



Algorithmic Social Sciences Research Unit

ASSRU

Department of Economics
University of Trento
Via Inama 5
381 22 Trento, Italy

DISCUSSION PAPER SERIES

4 – 2012/1

ALGORITHMIC COMPLEXITY OF FINANCIAL MOTIONS

Olivier Brandouy, Jean-Paul Delahaye, Lin Ma and Hector Zenil*

JANUARY 2012

* IHPST, Université de Paris 1 (Panthéon-Sorbonne), France, Department of Computer Science, University of Sheffield, UK & ASSRU *Honorary Founding Fellow*. Dr Zenil's fellow-authors are at, respectively, Sorbonne Graduate Business School (IAE), Université de Paris 1, France, Laboratoire d'Informatique Fondamentale de Lille, Université de Lille 1, France and École Universitaire de Management, Université de Lille 1, France.

Abstract

We survey the main applications of algorithmic (Kolmogorov) complexity to the problem of price dynamics in financial markets. We stress the differences between these works and put forward a general algorithmic framework in order to highlight its potential for financial data analysis. This framework is “general” in the sense that it is not constructed on the common assumption that price variations are predominantly stochastic in nature.

KEYWORDS: algorithmic information theory; Kolmogorov complexity; financial returns; market efficiency; compression algorithms; information theory; randomness; price movements; algorithmic probability.

CLASSIFICATION: C43, G11

In driving the decisions made by investors, information fuels financial markets. But the market has proven to be very complex in its dynamics and therefore very hard to predict. Market price movements in themselves are unpredictable or barely predictable. Price movements can be seen as the outcome of interactions between investors following rules in their quest to reap a benefit. It has been suggested that the market alone is complex enough, even when isolated from external stimuli (see e.g. [1]), yet external information can make it less or more predictable.

These concepts are at the heart of one of the most famous hypotheses in finance: the Efficient Market Hypothesis (EMH), painstakingly reconstructed from the recently rediscovered works of Bachelier [2] and subsequently refined. The concepts shaping this hypothesis, such as “information” or “randomness,” are undoubtedly of interest to computer scientists who have their own tradition of tackling these questions. Part of this tradition can be identified with the works of Shannon, and has provoked a burgeoning literature in finance. Another part of this tradition can clearly be linked with the works of Kolmogorov [3] and Chaitin [4]. This paper attempts to assess the existing works in these fields, highlighting salient divergences and proposing a general algorithmic framework as an alternative to the mainstream probabilistic one used in financial analysis.

This article is organized as follows: after a theoretical introduction to algorithmic complexity in section 1, we take a quick look at the relation between financial theories and the randomness of price variations in section 2. As this relation is studied by some existing works applying the notion of algorithmic complexity, we provide a survey of these works in section 3 and show why they failed to propose a general algorithmic framework for financial pattern tracking, which was not available until the publication of our work in [5] and [6]. The main contributions of these two works are then sketched in sections 4 and 5 respectively.

1 Algorithmic Information Theory

At the core of algorithmic information theory (AIT) is the concept of algorithmic complexity¹, a measure of the quantity of information contained in a string of digits (or more generally of symbols or integers). The algorithmic complexity of a string is defined as the length of the shortest algorithm that, when provided as input to a universal Turing machine (an idealized computer model), generates the said string. A string has maximal algorithmic complexity if the shortest computer program able to generate it is not significantly shorter than the string itself, perhaps allowing for a fixed additive constant. The difference in length between a string and the shortest algorithm able to generate it is the string’s degree of compressibility. A string of low complexity is therefore highly compressible, as the information that it contains can be encoded in an algorithm much shorter than the string itself. By contrast, a string of maximal complexity is incompressible. Such a string constitutes its own shortest description: there is no more economical way of communicating the information that it contains than by transmitting the string in its entirety. In algorithmic information theory a string is algorithmically random if it is incompressible.

Algorithmic complexity is inversely related to the degree of regularity of a string. Any pattern in a string constitutes redundancy: it enables one portion of the string to be recovered from another, allowing a more concise description. Therefore highly regular strings have low algorithmic complexity, whereas strings that exhibit little or no pattern have high complexity.

The algorithmic complexity $K_U(s)$ of a string s with respect to a universal Turing machine U is defined as the binary length of the shortest program p that produces as output the string s .

¹Also known as program-size complexity, Kolmogorov complexity, or Kolmogorov-Chaitin complexity.

$$(1) \quad K_U(s) = \min\{|p|, U(p) = s\}$$

Algorithmic complexity conveys the intuition that a random string should be incompressible: no program shorter than the string can produce it.

