Information Management and Business Review
Vol. 2, No. 6, pp. 252-258, June 2011

Martingales, Efficient Market Hypothesis and Kolmogorov's Complexity Theory

Amaresh Das
Southern University at New Orleans, New Orleans, LA, USA
adas2@cox.net

Abstract: Efficient market theory states that financial markets can process information instantly. Empirical observations have challenged the stricter form of the efficient market hypothesis (EMH). These empirical observations and theoretical considerations show that price changes are difficult to predict if one starts from the time series of price changes. This paper provides an explanation in terms of algorithmic complexity theory of Kolmogorov that makes a clearer connection between the efficient market hypothesis and the unpredictable character of stock returns.

Keywords: *Martingales, martingales difference sequence, Shannon's information theory, algorithmic complexity theory, efficient market hypothesis,*

1. Introduction

An investment theory that states it is impossible to "beat the market" because stock market efficiency causes existing share prices to always incorporate and reflect all relevant information. According to the efficient market hypothesis stocks always trade at their fair value on stock exchanges, making it impossible for investors to either purchase undervalued stocks or sell stocks for inflated prices. As such, it should be impossible to outperform the overall market through expert stock selection or market timing, and that the only way an investor can possibly obtain higher returns is by purchasing riskier investments. In finance, the efficient-market hypothesis asserts that financial markets are "information efficient".

Empirical analyses have consistently found problems with the efficient-market hypothesis, the most consistent being that stocks with low price to earnings (and similarly, low price to cash - flow or book value) outperform other stocks. Alternative theories have proposed that 'cognitive bases' (over confidence, overreaction, representative bias etc) these inefficiencies, leading investors to purchase over priced growth stocks rather than value stocks. Although the efficient-market hypothesis has become controversial because substantial and lasting inefficiencies are observed Beechey *et al*. (2000) consider that it remains a worthwhile starting point.

The efficient market hypothesis as formulated in economics and finance, first by Samuelson (1965) and then by Fama and French (1992), suggest that properly anticipated prices fluctuate randomly. Using the hypothesis of 'rational expectations' and market efficiency, he was able to demonstrate how yt+1 , the expected value of the price of a given asset at time t+1, is related to the previous values of prices y0, y1, ...yt through the relation

$$E[\, y_{t+1} \quad y_0, y_1, \ldots . y_t \,] = \text{yt} \qquad (1)$$

Stochastic processes obeying the conditional probability given in equation (1) are called martingales. Martingales are very important types of sequences, as shown in Davidson and Mackinnon (1993).

Definition 1
A sequence y_t of random variables is called a martingale if, for all t, E $|y_t|$ exists and is finite and for all t

$$E[\, y_{t+1} \mid y_0,\ y_1, \ldots y_t = y_t \qquad (2)$$

Martingales do crop up as such from time to time in econometrics. An important notion surrounding the concept is that of a martingale difference sequence.

Definition 2

A sequence $\{y_t\}$ is said to be a martingale difference sequence if[1] (3)

$$E[\, y_{t+1} \mid y_0, y_1, \ldots y_t] = 0$$

Theorem 1

If $\{y_t\}$ is a martingale difference sequence and there is an r ≥ 1 such that the series

$$\sum_{t=1}^{\infty} t^{-(1+r)} E(|y_t|^{2r}) \qquad (4)$$

Converges, then

$S_n \to 0$ almost certainly[2]

Between the cash flow dates, there is a constant drift in the asset price and most likely, there will also be a finite drift if fluctuations are included. Such drifts are present in most real markets. A modified statement then is that, up to the drift dF /dt, the market does not expect a net change of the true or fundamental, prices. However, deviations of a certain amplitude *y,* where *y = F (t) – F (0),* occur with probabilities *p (y),* which satisfy

$$\int_{-\infty}^{\infty} p(y)\, dy = 1 \qquad (5)$$

For all *t.* The expected profit from an investment is then

$$E(y) \equiv = \int_{-\infty}^{\infty} y\, p(y)\, dy > 0, \quad \text{so long as } dF/dt > 0 \qquad (6)$$

Such an investment is not a 'fair game' of chance because it has a positive expectation. However for a fair game of chance: $E(y) = 0$ (7)

The fair game (random walk) condition about the price changes observed in a financial market is equivalent to the statement that there is no way of making a profit on an asset by simply using its recorded history of its price fluctuations. This conclusion is the 'weak form' of the efficient

---

[1] This condition is very short because the condition implies not only the existence of the unconditional expectations E (y_t) but also that these are zero and the sequence is therefore centered.

