Journal of Quantitative Economics (2021) 19:1–21
https://doi.org/10.1007/s40953-020-00214-y

ORIGINAL ARTICLE

Information Theoretic Ranking of Extreme Value Returns

Parthajit Kayal[1] · Sumanjay Dutta[2] · Vipul Khandelwal[3] · Rakesh Nigam[1]

Published online: 12 July 2020

Abstract

This study highlights the importance of the information contained extreme value ratios (or returns) in the volatility estimation of financial assets. Most popular extreme value estimators like Parkinson (Journal of Business, 61–65, 1980), Garman Klass (Journal of business, 67–78, 1980), Rogers Satchell (The Annals of Applied Probability, 504–512, 1991) and Yang Zhang (The Journal of Business, 73 (3), 477–492, 2000) use a subset of all available extreme value ratios but not the full set. We examine if there are other extreme value ratios which contain more information than the most widely used ratios. This study shows empirically how much information is contained in various extreme value ratios of financial assets, using both real and simulated data. Using information theory, we find out their variability in relation to a uniform distribution in each quarter. We then rank them using the Kullback–Leibler metric (in accordance with a scoring methodology we developed in this study) to ascertain which set of ratios are more variable than others and thus may provide better estimation in computing volatility. We also calculate the rank of the matrix to identify the set of linearly independent ratios, for ascertaining the number of ratios that would be enough to generate a class of volatility estimators. The empirical results demonstrate that the need for incorporating other ratios in volatility estimation. We also observe that each dataset has other more informative ratios which are uniquely attributed to that dataset.

Keywords Extreme value estimators · Information theory · Volatility

JEL Classification G10 · G140 · G150 · G170

✉ Parthajit Kayal
parthajit@mse.ac.in

[1] Madras School of Economics (MSE), Behind Government Data Centre, Gandhi Mandapam Road, Kottur, Chennai 600025, India

[2] Indian Institute of Science, CV Raman Road, Bengaluru, Karnataka 560012, India

[3] Tata Consultancy Services (TCS), Sholinganallur, OMR, Chennai, 600119, India

Introduction

Finance involves the study of time and chance. Chance is defined by the *states of nature* whose occurrence is subject to probability. If these states of nature are quantifiable, it essentially becomes a study of risk. The Risk and portfolio selection, which still dominate the discourse in modern finance theory was given a mathematical framework by Markowitz (1952), which essentially said that the first and the second order moments were enough to make a portfolio choice that was optimal. The economic justification for the same came a few years later when Samuelson (1970) proposed an approximation theorem which held the higher order moments as redundant in making portfolio choices. Subsequent developments in the literature have revolved around the study of volatility though the measurement of volatility still remains an issue of much disagreement among academicians and practitioners alike. The simplest measure of volatility in finance is defined as the variance of returns. It is given by:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 \quad (1)$$

Volatility, despite being a very important summary statistic, suffers from a major drawback, as it does not convey anything about the direction of change. Apart from this, asset returns are anything but well behaved and issues arrives when we deal with a multi period setting (which distorts the Variance- Covariance structure of asset returns) and extreme events. The measurement of volatility is broadly divided into two classes: the Conditional and the Unconditional methods. The Conditional methods are dependent on the information set and the underlying model which captures the behavior of volatility.

$$Var(x) = E[x - \mu|\Omega]^2 \quad (2)$$

where Ω is the information set. It is also important to note that the different measures of volatility have been developed in response to the deviance from what is observed empirically. One such phenomenon is volatility clustering which Mandelbrot (1963) describes as large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. Put in other words, this essentially means that while returns themselves might be independent of their lagged values, the absolute returns $|r_t|$ or squared returns might show some correlations which might also be decaying in nature. That is, $corr(|r_t|,|r_{t+\tau}|) > 0$, for τ ranging from very short to longer time intervals. This empirical property was observed in the 90′s by Granger and Ding (1996) among others. Ding and Granger (1996) point to long-range time series dependence in volatility.

There has been an increasing consensus about treating financial volatility as a latent factor, something which cannot be observed directly. One way of doing this is to derive the volatility of the asset from its option's value using numerical methods. Ross (1981) suggests that option prices have expectations of future value of the underlying asset embedded in them. Transforming this to an eigenvalue problem can help in recovering the probabilities of the expected state prices and hence the expected volatility of the asset. Moreover, assuming that markets are efficient, the expected volatility of the asset should reflect itself in the actual volatility of the asset. All these methods, in principle, convey the basic idea that volatility in financial markets can only be indirectly estimated using its imprint on an underlying market process. With increasing complexity and the inclusion of market microstructure noise effects, the results lack clarity. By lack of clarity, we mean that as markets processes become increasingly integrated, traditional methods of computing volatility become less efficient in incorporating information from the market. In incorporating the dynamic nature of the price discovery process and hence the time varying volatility in financial markets, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family of models (Engle 1982; Bollerslev 1986), and the stochastic volatility (SV) models (Taylor 1986) are two commonly used alternatives that estimate and model time-varying conditional volatility of asset returns. These models however, come with their own flaws. Diebold (2002), and a few other author have come up with the objection that since the GARCH and SV models are formulated on the basis of only the closing prices, of the period under consideration, they fail to use the information that is embedded in the path of the pricing process, therefore rendering them inaccurate and inefficient. Illustrating this with an example, let us consider a turbulent day of trading in the financial markets. Information has flown into the market and since the spread of this information might not be as instantaneous as the Efficient Market Hypothesis would suggest, coupled with the noise in this process, it leads to high levels of intra-day volatility. In such a scenario, with frequent drops and recoveries of the markets, the traditional Close-to-Close volatility measure indicates a low level of volatility while the daily price range is more accurate in indicating that the volatility is high.