Even though K is uncomputable as a function, meaning that there is no effective procedure (algorithm) for calculating it (for every string), one can use the theory of algorithmic probability to obtain exact evaluations of $K(s)$. This can be done for strings s short enough, thus for which the halting problem can be solved for a finite number of cases due to the size (and simplicity) of the Turing machines involved.

1.1 Algorithmic probability

The concept of algorithmic probability is deeply connected to algorithmic complexity and it was first developed by Solomonof [7] and formalized by Levin [8]. The intuition behind algorithmic probability has to do with weighting past experience, with experience that is closer in time deemed more relevant.

Algorithmic probability assigns to objects an a priori probability in a strong universal and objective manner. This distribution has theoretical applications in a number of areas, including inductive inference theory and the time complexity analysis of algorithms. Its main drawback is that it is not computable and thus can only be approximated in practice, as was shown in [9].

Consider an unknown process producing a binary string of length k bits. If the process is uniformly random, the probability of producing a particular string s is exactly 2^{-k} , the same as for any other string of length k . Intuitively, however, one feels that there should be a difference between a string that can be recognized and distinguished, and the vast majority of strings that are indistinguishable as regards whether or not the underlying process is truly random.

Assume one tosses a fair coin 20×3 times and gets the following outcomes:

```
00000000000000000000
01100101110101001011
11101001100100101101
```

The first outcome would be very unlikely because one would expect a patternless outcome from a fair coin toss, one that resembles the second and third outcomes. In fact, it would be far more likely that a simple deterministic algorithmic process has generated the first string. The same could be said for the market: one usually expects to see few if any patterns in its main indicators. Algorithmic complexity can capture this expectation of *patternlessness* by defining what a random-looking string looks like. On the other hand, algorithmic probability predicts that random-looking outputs are the exception rather than the rule when the generating process is algorithmic.

There is a measure based on algorithmic probability which describes the expected output of an abstract machine when running a random program. A process that produces a string s with a program p when executed on a universal Turing machine U has probability $m(s)$ [8]. As p is itself a binary string, $m(s)$ can be defined as the probability that the output of a universal (prefix-free)

Turing machine² U is s when provided with a sequence of fair coin flip inputs interpreted as a program.

$$(2) \quad m(s) = \sum_{U(p)=s} 2^{-|p|} = 2^{-K(s)+O(1)}$$

i.e. the sum over all the programs p for which the universal Turing machine U outputs the string s and halts.

Levin's universal distribution is so called because, despite being uncomputable, it has the remarkable property (proven by Leonid Levin himself) that among all the lower semi-computable semi-measures, it dominates every other³. This makes Levin's universal distribution the optimal prior distribution when no other information about the data is available, and the ultimate optimal predictor (Solomonoff's original motivation [7] was actually to capture the notion of learning by inference) when assuming the process to be algorithmic (or more precisely, carried out by a universal Turing machine).

There is no general algorithm computing the algorithmic probability for every single string. However, one way to calculate an approximation of the algorithmic probability is to calculate an approximation of the universal distribution by running a large set of abstract machines, as we did in [6].

Following a short introduction to the subject in section 2 and a short survey in section 3, sections 4 and 5 will show how lossless compression algorithms as a measure for approximating algorithmic complexity, and a numerical evaluation of a distribution approximating Levin's distribution, will be used to tackle the question of the foundations and applications of algorithmic complexity to financial markets, in particular to the analysis of stock price movements.

2 Information and randomness within the modern finance corpus

Randomness is, on the one hand, pragmatically *modeled* in empirical finance with increasingly sophisticated techniques, and, on the other hand, *explained* using theoretical frameworks where information is one of the most prominent components.

2.1 A quick glance at the empirical landscape ...

To model randomness is usually a challenge, and empirical finance is no exception in this regard. It is traditional to root this quest in the works of [2] who was probably the first to embed price motions in a rigorous probabilistic framework. This heritage is undoubtedly important in finance (for example the works of [10] on the valuation of options are largely grounded in premises that can be directly linked to Bachelier) and in Mathematics (for instance, [11] even cites Bachelier as an

²A universal Turing machine is an abstraction of a general-purpose computer. Essentially, as proven by Alan Turing, a universal computer can simulate any other computer on an arbitrary input by reading both the description of the computer to be simulated and the input thereof from its own tape. Without loss of generality one can assume the domain of a Turing machine to be prefix-free, that is, no program for it is the beginning of any other valid program.