[2] The regularity condition is quite weak, because of the factor of $t^{-(1+r)}$. Note that

$$\sum_{t=1}^{\infty} t^{-(1+r)}$$

Converges for all r > 0. In particular the condition is satisfied if the $(2k)^{th}$ absolute moments of the y_t' s, $E(|y_t|^{2r})$ are uniformly bounded, by which we mean that there is a constant k, independent of t, such that $E(|y_t|^{2r}) < k$ for all t (see Stout (1974).

market hypothesis. This price changes are unpredictable from the historical time series of those changes.

Since the 1960s, a great number of empirical investigations have been devoted to testing the efficient market hypothesis. In the great majority of the empirical studies, the time correlation between price changes has been found to be negligibly small, supporting the efficient market hypothesis. However, it was shown in the 1960s that by using the information present in additional time series such as earnings/ price ratios, dividend yields and time-structure variables, it is also possible to make predictions of the rate of return of a given asset on a long time scale, much larger than a month. In actual markets, residual inefficiencies are always present. These empirical observations and theoretical considerations show that price changes are difficult to predict if one starts from the time series of price changes.

2. Kolmogorov Complexity Theory (KCT): Introductory remarks

The description of a fair game in terms of martingale is rather formal and mathematically complex. The problem of algorithmically constructing prices that reflect all available information has been studied and studied extensively by applying a computer. In this section, we will provide an explanation in terms of algorithmic complexity theory of Kolmogorov. For an excellent treatment of Kolmogorov, readers are asked to refer to Kolmogorov (1965), Cohn and Kumar (2007) and Ziegler and Koolen (2008). In order to describe Kolmogorov's contribution, we begin with the fundamental wok of Shannon's theory of information that provides much useful information in terms of bits or more generally, in terms of the normal amount of the complexity of structures needed to encode a given piece of information.

Shannon's Information theory: Suppose a random event has a discrete state $x_{1 \ldots Xn,}$ each with a probability $p1_{\ldots}P_{n,}$ the information value of x is a reduction of uncertainty

$$H(x) = -\sum_{j=1}^{n} P_j \log P(P_j) \qquad (8)$$

The right hand side is the entropy factor introduced by Boltzman (1974) is called the general function for information.

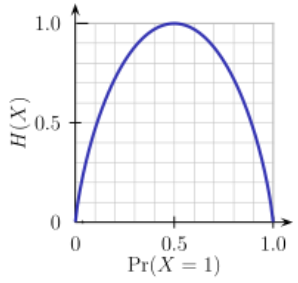
Analogously for a continuous distribution

$$H = -\int p(x) \log p(x)\, dx \qquad (9)$$

The most important result for Shannon's theory of information is the following:

$$R = H(x) - H_y(x) \qquad (10)$$

The amount $H_y(x)$ is the conditional entropy, also called, equivocation. The amount of information one receives would be equal to the amount of information sent minus the average rate of conditional entropy. As technology improves, the level of equivocation gradually reduces to more people. That means the value of information known to everyone is Zero. The laws that govern human activities, including mental activities, are the same as the physical laws that govern non-living objects. H is a function of probability of a given event. Value is decreasing function of probability. In information theory, *P* is the probability of some event



Note: Entropy of a Bernoulli trial as a function of success probability, often called the binary entropy function, $H_b(p)$. The entropy is maximized at 1 bit per trial when the two possible outcomes are equally probable, as in an unbiased coin.

