreme values, mainly at the beginning of the observed period. Both the ADF and KPSS results confirm the stationarity of the time series for a lag order of up to 15 and 9 observations, respectively.

As a result, the ARMA model will be used to forecast future returns. The results of the partial and autocorrelation functions indicate three lags are significant in both cases. An ARMA (3,3) with a maximum absolute value of the AIC of 14081.31 extracts the most information. The results are presented in the Appendix (Figures A3 and A4).

The model's coefficient of determination is near 0, suggesting that the returns are nonpredictable. Although the optimal model is used, in terms of statistical theory and coherence, the results fail to make an inference for the price returns. Thus, this signal shows that short-term prices are nonpredictable based on their returns, which induces weak-market efficiency for predicting future prices conceptually. The return prediction results are found in the Appendix (Figure A6).

As a summary the parametric results are considered robust and justified due to the following reasons: The stationarity, ACF and PACF tests, and other model performance-related tests are conducted before making a final decision about the optimal model to be used. This means that the choice of model is statistically justified, and even when accounting for the best statistical model for such data, the results indicate that the prices are unpredictable. The results are thus robust.

3.2.4. Predictability using nonparametric models (Signal 4)

Similar to the parametric models, the coefficient of determination, computed on the validation set, of the LSTM model is near 0, suggesting that one cannot rely on this model to make a prediction. Then, it is another sign that the returns of the considered market are unpredictable, which indicates the market efficiency. The return prediction results for the nonparametric model are found in the Appendix (Figure A5).

For the nonparametric deep model, several measures were taken to decide about the optimal architecture and hyperparameters values to be used and avoid the over- and underfitting phenomena using multiple techniques. However, the authors prefer to label these signal results as non-justified as it is difficult to interpret deep learning black-box models.

However, both parametric and nonparametric models are also considered conclusive regarding the efficiency of the considered market. In fact, for both models, the measure of performance applied on the testing data does not have the ability to predict the considered time series, which is a sign of weak-efficient market.

3.3. Aggregating individual signals and value of MI

The findings of the signal analysis and their interpretation are presented in Table 4. The first row corresponds to the prior probability and is interpreted as follows: before receiving any information, the regulator is indifferent in its judgment about whether the market is efficient or not. The probability of either state is thus uniformly distributed. As a result, these assumptions lead to nonconclusive and non-justified state.

Table 4. Signal analysis results

Signal	$\mathbb{P}\left[Efficient\right]$	$\mathbb{P}\left[non - Efficient\right]$	Interpretation
P_{prior}	0.5	0.5	Non-conclusive, non-justified
Measure of Shannon entropy (S1)	1	0	Conclusive, justified
Measure of extremes (S2)	1	0	Conclusive, justified
Measure of predictability, using parametric models (S3)	1	0	Conclusive, justified
Measure of predictability, using nonparametric models (S4)	0.75	0.25	Conclusive, non-justified

The results indicate show that the matrix is reasonably consistent (CR = 0.017 < 0.1), and the process of decision-making using AHP generated weights is valid. Also, the weights for signals 1 and 2 should be highest when extracting the value of information contained in the data. The pairwise matrix results for the four signals are presented in Table 5.

Table 5. AHP weight matrix

Signals	S_1	S_2	S_3	S_4	Weights w_i^* in %
S_1	1	1	1	2	28.57
S_2	1	1	1	2	28.57
S_3	1	1	1	2	28.57
S_4	0.5	0.5	0.5	1	14.29

The quantitative tools, conceptually deal with numbers and statistics, and this is the aim of the developed instrument. The test hypothesis for robustness and the result conclusiveness are all built, as quantitative measures by design. When one lack the necessary means to prove either (conclusiveness or justification and robustness), then the signal cannot be trusted in the same way as another signal that is measured, conclusive, and robust enough, and thus, it can be inferred that the analysis can be generalized in the latter case.

It is worth mentioning that the quantitatively justified models, such as signals 1, 2 and 3, will have a high importance score relative to other signals that are either unjustified, or not directly justified. This is the case of the signal results, with a process of justification composed of several steps, some of which are qualitative and others quantitative. This is the case of predictability signals, whereby several tests (steps) are conducted to justify and validate the results. The MI analysis results are presented in Table 6.