³Since it is based on the Turing machine model, from which the adjective *universal* derives, the claim depends on the Church-Turing thesis.

inspiration for the works of [12] on diffusion processes). For example, according to this tradition, prices S evolve in time following a geometric Brownian motion:

$$(3) \quad S(t) = S_0 e^{X(t)}$$

with

$$(4) \quad X(t) = \sigma W(t) + \mu t$$

In equation 4, $W(t)$ is a Brownian motion and satisfies, among other conditions, the condition that $W(t) \sim \mathcal{N}(0, t)$

One clearly sees that at the heart of this approach, randomness is modeled with a mere *iid* Normal. This approach makes sense if one invokes the central limit theorem: prices are additionally affected by independent economic events (see for example [13]). A crucial obstacle to using this kind of model efficiently is the problem with estimating the true unobservable volatility (σ) of a given asset.

Despite its strong roots, this approach suffers from several limitations: for example, volatility is not a constant and its fluctuations themselves deserve to be studied. This leads to a more complex set of models where non-linear stochastic processes describe higher conditional moments (ARCH models are good examples [14] or jump diffusion models [15]). Nevertheless, these attempts themselves fail to predict large market events such as the 1987 crash or the flash crash of 2010.

2.2 ... and an elliptical evocation of a theoretical monument, EMH

In addition to these empirical attempts to model randomness, finance proposes theoretical explanations linking information and price motions, notably within the Efficient Market Hypothesis. Even if it is definitely impossible to sum up fifty years of research on this topic (some of the major contributions in this field being, for example those of [16], [17], [18], [19], [20], [21], [22]), the general import of this approach is that the only thing that moves prices in financial markets (for example, stock prices), is the modification of Ω_t , the relevant set of information available at date t . Rational investors should only react to this modification. Thus market fluctuations simply reflect the world's own stochastic behavior.

It is equally impossible to sum up the enormous amount of criticism this framework has provoked in recent years. Some researchers claim that the EMH over-naturalizes a mere social construction, some that evidence shows that prices evolve significantly even when no major information is released to the market – [23] –.

This theoretical link between information and the randomness of price motions is exploited in a series of empirical works where concepts from information theory, such as Shannon's entropy [24], Kolmogorov complexity [3] and stochastic complexity [25] are used to highlight the similarities and dissimilarities between random strings and real world financial data. The following section offers a review of the major works shaping this new field at the crossroads of finance and information theory.

3 Information theory and finance

As mentioned previously, at least two traditions emerge within the information theory literature when considered from the point of view of its application in finance: the Shannonian and the Kolmogorovian tradition.

It is important to distinguish i) *Shannon's* information theory from ii) the *algorithmic* one, introduced by [3, 4, 8, 7]

i) In *probabilistic* terms, [24] proposed to measure the quantity of information contained in a random variable by the value of its *entropy*. This concept is defined by the following equation:

$$(5) \quad H(X) = - \sum_{i=1}^n p(x_i) \log_b p(x_i)$$

X in equation 5 denotes a random variable with n possible outcomes: $\{x_1, x_2, \dots, x_n\}$, $p(x_i)$ is the probability associated with x_i , and b , an arbitrarily chosen logarithmic base, whose common value is 2, in order to make the unit of $H(X)$ equal to “one bit”.

Exploiting this central concept, several researchers have tackled traditional questions in finance: for instance, [26] investigated the optimum of log-optimal portfolios. Another example is [27], who attempted to interpret investor behavior within the framework of Shannon's information theory.

From an empirical point of view, [28] estimated the Shannon entropy for NASDAQ and the Mexican IPC data, and found close resemblances between financial returns and random strings. [29, 30] undertook to measure financial risk with Shannon's entropy and illustrated their ideas with data from the Portuguese stock market.

The power of these empirical techniques lies in their ability to take into account high-order dependencies in financial dynamics. However, formulated within a *probabilistic* framework, they remain focused on statistical patterns, as classical econometric trend-detecting tools do.

ii) Departing from the probabilistic tradition, in *algorithmic* information theory, the quantity of information contained in a string, s , as mentioned previously, is measured by its algorithmic complexity, $K(s)$ (see equation 1). links the notions of compressibility and predictability, and permits the use of compression tools for financial trend tracking (if there are any such).

