Energy Economics 54 (2016) 302–312

Contents lists available at ScienceDirect

Energy Economics

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

On the predictability of energy commodity markets by an entropy-based computational method

F. Benedetto [a,b,*], G. Giunta [b,1], L. Mastroeni [a,2]

[a] Department of Economics, "Roma Tre" University, Via Silvio D'Amico 77, 00145 Rome, Italy
[b] Department of Engineering, "Roma Tre" University, Via Vito Volterra 62, 00146 Rome, Italy

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form 21 October 2015
Accepted 6 December 2015
Available online 23 December 2015

JEL classification:
C530
C630
G130
Q470

Keywords:
Computational methods
Entropy analysis
Market efficiency
Energy commodity markets
Risk management science

ABSTRACT

This paper proposes a novel computational method for assessing the predictability of commodity market time series, by predicting the entropy of the series under investigation. Assessing the predictability of a time series is the first mandatory step in order to further apply low-risk and efficient price forecasting methods. According to conventional entropy-based analysis (where the entropy is always ex-post estimated), high entropy values characterize unpredictable series, while more stable series exhibits lesser entropy values. Here, we predict (i.e. ex-ante) the entropy regarding the future behavior of a series, based on the observation of historical data. Our prediction is performed according to the optimum least squares minimization algorithm, usually used in many computational aspects of management science. Preliminary results, applied to energy commodity futures, show the effectiveness of the proposed method for application to energy market time series.

1. Introduction

The understanding of the dynamic behavior of financial market time series (FMTS), and in particular, of energy-commodity market time series (EMTS), is of great and crucial interest for market operators. The energy marketplace is growing in complexity and in dynamism. Electricity, natural gas, heating oil, crude oil, and emissions are exchange-traded commodities subject to frequent price fluctuations on a short- and long-term basis. Extreme levels of price volatility increase energy price risk, generating, in turn, for various companies and institutions, other types of risk. The observation of historical data as well as the analysis of their volatility (and price fluctuations) can be useful indicators of the dynamic characteristics of the series, in order to effectively perform forecasting procedures. In fact, volatility and risk analysis are strictly related to the amplitude of the series fluctuations: high volatility results in large deviations from the mean, hence stating high unpredictability for that series. Hence, assessing the predictability of a time series is a crucial mandatory step needed for applying efficient and low-risk forecasting methods. According to Pincus and Kalman (2004), EMTS may deviate from constancy exhibiting two different behaviors: (*i*) the series is characterized by high standard deviation, and (*ii*) the series shows many irregularities. It is really important to discriminate between these two cases because they lead to different conclusions about the predictability of the series. In fact, the degree of variation from the mean is not usually related with unpredictability, while the amount of irregularities drastically affects the further forecasting process, resulting in unpredictable series. In fact, as shown by (Tsakalozos et al., 2011), a forecasting method can be very effective even in the presence of time series with high deviation from the mean, but only if the series are characterized by little irregularities. For example, if it would be possible to ensure to an investor that the series of future prices would be characterized by a precise sinusoidal pattern (although characterized by high deviation from the mean), then future prices can be planned according to a precise forecasting strategy. An example of a practical forecasting strategy exploited in the presence of sinusoidal patterns is shown in (Giunta and Benedetto, 2012; Benedetto and Giunta, 2013). Conversely, in the presence of high irregularities, the time series become unpredictable and any further applied forecasting method drastically degrades its performance. But,

* Corresponding author. Tel.: +39 06 57337079; fax: +39 06 57337026.
E-mail addresses: francesco.benedetto@uniroma3.it (F. Benedetto), gaetano.giunta@uniroma3.it (G. Giunta), loretta.mastroeni@uniroma3.it (L. Mastroeni).
[1] Tel.: +39 0657337016; fax: +39 06 57337026.
[2] Tel.: +39 06 57335693; fax: +39 06 57335771.

http://dx.doi.org/10.1016/j.eneco.2015.12.009
0140-9883/

if we select only the most predictable time series, then we can apply any forecasting method, enhancing its prediction performance. In this paper, we are not performing any prediction of the time series values, conversely we are proposing a method for assessing if the series can be predicted or not.

In particular, we can address, as usual, to standard deviation (SD) as a measure of deviation from the mean, while we will use entropy as the metric for evaluating the irregularities and, hence, the predictability of a series. Entropy is a concept borrowed first from classical mechanics, and then from information theory. In mechanics, entropy is used to quantify disorder and uncertainty of dynamical systems, or in other words, is an expression of the randomness of a system (Jaynes, 1965), while in information theory, entropy is considered as a measure of the information content of the series under investigation (Shannon, 1948). In particular, the mathematician C. Shannon in his pivotal work "*A Mathematical Theory of Communications*", (1948), related the concept of entropy also to that of uncertainty, stating that information measures the degree of uncertainty exercised by the source in selecting the message to transmit (that is, here, the uncertainty of the EMTS behavior in the future). In practice, information is the removal of uncertainty: high values of the Shannon entropy results in an unpredictable series, while lower values mean less uncertainty and hence a more predictable behavior of that series. It is well known (see Martina et al., 2011) that the price formation of EMTS (e.g. crude oil market series) results from the combination of several actors (e.g., producers, governments, speculators). They originate effects that are then perceived at different time scales, spanning from several days to many years. Hence, the understanding of price formation through the mechanisms underlying production, speculation, commercialization and consumption helps in clarifying the dynamics of energy markets. In addition, Behmiri and Pires Manso (2013) showed that EMTS often exhibit a systematic pattern (such as autocorrelation) and have an exceedingly complex structural model. Hence, these series can be very well analyzed by an entropy-based approach. In particular, Darbellay and Wuertz (2000) demonstrated the usefulness of entropy concepts to characterize EMTS by showing that the salient advantage of the entropy approach resides in its ability to account for nonlinear dependences in the autocorrelation structure of the underlying system dynamics.