Among the class of unconditional volatility measures, the traditional volatility estimates are calculated using Close-to-Close or Open-to-Close returns. This kind of estimation does not incorporate the information contained in that particular day's price trajectory. Parkinson (1980) accounted for "High and Low prices" in its estimation and thus extended the regular volatility calculation. Beckers (1983) demonstrated empirically that daily price ranges contain new information when compared to the traditional measures. This was a significant improvement over the previously used estimators because of the superior quantum of information that it incorporated. The stochastic process usually assumed for a stock price is geometric Brownian motion which is given by the equation:

$$ds = \mu S dt + \sigma S dz \tag{3}$$

Under such assumptions, Ozturk et al. (2016) show that the extreme value estimators of volatility are substantially more efficient than the traditional methods. A major drawback of the Parkinson's estimator was that it did not take opening and closing prices into account. This issue was resolved by Garman and Klass (1980) as they considered those prices along with highs and lows into their estimation. Despite their superior information content, the two estimators were unable to handle non-zero drift. Rogers and Satchell (1991) developed an estimator which allowed for non-zero drift. However, a major disadvantage of these three estimators is the fact that they are not robust for opening jumps in prices. Yang-Zhang extended the Garman-Klass method called as GKYZ that allows for opening jumps hence it is a fair estimate, but assumes zero drift. Yang and Zhang (2000) derived the volatility estimator which has minimum estimation error. It is a weighted average of Rogers-Satchell, the Close-to-Open volatility and the Open-to-Close volatility and is a drift independent estimator (Yang and Zhang 2000). A common feature among all these estimators is the assumption that the security price follows a geometric Brownian motion, but with minor differences. Calculating volatility estimation makes use of extreme values. The input variables are (C_{t-1}, O_t, H_t, L_t, C_t). Here C_{t-1} denotes the previous day's closing price. O_t, H_t, L_t and C_t denote the Opening, High, Low and Closing prices of the current day.

In the present study, we constructed 10 ratios out of five input variables or the prices of today's trading day (making a unique pair out of each of these variables). These ratios are listed below:

Out of these ten ratios, we considered pairs of any two ratios at a time and arrived at 45 cross product ratios. Existing literature has used ratios R_2, R_3, R_4, R_8, etc. (R_1could be multiplied to R_2 and so on. Doing so, we arrive at ratios such as R_1R_2, R_2R_3, etc.) Added to the squared product of the ratios (such as $R_1R_{1,}$ $R_2R_{2,}$ etc.), we get 55 new ratios. The different unconditional Volatility estimators have used combinations of the different subsets of these 55 ratios. It is important to note that the current literature has not yet considered all possible combinations of these ratios and hence their efficacy in terms of conveying information remains untested. Hence, the hypothesis we propose is that the other ratios may contain more information and therefore more useful in estimating volatility. Thus, as a natural consequence of the current literature, it becomes important to investigate about the information content of the other ratios, that have not yet been assessed. Moreover, empirical evidence suggests that observed extreme values can provide a significant amount of information about the volatility within the trading period.

Review of Literature

The literature on extreme value volatility estimators begins with Parkinson (1980) who incorporated extreme values of price movements. His estimate of day to day volatility yielded more details of the intraday movements. This made his estimate a superior one in terms of efficiency (that is, the information content was higher) efficient than the widely used Close-to-Close estimator. The Parkinson estimator was based on the high and low values attained by the price process during the course of the day.

$$\sigma_{PK}^2 = \frac{1}{4ln2}\left(Ht - L_t\right)^2 \tag{4}$$

where Ht and Lt are log transformed prices (we shall use the same convention for other estimators as well.) It is evident from the above formula that as the spread between the high and low values of the price increases indicates higher levels of volatility. Acknowledging the many sided advantages of the Close- to-Close estimator such as its simplicity of usage (which is as important to a practitioner as to an academic) and its freedom from obvious sources of error and bias on the part of market activity, Garman and Klass (1980) pointed out that the crippling disadvantage of this estimator is its inability to incorporate Open, High, Low and Closing prices which could have contributed to higher efficiency in estimating volatility. Building on the Parkinsons estimator, they developed a volatility estimator that incorporated the high and low prices, and was an optimal combination of the Close-to-Close and the Parkinson volatility estimators. To illustrate this point, the Garman-Klass estimator is given below:

$$\sigma^2_{GK} = \frac{N}{n} \sum \frac{1}{2}\left(H_t - L_t\right)^2 - (2ln2 - 1)\left(C_t - O_t\right)^2 \quad (5)$$

A major advantage that this improved efficiency yields is that a significantly lesser number of observations (seven times fewer) are required to attain the same state of statistical precision as the Close-to-Close estimator. An important component of any price process that supposedly follows the Geometric Brownian Motion is the drift of the process. A drift in the motion is similar in function to what a trend does in a time series model. Not accounting for the drift, one may end up overestimating the volatility of security prices. To capture this trend like behaviour of the price process, Rogers and Satchell (1991) devised an estimator that allows for an arbitrary drift, given by:

$$\sigma^2_{RS} = \frac{N}{n} \sum \left(H_t - C_t\right)\left(H_t - O_t\right) + (L_t - C_t)(L_t - O_t) \quad (6)$$

A major drawback of this estimator is that it provides less precision than the other estimators and is only limited in its usage when the process involves a drift component. The Meilijson estimator (Meilijson 2008) is a slightly improved version of the Garman-Klass volatility estimator. The Meilijson estimator was observed to have a relatively higher efficiency compared to the Garman-Klass estimator. All of the studied estimators except for the Rogers-Satchell are derived under the assumption of zero drift.