Some authors have tried to apply these notions in empirical investigations of real world series. For example, [31, 32] exploited econometric models to predict financial returns and estimated the stochastic complexity⁴ of a stock market using the sum of the squared prediction errors they obtained from it. [33] measured the complexity of stock markets with the highest successful prediction rate⁵ (SPR) that one can achieve with different compression techniques. [34, 35] used the Variable Order Markov model (*VOM*, a variant of context predicting compression tools) to predict the direction of financial returns. They found a significant difference between the SPR obtained from financial data and that obtained from random strings. To establish a formal link between this result and the Efficient Market Hypothesis (EMH), the authors also simulated *VOM*-based trading rules on Forex time series and concluded that there were no abnormal profits.

⁴The notion of stochastic complexity is proposed by [25] replacing the universal Turing machine in the definition of Kolmogorov complexity by a class of *probabilistic* models.

⁵To obtain this successful prediction rate, at each step, the author uses compression algorithms to predict the direction of the next return, and calculates the rate of successful predictions for the whole series.

Exploiting another compression technique, [36] and [37] ranked stock markets all over the world according to their LZ index, an indicator showing how well the compression algorithm proposed by [38] works on financial returns.

Despite the expansive perspectives opened up by these pioneer works, two major problems in the aforementioned literature can be highlighted:

1. From a theoretical point of view, the frontier between the probabilistic framework and the algorithmic one is not clearly defined.

Algorithmic complexity works with *one given string* at a time, not a *set of strings* with probabilities generated by a given stochastic process. Hence, no probabilistic assumption is needed when one uses algorithmic complexity. This is an advantage and a strength of this algorithmic approach.

The estimation of successful prediction rates seems to suggest that price motions follow a certain distribution law. Despite the use of compression tools, this methodological choice, at least to a certain extent, reinstates a probabilistic framework. The general and non-probabilistic context is then lost, which is regrettable.

2. Certainly, this is not the case with [36] and [37], which compared the complexity of stock markets according to their LZ indexes. However, the discretization technique used in these papers remains open to discussion.

Actually, financial returns are often expressed in real numbers⁶, while compression tools only deal with discrete data. So, regardless of the compression tool used, a discretization process, which transforms real-number series into discrete ones, is always necessary.

To do this, [34, 35], as well as [36, 37], proposed to transform financial returns into 3 signals: “positive”, “negative”, or “stable” returns. After discretization, financial time series become ternary strings.

Undoubtedly, this radical change leads to a significant loss of information from the original financial series.

As [35] remarked themselves, “*the main limitation of the VOM model is that it ignores the actual value of the expected returns. That is, the version of the algorithm used is based on a ternary alphabet, and is thus limited to the forecasting of either “positive”, “negative”, or “stable” returns, regardless of the different amounts of the expected returns [35, 49].*”

This technical detail weakens the contribution of their algorithmic approach to finance. Other approaches—those working with all the information in a return series—are more in the spirit of algorithmic complexity and could deliver more applications for financial data studies.

If we compare the works cited above with traditional EMH tests, the former should be considered as “*algorithmic run tests*”,⁷ in the sense that they verify the possibility of predicting the sign of price variations.

⁶The classical formula $r_t = \log(p_{t+1}) - \log(p_t)$, followed by most financial works to calculate return series, delivers real-number outcomes.

⁷Run tests, also known as Geary tests, are a non-parametric procedure that compares sequences of “up” and “down” in market prices with the outcome of a random walk; for an application in Finance, see for example, [39]

However, the introduction of algorithmic complexity in finance could have wider implications. For example, [30] have claimed that the notion of complexity could become a measure of financial risk - as an alternative to “value at risk” or “standard deviation” - which could have general implications for portfolio management.

Given these two points, establishing a general algorithmic framework for price motions seems possible. This could have important implications not only for the EMH but also for other financial theories dealing with risk and uncertainty.

To accomplish this, [40], [5] and [6] proposed a general method for estimating the algorithmic complexity of financial time series. They showed that some structures, undetectable by statistical tests, can be tracked using algorithmic tools. The following sections sketch the main contributions of these works.

4 Kolmogorov complexity as a general indicator of financial randomness

As in the works cited above, to apply compression tools to financial data requires a “real-to-discrete” transformation. To save all relevant information in financial data, [40] introduced a discretization method with an adjustable precision level. For example, if an analyst finds it important to keep at least 3 decimal places for each financial return, she can choose the number of alphabets used in the discretization method and only discard the *unnecessary* portion of the initial information⁸.

This particular discretization method makes the algorithmic framework as general as its probabilistic alternative in the sense that, on choosing the right precision level, compression tools can be used to detect all kinds of financial structures, patterns in consecutive signs, in volatilities as well as in higher order moments of financial returns.