P = 1 – log P = 0 when *P* approaches zero, - *log P* approaches infinity; the value of information is very high. If the information is announced publicly and becomes known to people, the value of information is very low. Little profit can be made by trading such information. Warren buffet who has a very successful record for gaining and using insightful market information would not announce to public what stock he is going to buy. $H_y(x)$ offers the quantitative measure of information asymmetry. The amount of conditional entropy is determined by the correlation between the sender and the receiver. It is impossible to assume how much information one has from the information source. When x and y are independent $H_y(x) = H(x)$ and $R = 0$. No information can be transmitted between the two objects that are independent of each other. When the correlation between x and y are equal to one, $H_y(x) = 0$. No information loss occurs in transmission. The higher the correlation between the source and the receiver, the more information can be transmitted. Social impact of new product is gradually reached over the path of several decades. That is why the individual stocks and whole markets often exhibit cycles of different lengths. The whole analysis is in contrast to Grossman and Stiglitz where economic agents recognize the value of information instantly.

Communications over a channel - such as an ether net wire - is the primary motivation of information theory.[1] How much information can one hope to communicate over a noisy (or otherwise imperfect) channel? Let $p(y \mid x)$ be the conditional probability distribution function of y given x. We will consider $p(y \mid x)$ to be an inherent fixed property of our communications channel (representing the nature of the noise of our channel). Then the joint distribution of x and y is completely determined by our channel and by our choice of $f(x)$, the marginal distribution of messages we choose to send over the channel. Under these constraints, we would like to maximize the rate of information, or the signal, we can communicate over the channel. The appropriate measure for this is the mutual information, and this maximum mutual information is called the channel capacity and is given by:

$$C = \max I(x; y) \qquad (11)$$

This capacity has the property that is related to communicating at information rate R (where R is usually bits per symbol). For any information rate $R < C$ and coding error $\varepsilon > 0$, for large enough N, there exists a code of length N and rate ≥ R and a decoding algorithm, such that the maximal probability of block error is ≤ ε; that is, it is always possible to transmit with arbitrarily small block error. In addition, for any rate $R > C$, it is impossible to transmit with arbitrarily small block error. Optimal codes that can be found to transmit data over a noisy channel with a small coding error at a rate near the channel capacity.

Kolmogorov's Algorithmic Complexity Theory: Information as we have known by now says a random object $x \sim p(x)$ has a complexity (entropy) $H = \sum p(x) \log p(x)$ with the attendant interpretation that H bits are sufficient to describe X on the average. Algorithmic complexity says an object x has a complexity $k(x)$ equal to the length of the shortest (binary) program that describes x. It is a beautiful fact that these ideas are much the same. In fact, it is roughly true that $EK(x) = H$. Moreover, if we let $p_U(x) = p_r$ { *prints* x } be the probability that a given computer U prints x when given a random program, it can be shown that $\log(1/p_U(x)) = K(x)$ for all x, thus establishing vital link between the 'universal' probability measure p_U and the 'universal' complexity K. More on this later. As an author so eloquently put it: The concepts of information theory as applied to infinite sequences give rise very interesting investigations, which, without being indispensable as a basis of probability theory, can acquire a certain value in the investigation of the algorithm side of mathematics as a whole.

Different kinds of Kolmogorov complexity are available: the uniform complexity, prefix complexity, monotone complexity, time-bounded Kolmogorov complexity, and space-bounded Kolmogorov complexity. But, the most popular types of computational complexity are the 'time' complexity of a problem equal to the number of steps that it takes to solve an instance of the problem as a function of the size of the input (measured in bits) using the most efficient algorithm, Burgisser et al (1997), and Tucker and Jucker (2001) classifies different computational problems by complexity class. But whatever be the way of classification, the complexity theory addresses computational problems and not particular problem instances. In computational complexity theory, a problem refers to the abstract question to be solved. In contrast, an instance of this problem is a rather concrete utterance, which can serve as the input for a decision problem. The instance is a particular input to the problem, and the solution is the output corresponding to the given input.