Attempts at improving the Garman Klass estimators have been made, though these new class of information augmented estimators face the same limitations as their predecessors in the sense that they too, do not incorporate drift in their analysis. In an attempt to improve on this shortcoming of the estimators suggested by the previous literature, Yang and Zhang (2000) created a volatility measure that handles both opening jumps and drift. It is a combination of the Overnight volatility (defined as the Close-to-Open volatility) estimator and a linear combination of the Rogers-Satchell and the Open-to-Close volatility es- timators. The assumption of continuous prices does mean the measure tends to slightly underestimate the volatility. The specification of the Yang-Zhang estimator is given below:

$$\sigma^2_{YZ} = N(\sigma^2_1 + k\sigma^2_2 + (1 - k)\sigma^2_{RS} \quad (7)$$

where k is a constant. σ^2_1 and σ^2_2 measure the Overnight and the Open-to-Close volatilities. Empirical results suggest that variances measured by extreme value estimators efficiently approximate the true daily variance. They are proven to be highly efficient (Shu and Zhang, 2006). Todorova and Husmann (2012) demonstrated the same thing as above but they showed that estimators based on daily ranges are severely negatively biased due to discrete trading. The efficiency tests that are used to gauge the importance of and comparing the estimators are based on a simple intuition. The closer an estimator is to the historical volatility measure, the lesser is its information content and hence, it becomes less efficient. This idea is captured by the fact that the Relative Efficiency of an estimator is given by:

$$RE = \frac{\sigma^2_{EV}}{\sigma^2_{HV}} \tag{8}$$

In the above formula, the numerator and denominator denote the variances of any two given unbiased estimators, denoted by EV and HV. An important issue that needs to be addressed is about the sensitivity of these efficiency tests to the probability distribution of the realization of a process. Physics and Information theory (developed by Claude Shannon in the 1940s), propose to analyze any message in terms of any kind of uncertainty or any disorder, using the concept of entropy which is defined on the basis of the value of the logarithm of the number of likely equivalent messages. This concept has been used to address some of the fundamental questions in financial theory, especially in portfolio models, thus providing researchers with a kind of functional parallelism that can be used to study the working of financial markets in terms of relatively well defined natural processes. In this context, entropy has been described as an information measure by some researchers, in the sense that this measure gives an idea of what the true value of the information may be. Putting it in simpler terms, the entropy of a random variable can be defined as the amount of information required to describe it. Delving deeper into the notion of what entropy really means, one may ask that given the information that one may have about a certain random variable, is it possible to use it in some way to describe another random variable. To address this problem, one needs to be familiar with the idea of what is known as *Relative Entropy* in the Information Theory literature.

Relative entropy or the Kullback Leibler (KL) Distance (By distance, we do not mean the metric distance.) is defined as the distance between two probability distributions. In statistics, one encounters this idea in the form of the expected logarithm of the likelihood ratio. Another way of understanding this idea is to interpret the relative entropy D(P||Q) as the inefficiency in estimating the true distribution p as the erroneous distribution q. To further illustrate this point, if a code of description length H(p) is used to describe a certain message, one could use a code of description length H(q) instead, adding the term D(P||Q) to account for the distance between the two. These ideas about entropy and information have found use in the literature on financial theory spanning fields such as market microstructure theory, asset pricing and risk management. Philippatos and Wilson (1972) were the first two researchers who applied the concept of entropy for portfolio selection. Using the mean entropy approach, possible efficient portfolios were constructed from a randomly selected sample of monthly closing prices of 50 securities using data over a period of 14 years. Ong (2015) employed information theory to examine the dynamic relationships between stock returns, volatility and trading volumes for SP 500 stocks. Literature in this context has shown evidence that the information contained in the level of the extreme returns (which is lost while using the observed ranges for inferences) can also contribute to more efficient estimation of volatility (Horst et al., 2012).

The present study makes use of Shannon entropy and KL distance to measure the distance of the probability distribution of each ratio from the uniform distribution in each of the quarters. Unlike most information theoretic problems, this study does not address the issue of message extraction. The primary focus has been on developing a framework that is not parameter sensitive and could work well in terms of providing a ranking based on the information content. One thing that needs to be noticed is that most of the volatility estimators that are currently in usage (and have been analyzed in this study as well) address a statistical issue and do not capture any microstructure based signals that the price process may have incorporated. Thus, this discrepancy compels us to develop only a ranking method and the scope for unraveling the implicit market process beneath the explicit volatility measure is limited.

Data and Methodology

This study uses both the real data as well as the simulation generated data for seventy two quarters (on a daily basis) and two commodities viz. Crude Oil and Gold Futures. Two things need to be understood here. Firstly, the choice of assets is arbitrary here. That is, any other commodity could have been chosen and the applicability of the KL distance is not sensitive to the choice of the asset. Secondly, the period under consideration is not fixed either. Any other time span would have worked well. The only reason to consider a sufficiently long time series data is to observe patterns that are stable and not subject to fluctuations that are often associated with the short run. The real data was extracted from the Investing.com website database which is a global financial portal. The simulation process used the Geometric Brownian Motion for simulating the price trajectories for these commodities (assuming that prices follow a random walk like movement in continuous time, thus allowing for market efficiency). Eventually, the data series was further divided into seventy two quarters. The commodities have been listed with their characteristics below.

Commodity	Group	Unit
Gold Futures-Jun 18(GCM8)	Metal	1 Troy Ounce
Crude Oil WTI Futures (CLM 8)	Energy	1 Barrel

Five key variables need to be defined in this context. Open is the price of the stock at the beginning of the trading day (it need not be the closing price of the previous trading day), High is the highest price of the stock on that trading day, Low the lowest price of the stock on that trading day, and Close the price of the stock at closing time. The ratios under assessment in our study have all been derived from these five input variables namely C_{t-1}, O_t, H_t, L_t, and C_t from which the estimators have been derived.