To measure the complexity of an “n-length” string, s , [40] used the best lossless compression rate one can obtain from s , with the compression rate, CR , defined by equation (6):

$$(6) \quad CR = \frac{n - K_U(s)}{n}$$

where $K_U(s)$ denotes the Kolmogorov complexity of s as defined by equation (1). Hence, we have $\lim_{n \rightarrow \infty} CR = 0$, if s is random.

Thus, for a finite string, the best lossless CR is a good indicator of its complexity. The longer this string, the better the indicator.

Following this principle, [5] estimated the complexity of real-world financial returns. Data used in this study cover the period from 05/02/2001 to 09/02/2001 and are observed on a tick-by-tick frequency. As tick-by-tick data are only available for individual securities but not for indexes, the author chose 21 stocks from the Dow-Jones Industrial Average (hereafter DJ) and 30 from NASDAQ-100 as proxies for their corresponding markets (NYSE and NASDAQ). This choice is based on the number of transactions observed for each stock during the sample period. More precisely, a stock is selected if it belongs to DJ (or NASDAQ-100) *and* it registers more than 10000 price variations during the sample period. All these tick-by-tick data are extracted from the database “Trade and Quote II” (TAQ II) commercialized by Euronext-NYSE.

⁸That is to say the 4th, 5th, 6th ... figures after the point for each financial return.

To estimate the complexity of each stock, [5] followed a 3-step process:

1. Discretize the return series of each stock at an adequate precision level. In [5], at least four decimal places are preserved for each return.
2. Compress the discretized return series, denoted by s , with 3 compression tools: Huffman, Gzip and PAQ8o8⁹, and record the best lossless CR achieved by these algorithms.
3. If s is compressible, erase well-known structures¹⁰ in s with lossless arithmetic transformation, and repeat the whole process in order to see if s remains compressible. If it does, we may conclude that unknown structures are present in s .

The results of this process are reported in Tables 1 and 2. As is evident in column “CR”, tick-by-tick returns are compressible by algorithmic tools, which indicates the presence of patterns in these financial series. This result is congruent with econometric works that document stylized facts in finance (or put differently, patterns in moments of order 2 and beyond).

Thus, [5] used a progressive discretization process to erase volatility clusters in tick-by-tick series. Successive returns after this process are then tested by the same compression algorithms as in the preceding step. The best lossless CR are reported in the last columns of Tables 1 and 2.

Here one notes that even without volatility clusters, tick-by-tick returns remain compressible by lossless compression algorithms. This suggests the presence of unknown patterns in tick-by-tick data. Further work is necessary to identify the nature of these unknown structures and to ascertain whether the underlying structures can be of use in designing profitable trading rules. This is the main limitation of this algorithmic method: the patterns detected using general compression tools could be irrelevant to financial trading. This can be addressed in future work on designing compression tools which exploit profitable structures only.

As a general indicator of financial randomness, CR can also help to understand the relation between market microstructure and the speed at which relevant information is diffused:

One can see from the result tables that the mean of estimated CR(s) are quite close between the NYSE and the NASDAQ¹¹, both with and without volatility clusters. Actually, the NYSE is (mainly) an order-driven market and the NASDAQ is a price- driven one. IF this microstructural divergence affected the information diffusion speed on these two markets, one would have observed a significant difference in their CR. However, according to these algorithmic tests, it seems that being price-driven or order-driven has little impact on the quantity of information contained in price sequences, or on the randomness of price variations.

5 The algorithmic-probability approach to the market deviation from log-normal

When analysis is performed over short strings however, for example, when closing prices are encoded with one bit per day, lossless compression algorithms do not allow finer short period inspections because short strings are already too short to compress them further. In [9] we provide an alternative method (to compression) for approximating the algorithmic complexity of strings. Like long durations, weak complexities for short strings are tricky to evaluate. Paradoxically, the evaluation

⁹[40] includes a good description of the 3 compression tools.

¹⁰Such as the famous fat tail or volatility clustering phenomenon.

¹¹These similarities are validated by Kolmogorov-Smirnov tests.