Definition 3:
The definition of algorithmic complexity

$$k(x) = \min_{U(p)=x} l(p) \tag{12}$$

of a string *x* as the length of the shortest program for a universal computer U to output *x* have immediate impact.

Theorem 2*:*
Kolmogorov theorem, which is known as the *invariance theorem,* says that the notion of complexity can be made fairly independent of the choice of the interpreter, that is, there are 'asymptotically optima' interpretation $U(p)$ with the property that for any other computable partial function $\varphi(p)$ we have the inequality

$$K_U(x) \le K_\varphi(x) + c_\varphi \tag{13}$$

where the constant *c* does not depend on *x.*
Kolmogorov considered an arbitrary computable partial function φ and defined $k_\varphi(x) = x$; the function φ acts as the decoder or interpreter (when information is reached to the receiver) of the description *p*. The complexity of x is the length of the shortest description, with respect to the interpreter φ.
This theorem can be proved simply by choosing *U* as the machine universal as defined above. Indeed the universality of the of *U implies* that there is a binary string q_φ with the property that for all p we have

$$\varphi(p) = U(q_\varphi p) \tag{14}$$

It is the algorithmic complexity theory that makes a clearer connection between the efficient market hypothesis and the unpredictable character of stock returns. Such a connection is now supported by the property that a time series that has a dense amount of non-redundant economic information (as the efficient market hypothesis requires for the stock returns time series) exhibits statistical features that are almost indistinguishable from those observed in a time series that is random[3].

[3] An axiomatic approach to Kolmogorov is based on Blum axioms (Blum, 1967). It is possible to treat different kinds of Kolmogorov complexity as particular cases of axiomatically defined generalized Kolmogoov complexity. Instead, of proving similar theorems, such as the basic invariance theorem, for each particular measure, it is possible to easily deduce all such results from one corresponding theorem proved in the axiomatic setting. This is a general advantage of the axiomatic approach in mathematics. The axiomatic approach to Kolmogorov complexity was also applied to software metrics, e.g., Debnath and Burgin, (2003) and Burgisser, *et al* (1997).

3. Efficient Market Hypothesis and KCT

Within the algorithmic complexity theory, as we have now seen, the complexity of a given object coded in an n - digit binary sequence is given by the bit length $K^{(n)}$ of the shortest computer program that can print the given symbolic sequence. Such an algorithm does exist and is asymptotically optimal (invariance theorem). To illustrate this concept suppose that as a part of space exploration we want to transport information about the scientific and social achievement of the human race to regions outside the solar system. Among the information blocks we include, we transmit the value of π expressed as a decimal carried out to 125,000 places and the time series of the daily values of the Dow Jones industrial average between 1898 and the year of the space exploration (approximately 125,000 digits). To minimize the amount of storage space and transmission time needed for these two items of information, we write the two number sequences using, for each series, an algorithm that makes use of the regularities present in the sequence of digits. The best algorithm found for the sequence of digits in the value of π is extremely short. In contrast, an algorithm with comparable efficiency that has not been found for the time series is a non-redundant time series.

Just as Shannon's theory describes the maximum possible efficiency of error correcting methods and various levels of noise interference and data corruption, within algorithmic complexity theory, a series of symbols is considered inefficient or unpredictable if the information embodied cannot be 'compressed' or reduced to a more compact form. This statement is made more formal by saying that the most efficient algorithm reproducing the original series of symbols has the same length as the symbol sequence itself.
Algorithm complexity theory, therefore, will help us understand the behavior of a financial time series[4]. In particular,

- Algorithm complexity theory will make a clear connection between the efficient market hypothesis and the unpredictable character of stock returns. Such a connection is now supported by the fact that a time series that has a dense amount of non - redundant economic information (as the efficient market hypothesis requires) exhibits statistical features that are almost indistinguishable from those observed in a time series that is random.
- Measurement of the deviation from randomness provides a tool to verify the validity and Limitations of the efficient market hypothesis.
- From the point of view of algorithmic complexity theory, it is not possible to determine between trading on 'noise' and trading on information where we use information to refer to fundamental information concerning the traded asset, internal or external to the market. Algorithmic theory detects no difference between a time series carrying a large number of non-redundant economic information and a pure random process.