The present study examines various extreme value ratios developed in the volatility estimation literature (as well as those which have not yet been analyzed) from an information theoretic point of view. The entropy of a random variable measures uncertainty in probability theory. One way of appreciating this idea, is to define something known as the *occurrence probability* of an event. A smooth pattern of price movement is less likely to yield us some news compared to a pattern where some shocks are visible. Since shocks are less likely to occur than less volatile price changes, they tend to carry more information and hence higher entropy. In a more formal setting in the field of information, entropy represents the loss of information of a physical system observed by an outsider, but within the system, it represents countable information. The purpose of entropy metric is to measure the amount of information. Its applications in finance can be regarded as extensions of both information and probability entropies. In our context, it is also very important to introduce another useful principle, which is known as the Minimum Cross-Entropy Principle (MCEP). This principle was developed by Kullback and Leibler (1951), and it has been one of the most important entropy optimization principles. Consider a data set that we have for which the actual statistical distribution is given by $P = p(x)$. We propose a distribution $Q = q(x)$ for modeling the data set (a traditional example would be to use a least-squares line fit for Q). We would prefer having a measure which can tell us something about how well our model matches the actual distribution.

Information theory defines information in terms of encrypted message in a given signal or message (which is expressed in terms of codes) and introduces Shannon entropy, i.e., entropy in the discrete case. Shannon entropy could both be defined as the "expected value of the information of the distribution" and the number of bits one would need to reliably encode a message. The Shannon

entropy of a probability measure on a finite set X is given by:

$$H(x) = -\sum P(x)log_2P(X) \tag{9}$$

where, $\sum$P(x) = 1.

When dealing with continuous probability distributions, a density function is evaluated at all values of the argument. Given a continuous probability distribution with a density function f(x), we can define its entropy as:

$$H(x) = -\int f(x)lnf(x)dx \tag{10}$$

Thus, the entropy of a probability distribution is just the expected value of the information of the distribution.

As discussed above, we essentially want to device a criterion that measures the extent to which our fitted model is close to the real world. In this context, we defined what is known as the *Kullback Cross Entropy*. It measures the distance between two probability distributions, P and Q. If we have no other information other than that each $p_i \geq 0$, and the sum of the probabilities is unity, we have to assume the uniform distribution due to Laplace's principle of insufficient reasons. It is a special case of the principle of maximum uncertainty according to which the most uncertain distribution is the uniform distribution. Therefore in this study, we have taken the second distribution as the uniform distribution, which is the most random or uncertain distribution. Kullback and Leibler proposed the Kullback cross-entropy which is defined as:

$$H(p||q) = -\sum p_i ln\left(\frac{p_i}{q_i}\right) \tag{11}$$

Relative entropy (or KL divergence) proves to be the key to information theory in the continuous case, as the notion of comparing entropy across probability distributions retains value.

Kullback's cross-entropy can be considered as an "entropy distance" between the two distributions p(x) and q(x). It is not a true metric distance, but it satisfies $S(p, p) = 0$ and $S(p, q) > 0$, whenever p is not equal to q. (Here, S(p,q) refers to the entropy distance between any two given distributions p and q. Similarly for S(p,p). It follows intuitively that S(p,p) = 0.) Kullback's Principle of Minimum Cross- Entropy (MCEP) states that out of all probability distributions satisfying given constraints, we should choose the one that is closest to the least prejudiced posterior density p(x). In other words, the objective is to minimize the following expression:

$$H(p||q) = -\sum p_i ln\left(\frac{p_i}{q_i}\right) \tag{12}$$

When Q = U we have the following result:

$$H(p||q) = lnN - S(P) \tag{13}$$

Here, S(P) denotes the entropy distance. Coming back to the idea behind using the Uniform Distribution, one needs to revisit the notion of entropy. Maximum entropy principle emerged in statistical mechanics. If nothing is known about a distribution except that it belongs to a certain class, then distribution with the largest entropy should be chosen as the default. It is due to two reasons:

- Maximizing entropy minimizes the amount of prior information built in to the system.
- Many physical systems tend to move towards maximal entropy configurations over time.

Uniform distribution entropy calculation: Suppose we have $P\ (X) = \frac{1}{N}$ where X takes the values $x_1, x_2, \ldots, x_N$.

$$H(x) = -\sum P\left(X = x_N\right) log_2 P\left(X = x_N\right) = log_2 N \tag{14}$$

We should note that this is actually the maximum value for the entropy. This can be demonstrated using Gibbs' inequality, or just by finding the maximum of the function $f(x) = xlnx$. Uniform distribution has the highest entropy and thus the maximal uncertainty and low information. The probability density function on x_1, $x_2, \ldots, x_N$ with maximum entropy turns out to be the one that corresponds to the least amount of knowledge of $x_1, x_2, \ldots, x_N$ which, in other words, is the Uniform distribution. A well-known result is that the Maximum Entropy method is the special case of MCEP in which the target distribution is uniform and in our study, we utilize this result.

A well-known result is that the Maximum Entropy method is the special case of MCEP in which the target distribution is uniform and in our study, we utilize this result.