Table 1: The best lossless CR obtained from the 21 most liquid stocks in DJ (05/02/2001-09/02/2001)

	Stock	Length	CR	CR (without volatility clustering)
1	AXP (Amer Express Inc)	14592	16,91%	4,06%
2	BMY(Bristol-Myers Squibb)	10752	10,03%	2,53%
3	C (Citigroup)	19712	20,48%	5,17%
4	DD (E.I. du Pont de Nem)	11008	18,01%	3,71%
5	DIS (Walt Disney Company)	14848	18,68%	5,74%
6	F (Ford Motor)	12032	17,25%	4,29%
7	GE (Gen Electric Co)	33024	24,62%	7,11%
8	HD (Home Depot)	20992	21,31%	4,77%
9	IBM (Intl Business Mach)	23552	19,54%	3,51%
10	JPM (JP Morgan Co)	15360	16,60%	4,08%
11	KO (Coca Cola)	11264	23,38%	3,81%
12	MCD (Mcdonalds)	13059	17,88%	5,27%
13	MO (Altria group)	16384	21,68%	6,49%
14	MRK (Merck Co Inc)	12544	14,25%	3,13%
15	PFE (Pfizer Inc)	25600	21,78%	6,90%
16	PG (Procter Gamble Co)	10240	11,62%	2,95%
17	T (ATT Inc)	17152	21,71%	5,33%
18	TXN (Texas Instruments)	34304	20,57%	5,39%
19	VZ (Verizon Commun)	12544	18,33%	4,41%
20	WMT (Wal-Mart)	18944	19,23%	5,34%
21	XOM(Exxon Mobil Co)	13056	16,50%	4,98%
	Mean		18,59%	4,71%
	Standard deviation		3,60%	1,25%

In the column labeled “Length”, the number of transactions registered for each security is displayed; in column “CR”, the best lossless compression rate obtained from the discretized string is shown, and in the last column, the best lossless CR obtained from return series without volatility clusters.

Table 2: The best lossless CR obtained from the 30 most liquid stocks in NASDAQ-100 (05/02/2001-09/02/2001)

	Stock	Length	CR	CR (without volatility clustering)
1	Apple Inc. (AAPL)	31488	11, 51%	3, 44%
2	Adobe Systems Incorporated (ADBE)	44032	18, 94%	5, 28%
3	Akamai Technologies, Inc. (AKAM)	20480	12, 84%	3, 65%
4	Altera Corporation (ALTR)	44544	14, 86%	4, 25%
5	Amazon.com, Inc. (AMZN)	47104	13, 41%	3, 42%
6	Amgen Inc. (AMGN)	47872	16, 04%	5, 00%
7	Applied Materials, Inc. (AMAT)	111872	18, 69%	5, 32%
8	Broadcom Corporation (BRCM)	260352	25, 47%	6, 18%
9	Check Point Software Technologies Ltd. (CHKP)	97280	21, 90%	4, 98%
10	Cisco Systems, Inc. (CSCO)	559872	23, 65%	4, 69%
11	Citrix Systems, Inc. (CTXS)	30720	12, 81%	4, 18%
12	Costco Wholesale Corporation (COST)	29184	13, 64%	3, 89%
13	Dell Inc. (DELL)	118272	16, 80%	4, 26%
14	EBay Inc. (EBAY)	38144	15, 14%	4, 58%
15	Flextronics International Ltd. (FLEX)	52224	16, 18%	4, 49%
16	Intel Corporation (INTC)	190720	18, 55%	4, 78%
17	Intuit, Inc. (INTU)	34304	17, 32%	4, 59%
18	KLA Tencor Corporation (KLAC)	62208	16, 73%	4, 33%
19	Lam Research Corporation (LRCX)	19968	8, 62%	2, 47%
20	Linear Technology Corporation (LLTC)	37376	14, 38%	3, 92%
21	Maxim Integrated Products (MXIM)	37376	13, 46%	3, 80%
22	Microsoft Corporation (MSFT)	180992	20, 29%	5, 13%
23	NetApp, Inc. (NTAP)	174848	24, 06%	6, 14%
24	Oracle Corporation (ORCL)	195840	20, 23%	4, 91%
25	Paychex, Inc. (PAYX)	22016	12, 55%	3, 59%
26	Qualcomm Incorporated (QCOM)	144128	23, 38%	5, 92%
27	Starbucks Corporation (SBUX)	21760	11, 51%	3, 34%
28	VeriSign, Inc. (VRSN)	72192	20, 75%	4, 96%
29	Xilinx, Inc. (XLNX)	72192	18, 83%	5, 03%
30	Yahoo! Inc. (YHOO)	95488	19, 69%	5, 00%
	Mean		17, 07%	4, 57%
	Standard deviation		4, 25%	0, 83%

methods require colossal calculations. For short strings, this novel method is stable enough and conforms to our idea of complexity; for long strings, it is guaranteed to converge to the algorithmic complexity due to the so-called invariance theorem (for details see this introductory book on the topic [41]).