Table 6. Comparison between AHP and equal weight MI results

Signals	AHP weights %
Prior Entropy (no information)	1
Posterior Entropy	0.1352
Mutual Information	86.4787%

The analysis of the four proposed signals, has contributed to the gain of information, thus reducing uncertainty when assessing market effectiveness by a value of 86.5%. This reduction is measured relative to the pre-signal analysis situation, where no price data were available and analyzed using the proposed tool. In addition, all signal results point to the weak-form efficiency of gas prices reported in the UK natural gas wholesale market, in the period of 2011- to the end of 2020.

4. Discussion and conclusion

MI conceptually describes how different signal results contribute to the reduction of uncertainty when assessing market efficiency. Our analysis began with the selection of models based on data-related assumptions, mathematical theory, and the type of signal that the authors find useful for the measurement. Such models can always be improved, and larger dataset could produce different results. To have a solid interpretation that leads to policy recommendations, the results must be as conclusive and justified as possible. Careful application of the proposed tool will feed the regulator with information about the state of the market. The analysis of the four proposed signals, has contributed to the gain of information, thus reducing uncertainty when assessing UK gas market effectiveness by a value of 86.5%. In fact, and with minimal access to data, the UK natural gas hub is proved to be in a weak form efficiency, a sign of a functional market where prices are transparent and reflect available information for all stakeholders.

To the best of the knowledge of the authors, it is the first time that a multitude of mathematical models are used, with the aim to refute (or not) a hypothesis regarding the state of a commodity market. The main advantage can cover several statistical and mathematical modeling approaches before inferring about the global state of the commodity market. Additionally, the analysis is done, based on a limited available dataset, and using classical models (low-level of computational complexity, i.e., very few model-specific input assumptions are needed, such as, but not limited to, hyper-parameters, assumptions about random variable distribution, etc.). Thus, one can make a consistent and accurate inference about the state of the commodity market. Moreover, the AHP technique is applied with a minimal subjective dimension. In fact, all the weights are computed, using a defined scientific approach, thus limiting subjectivity.

Future research should investigate broadening the concept and dealing with conclusive and non-justified signals. The limitation of this study is identifying an appropriate model to measure a market signal, leading to conclusive, but unjustifiable (in terms of closed mathematical and statistical arguments) results. The latter complicates the task of assigning a probability measure to the signal; however, dropping such a signal might lead to a false interpretation of the status of the market, making it difficult for the user to drop or retain the signal in the first place, and assign the probability in the second step.
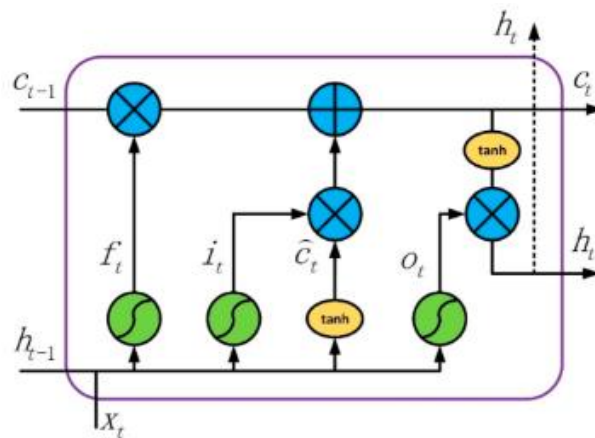
Appendix A
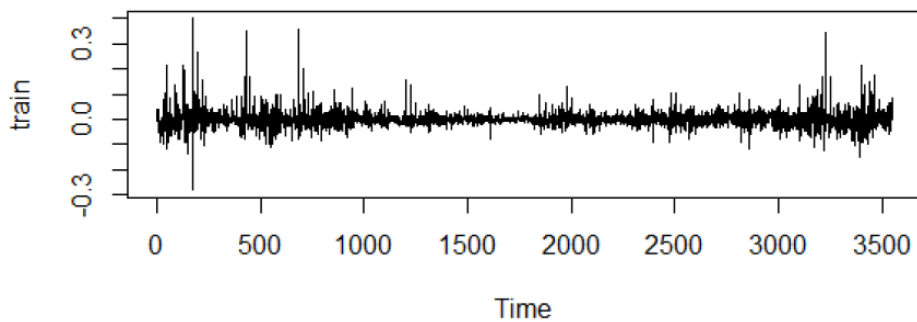