In our study, we use the KL metric to know the extent of divergence from uniform distribution for each ratio. Then, we ranked the ratios on the basis of their distance from the uniform distribution. As known from the above discussed principles, a larger distance from the uniform distribution indicates a higher information content. And hence, a higher rank. This process is repeated for each quarter and we get a frequency distribution like ranking system (Since a ratio might attain different ranks and a single rank multiple times as well.) After developing a rank matrix for all the ratios, we calculate the rank of the matrix to know how many ratios are linearly independent as it tell us that how many ratios could be used to describe the entire family of ratios (for the sake of parsimony). Another way of addressing this issue would have been to use Principal Components Analysis (PCA) but since we are not interested in these linear combinations, finding the rank of the matrix suffices for our needs.

An important issue that needs to be addressed is that while we might rank a ratio higher on the basis of it having attained a rank of 1 on a few occasions, it might actually have performed badly in the other quarters. If that is the case, one needs to take into account the distributions of the rankings. In other words, rankings need to be weighted by their occurrences for each ratio. In the subsequent section, we propose a way to do so.

Empirical Evidences

Real Data

Our analysis focuses on two kinds of data. To find out the most informative ratios, we initially worked on a real data set. We first analyzed the crude oil futures data. For the simple ratios (that is, the 10 fundamental ratios) and cross and squared ratios, we have the following results (Table 1).

One way of ranking the ratios could be on the basis of the best ranks attained by them. This is done by counting the number of times any ratio attained a given rank from 1 to 10, during the 72 quarters in consideration. Table 2 provides an illustration for this methodology. The above table illustrates one way of doing that. But while Table 3 talks about the first two *best* ranks attained by a given simple ratio, it

Table 1 Defining the Ten Fundamental Ratios

Ratio	Definition
R_1	$ln(C_{t-1}/O_t)$
R_2	$ln(H_t/O_t)$
R_3	$ln(L_t/O_t)$
R_4	$ln(C_t/O_t)$
R_5	$ln(C_t/H_t)$
R_6	$ln(C_t/L_t)$
R_7	$ln(C_t/C_{t-1})$
R_8	$ln(H_t/L_t)$
R_9	$ln(H_t/C_{t-1})$
R_{10}	$ln(L_t/C_{t-1})$

Table 2 Ranking of Crude oil Ratios according to the KL Criterion

Ratios	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10
R_1	72	0	0	0	0	0	0	0	0	0
R_2	0	0	1	7	10	12	9	17	12	4
R_3	0	0	1	11	24	13	9	8	5	1
R_4	0	0	0	7	7	8	7	15	15	13
R_5	0	0	1	2	0	2	6	5	12	44
R_6	0	0	2	15	16	17	10	6	5	1
R_7	0	0	2	16	5	9	20	11	6	3
R_8	0	0	2	10	7	10	11	9	17	6
R_9	0	49	23	0	0	0	0	0	0	0
R_{10}	0	23	40	5	3	1	0	0	0	0

Table 3 Crude oil simple ratios summarized

Ratios	Highest Rank	Second Highest Rank
R_1	Rank 1	None
R_2	Rank 8	Rank 6/Rank 9
R_3	Rank 5	Rank 6
R_4	Rank 8 /Rank 9	Rank 10
R_5	Rank 10	Rank 9
R_6	Rank 6	Rank 5
R_7	Rank 7	Rank 4
R_8	Rank 9	Rank 7
R_9	Rank 2	Rank 3
R_{10}	Rank 3	Rank 2

Table 4 Ranking of Gold Futures Ratios according to the KL Criterion

Ratios	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10
R_1	17	9	6	2	3	4	6	14	6	5
R_2	35	6	6	6	4	0	1	3	9	2
R_3	0	5	10	14	17	15	5	4	2	0
R_4	4	8	19	8	4	5	6	9	9	0
R_5	12	34	13	8	4	5	6	9	9	0
R_6	0	1	3	11	20	21	6	8	2	0
R_7	0	0	0	1	5	8	11	14	18	15
R_8	2	3	3	10	6	10	2	1	4	31
R_9	0	1	2	7	6	7	7	7	17	18
R_{10}	2	5	10	5	4	1	27	12	5	1

Table 5 Gold Futures simple ratios summarized

Ratios	Highest Rank	Second Highest Rank
R_1	Rank 1	Rank 8
R_2	Rank 1	Rank 9
R_3	Rank 5	Rank 6
R_4	Rank 3	Rank 8/Rank 9
R_5	Rank 2	Rank 3
R_6	Rank 6	Rank 5
R_7	Rank 9	Rank 10
R_8	Rank 10	Rank 4/Rank 6
R_9	Rank 10	Rank 9
R_{10}	Rank 7	Rank 8

does not take into account the ranking distribution of the individual ratios across 72 quarters. The problem remains the same with the derived ratios. Thus, for analyzing the square and cross product ratios, we developed a scoring methodology. We shall illustrate this with an example. Suppose, that a ratio (named X) attained the first rank for 60 quarters, the second rank for 10 quarters and the third rank for the last two quarters. Now, we need to give weights to all of these ranks. So, for the sake of simplicity, let the weights be linearly decreasing in nature. (We must be mindful of the fact that there are 55 ratios, and hence 55 rankings). Now, these rankings could be weighted by their frequencies of occurrence. Also, higher ranks need to be rewarded. This means that Rank 1 could be assigned a score of 55, Rank 2 with 54, and so on. So, X should be scored as[1] (Tables 4, 5, 6, 7, 8, 9, 10, 11):

[1] We could not include the ranking scores of all the ratios due to brevity. However it can be made available on request.