On the other hand, we know that in turbulent price periods, what traders often end up doing is to leave aside the *any day is like any other-normal-day* “rule” derived from Brownian motion models, and fall back on their *intuition*, reproducing and falling into recurrent behavior leading to violent price changes (traditionally in the negative), which leads to their unwittingly following a model we believe to be better fitted to reality and hence to be preferred at all times, not just in uncertain times.

Using frequency distributions of daily closing price time series of several financial market indexes, we investigated and reported [6] on whether the bias away from an equiprobable sequence distribution found in the data may account for some of the deviation of financial markets from log-normal, and if so for how much of the said deviation and over what sequence lengths. We did so by comparing the distributions of binary sequences from actual time series of financial markets and series built up by purely algorithmic means.

The question is whether the market could be considered a rule-based system with an *algorithmic* component, despite its apparent randomness, so that the theory of algorithmic probability could account for the deviation of market data from log-normal. The basic assumption is not that alien to the mechanics of the stock market, where computers play an increasingly significant role in decision making (replacing human decisions) based on parameters both internal and external to the market but deterministic in nature and in agreement with an algorithmic view of the market. And even if undertaken by humans, the assumption is compatible with rational choice theory in that humans follow certain basic rules (whether rational or not) in their quest to maximize profit.

When observing a certain phenomenon, its outcome s can be seen as the result of a process P . One can then ask what the probability distribution of P generating s looks like. A probability distribution of a process is a description of the relative number of times each possible outcome occurs in a number of trials.

According to Levin’s distribution, in a world of computable processes, patterns which result from simple processes are relatively likely, while patterns that can only be produced by very complex processes are relatively unlikely. Algorithmic probability would predict, for example, that consecutive runs of the same magnitude, i.e. runs of pronounced falls and rises, and runs of alternative regular magnitudes have a greater probability than random-looking changes.

If one fails to discern the same simplicity in the market as is to be observed in certain other real world data sources [6], it is likely due to the dynamic of the stock market, where the exploitation of any regularity to make a profit results in the deletion of that regularity. Yet these regularities may drive the market and may be detected upon closer examination.

In a world of computable processes, Levin’s universal distribution establishes that patterns which result from simple processes (short programs) are likely, while patterns produced by complicated processes (long programs) are relatively unlikely, and that these patterns follow a power law distribution.

In economics the dynamics of the data differ from the dynamics of other empirical data in that patterns are quickly erased by economic activity itself, in the search for an economic equilibrium. But assuming an algorithmic hypothesis, that is, that there is a rule-based—as opposed to a purely stochastic—component in the market, one could apply the tools of the theory of algorithmic information, just as assuming random distributions led to the application of the traditional machinery

of probability theory.

Using a simulation of Turing machines to reproduce what the market would look like if it were all algorithmic in nature, what we found is that there are correlations with different degrees of significance [6] between the largest price changes in the empirical distribution of stock market price movements and the algorithmic empirical distribution (that we use as an approximation of Levin's universal distribution).

As expected, the algorithmic approach suggests that the tail of the distribution shows a stronger correlation among the elements themselves than the elements covered by the normal curve (accumulated at the centre of the Gaussian distribution from the Brownian motion model), which leads to the expected conclusion that no day in the market is like any other, but that certain days are more likely to be like certain others (e.g. a cascade of crashes).

Hence, departures from normality could be accounted for by the algorithmic component acting in the market, as is consonant with some empirical observations and common assumptions in economics, such as rule-based markets and agent modeling.

The algorithmic model in [6] predicts a greater incidence of simple signatures in agreement with the market if minor fluctuations derived from the Brownian motion model are regarded as still (stable) times under the algorithmic model.

The algorithmic model also predicts that random-looking signatures of higher volatility will occur more frequently if they are already occurring, a signature in unstable times where Brownian motion no longer holds. Our empirical samples show that given the weak to strong correlations, it is indeed the case that a small component of the price variations in financial markets may follow rules, and that the upshot may be the hidden rules and trends underlying and driving the market.

6 Concluding remarks

The most obvious feature of financial markets is the apparent randomness with which prices tend to fluctuate and which most standard models try to capture. Nevertheless, the very idea of chance in financial markets clashes with our intuitive sense of the processes regulating the market. Traders do not just follow hunches, but act in accordance with specific rules, and even when they do appear to act on intuition, their decisions are not random but instead follow from the best of their knowledge of the internal and external state of the market. For example, traders copy other traders, or take the same decisions that have previously worked, sometimes reacting against information and sometimes acting in accordance with it.