This paper exploits a computational method based on entropy analysis, under the meaning borrowed from information theory, to study the predictability of EMTS. There are plenty of works exploiting the concept of entropy applied to many aspects of management science and in particular to the analysis of financial markets time series. For example, the validity of the entropy approach in analyzing financial time series is demonstrated by the work of Darbellay and Wuertz (2000). Then, in the work by Pincus and Kalman (2004), an empirical method for evaluating the entropy of a series is proposed, namely the *approximate entropy* (*ApEn*). In particular, the authors use the approximate entropy technique as a marker of market stability, with rapid increases possibly foreshadowing significant changes in a financial variable. Entropy has been also used to quantify efficiency in foreign exchange markets (Oh et al., 2007) and stock markets (Risso, 2008, 2009; Zunino et al., 2009; Gradojevic and Gencay, 2011). Recently, studies focusing on energy commodity markets have been carried out under the entropy-based approach. As an instance, an entropy analysis of crude oil price dynamics is revealed in Martina et al. (2011), while evidences from informational entropy analysis in evaluating the efficiency of crude oil markets were discussed in Ortiz-Cruz et al. (2012). Then, in Kristoufek and Vosvrda (2014) a market efficiency index based on the ApEn metric is discussed for application to several energy commodities. However, all the aforementioned works evaluate the entropy of historical data and, applying *ex-post* considerations, try to declare the predictability of the series, i.e. they implicitly assume that the series under investigation are characterized by a stationary behavior. This means that they suppose that the past statistical features of the analyzed series remain unaltered also in the future.

In this paper, we move further by proposing an algorithm to predict the entropy regarding the future behavior of FMTS, (and in particular of EMTS), based on the observation of historical data. We do not estimate the entropy of the analyzed series; rather, we predict the entropy of the next time interval of the series. We remove the assumption of stationary series, assuming that its statistical features change in time and influence the future behaviors of the series itself. Then, according to the conventional entropy analysis, see for example the works of Pincus and Kalman (2004); Martina et al. (2011), and Ortiz-Cruz et al. (2012), if we predict high entropy values we are facing with unpredictable series. Conversely, more stable market time series exhibit lesser-predicted entropy values. Our algorithm exploits the concept of entropy under an information theory viewpoint, recalling the *maximum entropy theory* to evaluate the entropy estimation. In addition, our prediction is performed according to optimum prediction methods usually used in computational methods applied in management science, such as the least squares minimization scheme.

The remainder of this work is organized as follows. Section 2 discusses the basic frameworks about energy market-based entropy analysis. The first half of the section is dedicated to the approximate entropy method, while in the second half the maximum entropy theory is depicted. Then, our proposed entropy estimator is shown in details in Section 3, with all the mathematical derivations. Section 4 first discusses the validation of the proposed computational method exploiting two controlled experiments, and then shows the results and discussions about the application of our method to EMTS. Finally, our conclusions are depicted in Section 5.

2. Energy market-based entropy analysis

2.1. Approximate entropy

Conventional entropy theories are usually related to infinite data series, corresponding to an infinitely accurate precision and resolution for entropy evaluation (Dorfman, 1999). However, practical data are finite time series data, sampled with a sampling rate T_s and characterized by limited resolution. The problem is that accurate estimation of the series entropy requires a big amount of data to be processed, and the results will be greatly influenced by the system noise. To overcome these limitations, Pincus (1991) introduced the approximate entropy (ApEn) method, to numerically quantify the entropy content of a finite time series, as a measure of the regularity of the series itself. The regularity of the series clearly reflects in the predictability of the series, (Pincus et al., 1991). ApEn is able to obtain the entropy estimation by modifying an exact regularity statistic, namely the maximum entropy (or *Kolmogorov–Sinai* entropy). ApEn was initially developed to analyze medical data, such as heart rate in Pincus et al. (1991), and to study physiological time-series in Pincus and Goldberger (1994). Later, ApEn spreads its applications also in finance, as shown in Pincus and Kalman (2004). The ApEn computations are conceptually simple and are based on the likelihood that templates in the time series which are similar remain similar on next incremental comparisons, (Martina et al., 2011). In other words, the presence of repetitive patterns of fluctuation in a time series renders it more predictable than a time series in which such patterns are absent. Hence, a time series containing many repetitive patterns has a relatively small ApEn, while time series with large ApEn should have high irregular fluctuation (see Fig. 1).

According to Pincus and Kalman (2004), ApEn needs two input parameters to be specified in order to evaluate the approximate entropy of a given time series: a block or run length m, and a tolerance window r. Then, the ApEn procedure first measures the logarithmic frequency that runs of patterns that are close (within the tolerance window r) for m contiguous observations remain close (within the same tolerance r) on the next incremental comparison. A detailed and more formalized definition of the ApEn method is given in Pincus (1991). Widely used