Table 6 Comparing ratios of both data sets

Common Simple ratios	Common squared and cross ratios
R_1	R_2R_2
R_3	R_2R_9
	R_4R_8
	R_7R_8
	R_8R_8

Table 7 Ranking of Crude oil Ratios according to the KL Criterion

Ratios	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10
R_1	71	0	1	0	0	0	0	0	0	0
R_2	1	0	0	7	10	12	9	17	12	4
R_3	0	0	1	11	24	13	9	8	5	1
R_4	0	0	0	7	7	8	7	15	15	13
R_5	0	0	1	2	0	2	6	5	12	44
R_6	0	0	2	15	16	17	10	6	5	1
R_7	0	0	2	16	5	9	20	11	6	3
R_8	0	0	2	10	7	10	11	9	17	6
R_9	0	49	23	0	0	0	0	0	0	0
R_{10}	0	23	40	5	3	1	0	0	0	0

Table 8 Crude oil simple ratios summarized

Ratios	Highest Rank	Second Highest Rank
R_1	Rank 1	Rank 3
R_2	Rank 8	Rank 6/Rank 9
R_3	Rank 5	Rank 6
R_4	Rank 8 /Rank 9	Rank 10
R_5	Rank 10	Rank 9
R_6	Rank 6	Rank 5
R_7	Rank 7	Rank 4
R_8	Rank 9	Rank 7
R_9	Rank 2	Rank 3
R_{10}	Rank 3	Rank 2

$$\frac{60 * 55 + 10 * 54 + 2 * 53}{72} = 54.805$$

Firstly, we carried out this procedure for the simple ratios. We found out that R_1, R_3, R_6, R_9 and R_{10} were the most informative ratios. We carried out this procedure for all the 55 ratios and found out that R_1R_1, R_1R_7, R_1R_9, R_1R_{10}, R_2R_2, R_2R_9, R_4R_8,

Table 9 Ranking of Gold Futures Ratios according to the KL Criterion

Ratios	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10
R_1	16	9	6	2	3	4	6	14	6	6
R_2	35	6	6	6	4	0	1	3	9	2
R_3	0	5	10	14	17	15	5	4	2	0
R_4	4	8	19	8	4	5	6	9	9	0
R_5	12	34	13	8	4	5	6	9	9	0
R_6	0	1	3	11	20	21	6	8	2	0
R_7	0	0	0	1	5	8	11	14	18	15
R_8	2	3	3	10	6	10	2	1	4	31
R_9	0	1	2	7	6	7	7	7	17	18
R_{10}	3	5	10	5	4	1	27	12	5	0

Table 10 Gold Futures simple ratios summarized

Ratios	Highest Rank	Second Highest Rank
R_1	Rank 1	Rank 8
R_2	Rank 1	Rank 9
R_3	Rank 5	Rank 6
R_4	Rank 3	Rank 8/Rank 9
R_5	Rank 2	Rank 3
R_6	Rank 6	Rank 5
R_7	Rank 9	Rank 10
R_8	Rank 10	Rank 4/Rank 6
R_9	Rank 10	Rank 9
R_{10}	Rank 7	Rank 8

Table 11 Comparing ratios of both data sets

Common Simple ratios	Common squared and cross ratios
R_1	R_2R_2
R_5	R_2R_9
	R_4R_8
	R_7R_8
	R_8R_8

R_7R_8, R_8R_8, R_8R_9 and R_8R_{10} were the most informative ratios (See Tables 12 and 13 in Appendix A).

We did a similar exercise for Gold Futures data. The results are shown below:

Our analysis of the above two tables tells us that R_1, R_2, R_3 and R_5 are the most informative simple ratios (using the first two best ranks technique, discussed earlier). We developed a ranking matrix for the derived ratios which was then used for

Table 12 Scores for Crude Oil Simple Ratios

Ratio	Score
R_1	10
R_2	4.05
R_3	5.07
R_4	3.43
R_5	1.38
R_6	5.16
R_7	4.93
R_8	4.01
R_9	8.68
R_{10}	8.12

Table 13 Scores for 20 most informative Crude Oil derived ratios

Ratio	Score	Ratio	Score
R_1R_1	55	R_3R_7	28.33
R_1R_7	40.14	R_7R_5	27.65
R_1R_9	39.29	R_5R_6	26.91
R_1R_{10}	37.81	R_4R_5	26.15
R_2R_2	36.56	R_6R_7	24.13
R_2R_9	34.37	R_4R_4	22.31
R_4R_8	33.12	R_3R_3	20.18
R_7R_8	31.87	R_1R_8	19.50
R_8R_9	31.85	R5R10	16.22
R_8R_{10}	30.51	R_2R_5	14.31

Table 14 Scores for Gold Futures Simple Ratios

Ratio	Score
R_1	6.03
R_2	7.65
R_3	6.05
R_4	5.87
R_5	8.52
R_6	5.37
R_7	2.97
R_8	3.84
R_9	3.42
R_{10}	6.30

Table 15 Scores for 20 most informative Gold Futures derived ratios

Ratio	Score	Ratio	Score
R_1R_1	43.31	R_3R_7	29.33
R_1R_7	40.12	R_7R_5	27.35
R_1R_9	39.35	R_5R_6	26.92
R_1R_{10}	37.61	R_4R_5	26.10
R_2R_2	35.56	R_6R_7	24.31
R_2R_9	33.37	R_4R_4	22.32
R_4R_8	33.12	R_3R_3	18.10
R_8R_8	31.82	R_1R_8	15.40
R_7R_9	31.69	R5*R*10	12.88
R_8R_{10}	29.51	R_2R_5	11.21

formulating the scores for them. We observed that R_1R_1, R_1R_7, R_1R_9, R_1R_{10}, R_2R_2, R_2R_9, R_4R_8, R_7R_7, R_7R_8, R_8R_8 and R_8R_9 were the most informative ratios. Also, we found out that R_1, R_2, R_4, R_5 and R_{10} were the most informative simple ratios (See Tables 14 and 15 in Appendix A). After calculating the rankings, we developed the rank matrices.