4. Conclusion

Investor performance and market patterns are primarily information driven. However, theories of finance offer little guidance in identifying informed investors and in distinguishing between securities with scarce information and those with widely available information. Most empirical evidences about market behaviors documented in the literature can be explained by Kolmogorov's algorithm complexity information theory that can be generalized by Shannon's

[4] Decision problems are one of the central objects of study in computational complexity theory. Malajovich (2002) , Basu et al (2006) , and Tucker and Jucker (2001)view decision problem as a formal language where the members of the language are instances whose answer is yes, and the non-members are those instances whose output is no. The objective is to decide, with the aid of an algorithm whether a given input string is member of the formal language under consideration. If the algorithm deciding this problem returns the answer, *yes*, the algorithm is said to accept the input string, otherwise it is said to reject the input. The input is actually an arbitrary graph.

entropy theory of information. Investor performance and market patterns are the results of information processing by investors of different sizes with different background knowledge.

This property of financial time series looking unpredictable and their future values being essentially impossible to predict, is not a manifestation of the fact that the time series of financial asset prices does not reflect any valuable and important economic information. Indeed the opposite is true. The time series of prices in a financial market carries a large amount of non-redundant information. Because the quantity of information is so large, it is difficult to extract a subset of economic information associated with specific aspect. The difficulty in making predictions is thus related to an abundance of information of financial data, not to a lack of it. Therefore, what is needed an efficient algorithm that will help us understand better the behavior of a financial time series.

References

Basu, S., Pollack, R. & Roy, M. F. (2006). Algorithm in Real Algebraic Geometry (Algorithm and Computational Mathematics) Springer–Verlag, New York.

Beechey, M. Gruen, D. & Vickrey, J. (2000). The Efficient Market Hypothesis: A Survey. Reserve Bank of Australia.

Blum, M. (1967). On the Size of Machines. *Information and Control*, 11(3), 257-265.

Boltzmann, L. (1974). The Second Law of Thermodynamics. Popular Schriften, Essay 3, Address to a formal meeting of the Imperial Academy of Science, 29 May 1886, reprinted in Ludwig Boltzmann, Theoretical Physics and Philosophical Problem, S. G. Brush (Trans). Boston: Reidel. (Original work published 1886).

Burgisser, P., Clausen, M. & Shokrollahi, M. A. (1997). *Algebraic Complexity Theory*. Springer–Verlag, New York.

Cohn, H. & kumar, A. (2007). Universally Optimal Distribution Points in Spheres. *Journal of American Mathematical Society,* 20(1), 99-148.

Davidson, R. & Mackinnon, J. (1993). Estimation and Inference in Econometrics Oxford University Press.

Debnath, N. C. & Burgin, M., (2003). *Software Metrics from the Algorithmic Perspective,* Proceedings of the ISCA 18th International Conference Computers and their Applications, Honolulu, Hawaii, 279-282.

Fama, E. & French, K. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance,* 47(2), 427-465.

Kolmogorov, A. N. (1965). Three Approaches to the Quantitative Definition of Information, *Problems of Information Transmission*, 1(1), 1-7.

Malajovich, G. (2002). Lower Bounds for Some Decision Problems over C. *Theoretical Computer Science*, 276(1-2), 425-434.

Samuelson, P. A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6(2), 41-49.

Stout, W. F. (1974). Almost Sure Convergence, New York, Academic Press.

Tucker J. V. & Jucker, J. I. (2001). Computable Functions and Semi Computable Sets on Many-Sorted Algebras, Handbook of Logic in Computer Science, Logic and Algebraic Methods, Oxford University Press.

Ziegler, M. & Koolen, W. (2008). Kolmogorov Complexity Theory over the Real's. Electronic Notes *Theoretical Computer Science*, 221, 153-169.