Figure A1. LSTM architecture
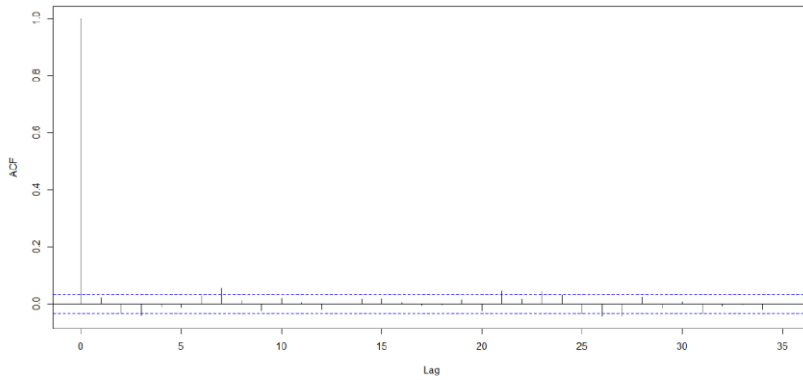

Figure A2. Gas returns plot

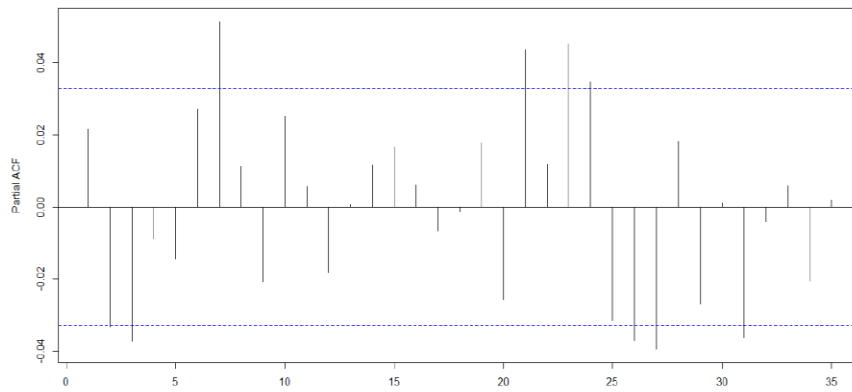**Figure A3.** Auto-correlation function results (ACF)

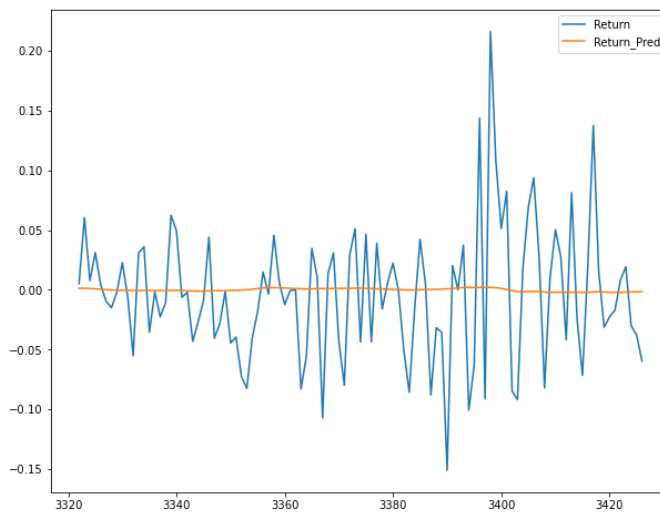**Figure A4.** Partial auto-correlation function results

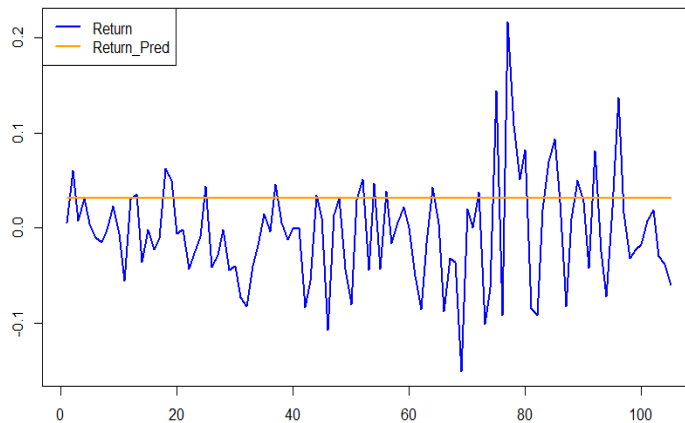**Figure A5.** Prediction (in orange) vs. real returns - testing data in LSTM

Figure A6. Prediction (in orange) vs. real returns for testing data in ARMA

References

Akcora, B., & Kocaaslan, O. K. (2023). Price bubbles in the European natural gas market between 2011 and 2020. Resources Policy, 80, 103-136.