To be able to compare patterns with one similarity threshold the data was normalized. The scaling was done by dividing opening, highest, lowest, closing price for every day and closing price for previous day with the today's opening. For the Crude Oil data. We then calculated the rank of the matrix comprising of 10 simple ratios. The rank was observed to be 5. Further, we calculated the rank of matrix comprising of 55 squared and cross ratios. The rank was observed to be 11. Similarly for the Gold Futures Data, the rankings for the two corresponding matrices were found to be 5 and 11 respectively. We summarize below the most common informative ratios for both the data:

Simulated Data

We first analyzed the crude oil futures data. For the simple ratios (that is, the 10 fundamental ratios) and cross and squared ratios, we know that that the ranking technique remains the same. Also, we used the same scoring method. The reason behind using simulated data was to understand the behaviour of the ratios in the case of multiple realizations. (It should be noted that in theory, a time series data is just one of the infinite realizations of a process, each beginning infinitely long back in time and continuing in perpetuity) we got the following results:

Using the scoring technique, as discussed in the previous subsection, we found out that R_1, R_5, R_6, R_9 and R_{10} are the most informative among the simple ratios. Further, we proceeded to develop a ranking for the derived ratios and thereafter, the scores were formulated. We found that R_1R_1, R_1R_7, R_1R_9, R_1R_{10}, R_2R_2, R_2R_9, R_4R_8, R_7R_8, R_8R_8, R_8R_9 and R_8R_9 were the most informative ratios (See Tables 16 and 17 in Appendix B).

Table 16 Scores for Crude Oil Simple Ratios

Ratio	Score
R_1	9.97
R_2	4.15
R_3	5.06
R_4	3.43
R_5	1.18
R_6	5.26
R_7	4.93
R_8	4.21
R_9	8.38
R_{10}	8.22

Table 17 Scores for 20 most informative Crude Oil derived ratios

Ratio	Score	Ratio	Score
R_1R_1	42.21	R_3R_7	28.33
R_1R_7	41.87	R_7R_5	27.65
R_1R_9	40.01	R_5R_6	26.91
R_1R_{10}	36.85	R_4R_5	26.15
R_2R_2	36.56	R_6R_7	24.13
R_2R_9	34.37	R_4R_4	22.31
R_4R_8	33.12	R_3R_3	20.18
R_7R_8	31.87	R_1R_8	19.50
R_8R_9	31.85	R5R10	16.22
R_8R_{10}	30.51	R_2R_5	14.31

It is evident from the crude oil data that there is a possibility for using other ratios in calculating new and more efficient (in terms of information content) volatility estimates That is, this finding opens up the scope for creating new combinations of ratios which remain hitherto untouched by the literature. Having developed the rankings for the crude oil data, it was important to see whether other assets show similar patterns (that is, the possibility of using other ratios in volatility estimation). Therefore, we did a similar exercise for Gold Futures data, and found out the following results as shown below:

Having developed a ranking for the simple ratios, we intended to see if the square and cross product ratios provided us with some information. In other words, were there any derived ratios that ranked higher than the other ratios when ranked for seventy two quarters. If that indeed was the case, we could possibly incorporate them in estimating volatility. Therefore, we first developed a ranking for the derived ratios, and then the scores for each one of them. It was observed that R_1R_1, R_2R_7, R_2R_6, R_1R_{10}, R_2R_2, R_2R_9, R_4R_8, R_7R_8, R_8R_8, R_8R_9 and R_8R_{10} were the most informative ratios. Similarly, for the Simple ratios, R_1, R_2, R_3, R_5 and R_{10} were found to be

Table 18 Scores for Gold Futures Simple Ratios

Ratio	Score
R_1	6.03
R_2	7.69
R_3	6.15
R_4	5.47
R_5	8.42
R_6	5.31
R_7	2.91
R_8	3.74
R_9	3.12
R_{10}	6.25

Table 19 Scores for 20 most informative Gold Futures derived ratios

Ratio	Score	Ratio	Score
R_1R_1	42.81	R_3R_7	28.33
R_1R_7	40.12	R_7R_5	27.35
R_1R_9	39.35	R_5R_6	26.92
R_1R_{10}	37.21	R_4R_5	26.10
R_2R_2	35.46	R_6R_7	24.31
R_2R_9	33.37	R_4R_4	21.62
R_4R_8	33.12	R_3R_3	18.03
R_8R_8	31.82	R_1R_4	15.41
R_7R_9	31.69	R5R10	12.88
R_8R_{10}	29.51	R_2R_3	10.21

the best ones in terms of their information content (See Tables 18 and 19 in Appendix B). Further, we proceeded to calculate the rank of the matrices for the simulated data for both of the commodities. We found out that the ranks were 5 and 11 respectively for both the commodities. In other words, from a larger set of ratios, only a few could be termed as unique in the sense that no other ratios could be used to express them. Furthermore, we wanted to see if there was any commonality in the ranking patterns of the two commodities. In order to check that, we looked for the most common and informative ratios among the commodities. What we found out that the results were the same as that from the real data. We summarize below the most common informative ratios for both the assets:

Therefore, we can conclude that the real and the simulated data are not at variance with each other, to a large extent. This holds true for both the simple and the derived ratios. So, the rankings that we observe for both of the commodities could be expected to be stable. Also, one important thing that needs to be reiterated here is that these rankings are data dependent. Therefore, they might vary from commodity to commodity. Thus, the rankings shown above only reflect the fact that more informative ratios might exist (without strongly emphasizing on what these ratios might be).

Conclusion

The paper addresses certain issues regarding volatility estimation using extreme value estimators employing information theoretic measures which as never been recognized as a possible approach to information assessment of estimators in the literature. These include the use of entropy and and Kullback–Leibler distance. Kullback–Leibler (KL) divergence metric was used to give an estimate of the amount of information contained in extreme value ratios. Since uniform distribution has highest entropy value and contains least information, the extent of deviation from uniform distribution measured the degree of variability in the ratio (both simple and cross product ratios). The five input observations were employed to form a set of 10 simple ratios and subsequently 55 squared and cross product ratios. We looked at the ratios which have minimum entropy and show greater divergence from maximum entropy distribution i.e., the uniform distribution. The empirical analysis was conducted on two simulated data sets for Gold futures and Crude oil futures. Both the data sets showed similar results in the following way: The results on the rank of matrix of both sets of ratios indicate that out of ten simple ratios, five are linearly independent and out of fifty-five ratios, only eleven are linearly independent. This was true for both the data sets. It shows that the eleven ratios cannot be represented as a linear combination of the remaining ratios in the set. Those eleven constitute the important ratios in the set of fifty five ratios. The widely used traditional estimators employed the subset of these 55 ratios (i.e. 55 squared cross product) and contained not more than six ratios in estimating volatility. On the basis of results arrived, we can conclude that the other ratios also contain more information and may be used in calculating extreme value estimators.)

In addition to that, both the results gave strikingly different results when we ranked these ratios according to the KL metric. The five more informative simple ratios in both data sets had only two ratios in common, namely R_1 and R_3. The eleven more informative squared and cross product ratios in both data sets had only five ratios in common namely, R_2R_2, R_2R_9, R_4R_8, R_7R_8 and R_8R_8. R_8R_8 is used in Parkinson estimator. R_8R_8 and R_4R_8 are used in GK estimator. What must be noted is that these results hold true for both the simulated and the real data, in this case. We cannot be certain about whether for different assets different sets of informative ratios exist or not. (Which also leads to the idea that there must be further investigation into the information content of different ratios on an asset to asset basis.) The traditional estimators used a single formula for all the stock prices which may not be the case. If one varies their data sets, the ratios for estimating volatility may be changed in computing

volatility for better efficiency. A question remains though. If we recognize the fact that there could more informative ratios on an asset to asset basis, what combinations of them could give us more efficient estimators (and possibly lead to the creation of a family of estimators).

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix

Appendix A: Score Tables for Real Data

See appendix Tables 12, 13, 14, 15, 16.

Appendix B: Score Tables for Simulated Data

See appendix Tables 16, 17, 18, 19.

References

Alizadeh, S., M. Brandt, and F.X. Diebold. 2002. Range-Based Estimation of Stochastic Volatility Models. *Journal of Finance* 57: 1047–1092.
Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedas- ticity. *Journal of Econometrics* 31 (3): 307–327.
Beckers, S. 1983. Variances of security price returns based on high, low, and closing prices. *Journal of Business* 56 (1): 97–112.
Ding, Z., and C. Granger. 1996. Varieties of long memory models. *Journal of Econometrics* 73 (1): 61–67.
Engle, R. 1982. Autoregressive Conditional Heteroscedasticity with Esti- mates of the Variance of United Kingdom Inflation. *Econometrica* 50 (4): 987–1007.
Garman, M.B., and M.J. Klass. 1980. On the estimation of security price volatilities from historical data. *Journal of business* 53 (1): 67–78.
Horst, E.T., A. Rodriguez, H. Gzyl, and G. Molina. 2012. Stochastic volatil- ity models including open, close, high and low prices. *Quantitative Finance* 12 (2): 199–212.
Kullback, S., and R.A. Leibler. 1951. On information and sufficiency. *The Annals of Mathematical Statistics* 22 (1): 79–86.
Mandelbrot, B. 1963. The Variation of Certain Speculative Prices. *The Journal of Business* 36 (4): 394–419.
Markowitz, H. 1952. Portfolio Selection. *The Journal of Finance* 7 (1): 77–81.
Meilijson, I. (2008). The garman-klass volatility estimator revisited. arXiv preprint arXiv:0807.3492.
Ong, M.A. 2015. An information theoretic analysis of stock returns, volatility and trading volumes. *Applied Economics* 47 (36): 3891–3906.
Ozturk, H., U. Erol, and A. Yuksel. 2016. Extreme value volatility estimators and realized volatility of istanbul stock exchange: Evidence from emerging market. *International Journal of Economics and Finance* 8 (8): 71.
Parkinson, M. 1980. The extreme value method for estimating the variance of the rate of return. *Journal of business* 53 (1): 61–65.
Philippatos, G.C., and C.J. Wilson. 1972. Entropy, market risk, and the selection of efficient portfolios. *Applied Economics* 4 (3): 209–220.
Rogers, L.C.G., and S.E. Satchell. 1991. Estimating variance from high, low and closing prices. *The Annals of Applied Probability* 1 (4): 504–512.
Ross, S.A. (1981). The Recovery Theorem. *NBER Working Paper No. 17323*.

Samuelson, P.A. 1970. The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments. *The Review of Economic Studies* 37 (4): 537–542.

Shu, J., and J.E. Zhang. 2006. Testing range estimators of historical volatil- ity. *Journal of Futures Markets* 26 (3): 297–313.

Taylor, S.J. 1986. *Modeling financial time series*. New York: Wiley.

Todorova, N., and S. Husmann. 2012. A comparative study of range-based stock return volatility estimators for the german market. *Journal of Futures Markets* 32 (6): 560–586.

Yang, D., and Q. Zhang. 2000. Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business* 73 (3): 477–492.