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1998). Records. John Wiley & Sons. New York.

Bohl, M. T., Pütz, A., & Sulewski, C. (2021). Speculation and the informational efficiency of commodity futures markets. Journal of Commodity Markets, 23, 100-129.

Broadstock, D. C., Li, R., & Wang, L. (2020). Integration reforms in the European natural gas market: A rolling-window spillover analysis. Energy Economics, 92, 104-139.

Chen, Y., Wang, C., & Zhu, Z. (2022). Toward the integration of European gas futures market under COVID-19 shock: A quantile connectedness approach. Energy Economics, 114, 106-128.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. ArXiv Preprint ArXiv:1502.02127.

Franek, J., & Kresta, A. (2014). Judgment scales and consistency measure in AHP. Procedia Economics and Finance, 12, 164–173.

Hamie, H., Hoayek, A., & Auer, H. (2018). Modeling the price dynamics of three different gas markets-records theory. Energy Strategy Reviews, 21, 121–129.

Hamie, H., Hoayek, A., Kamel, M., & Auer, H. (2020). Northwestern European wholesale natural gas prices: Comparison of several parametric and non-parametric forecasting methods. International Journal of Global Energy Issues, 42(3–4), 259–284.

Heather, P. (2010). The Evolution and Functioning of the Traded Gas Market in Britain. OIES Paper NG44.

Heather, P. (2012). Continental European Gas Hubs: are they fit for purpose? OIES Paper NG63.

Heather, P. (2019). European traded gas hubs: a decade of change. OIES Insight OEI55.

Helm, D., & Jenkinson, T. (1997). The assessment: Introducing competition into regulated industries. Oxford Review of Economic Policy. doi:10.1093/oxrep/13.1.1.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Lee, J., & Strazicich, M. C. (2003). Minimum Lagrange multiplier unit root test with two structural breaks. Review of Economics and Statistics, 85(4), 1082–1089.

Lesne, A. (2014). Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics. Mathematical Structures in Computer Science, 24(3), e240311. doi:10.1017/S0960129512000783.

Lindström, E., & Regland, F. (2012). Modeling extreme dependence between European electricity markets. Energy Economics, 34(4), 899–904.

Martínez, B., & Torró, H. (2023). Theory of storage implications in the European natural gas market. Journal of Commodity Markets, 8(1), 100-121.

Meucci, A. (2011). 'The Prayer'Ten-Step Checklist for Advanced Risk and Portfolio Management. Ten-Step Checklist for Advanced Risk and Portfolio Management. doi: 10.2139/ssrn.1753788.

Mishra, V., & Smyth, R. (2016). Are natural gas spot and futures prices predictable? Economic Modelling, 54, 178–186.

Narayan, P. K., & Popp, S. (2010). A new unit root test with two structural breaks in level and slope at unknown time. Journal of Applied Statistics, 37(9), 1425–1438.

Nick, S. (2016). The informational efficiency of European natural gas hubs: Price formation and intertemporal arbitrage. The Energy Journal, 37(2). 25-42. doi: 10.5547/01956574.37.2.snic.

Pachón-Suescún, C. G., Pinzón-Arenas, J. O., & Jiménez-Moreno, R. (2020). Abnormal gait detection by means of LSTM. International Journal of Electrical & Computer Engineering (2088-8708), 10(2), 12-26.

Papież, M., Rubaszek Michałand Szafranek, K., & Śmiech, S. (2022). Are European natural gas markets connected? A time-varying spillovers analysis. Resources Policy, 79, 103-129.

Saaty, T. L. (2004). Fundamentals of the analytic network process—multiple networks with benefits, costs, opportunities and risks. Journal of Systems Science and Systems Engineering, 13(3), 348–379.

Uribe, J. M., Mosquera-López, S., & Arenas, O. J. (2022). Assessing the relationship between electricity and natural gas prices in European markets in times of distress. Energy Policy, 166, 113-128.