These deterministic processes could leave signatures (patterns) on financial data. To reveal their presence, algorithmic tools constitute a good alternative to stochastic models. In this paper, we have surveyed the principal applications of algorithmic (Kolmogorov) complexity to the problem of financial price motions and showed the relevance of the algorithmic framework to structure tracking in finance. Some empirical results are also provided to illustrate the power of the proposed estimators to take into account patterns in stock returns. Of course, the empirical tools reviewed above are only some of the uncountable possibilities opened up by the theory of algorithmic complexity. Just as one can always design new statistical tests for structure detection, the development of new algorithmic tools could enlarge the scope of patterns taken into account by the algorithmic framework and hence improve our comprehension of financial price dynamics.

References

- [1] S. Wolfram. *A New Kind of Science*. Wolfram Media, 2002.
- [2] L. Bachelier. *Théorie de la spéculation*. PhD thesis, Annales scientifiques de l'Ecole Normale Supérieure Série, 1900.
- [3] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Prob. Info. Trans.*, 1(1):3–11, 1965.
- [4] G. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, 1987. Cambridge Tracts in Theoretical Computer Science 1.
- [5] L. Ma. *Structures et hasard en finance: une approche par la complexité algorithmique de l'information*. Ph.d. thesis, University of Lille 1, 2010.
- [6] H. Zenil and J.-P. Delahaye. An algorithmic information theoretic approach to the behaviour of financial markets. *Journal of Economic Surveys*, 25(3):431–463, 2011.
- [7] R. Solomonoff. A formal theory of inductive inference: Parts 1 and 2. *J. Information and Control*, 7(1):224–254, 1964.
- [8] L. Levin. On a concrete method of assigning complexity measures. *Doklady Akademii nauk SSSR*, 18(3):727–731, 1977.
- [9] J.-P. Delahaye. Numerical evaluation of the complexity of short strings: A glance into the innermost structure of algorithmic randomness. *Applied Mathematics and Computation*, forthcoming.
- [10] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54, 1973.
- [11] J.-M. Courtault, Y. Kabanov, B. Bru, P. Crepel, I. Lebon, and A. Le Marchand. Louis bachelier: To the centenary of théorie de la spéculation. *Mathematical Finance*, 10(3):341–353, 2000.
- [12] A.N. Kolmogorov. Über die analitischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(3):415–458., 1931.
- [13] M. F. M. Osborne. Brownian motion in the stock market. *Operations Research*, 7:145–173, 1959.
- [14] R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50, 1982.
- [15] D.S. Bates. Post-'87 crash fears in the s&p 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238, 2000.
- [16] P.A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, vol. 6:pp. 41–49, 1965.

- [17] E. Fama. Efficient capital markets : A review of theory and empirical work. *Journal of Finance*, vol. 25:pp. 383–417, 1970.
- [18] P. A. Samuelson. Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, vol. 4:pp. 369–374, 1973.
- [19] S.F Leroy. Efficient capital markets and martingales. *Journal of Economic Literature*, vol. 27:pp. 1583–1621, 1989.
- [20] E. Fama. Efficient capital markets : Ii. *Journal of Finance*, 46(5):1575–1617, 1991.
- [21] E. Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283 – 306, 1998.
- [22] A. Timmermann and C. Granger. Efficient market hypothesis and forecasting. *International Journal of Forecasting*, 20(1):15–27, 2004.
- [23] D.M. Cutler, J.M . Poterba, and L.H . Summers. What moves stock prices? *The Journal of Portfolio Management*, 15(3):4–12, 1989.
- [24] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948.
- [25] J. Rissanen. Stochastic Complexity and Modeling. *The Annals of Statistics*, 14(3), 1986.
- [26] T.M. Cover and J.A. Thomas. *Elements of Information Theory*, chapter 15: Information Theory and the Stock Market, pages 459–481. John Wiley & Sons, Inc., 1991.
- [27] J. Chen. Information theory and market behavior. Finance 0503009, EconWPA, March 2005.
- [28] R. Mansilla. Algorithmic complexity in real financial markets. *Physica A*, 301(1-4):19, 2001.
- [29] A. Dionisio, R. Menezes, and D. Mendes. An econophysics approach to analyse uncertainty in financial markets: an application to the portuguese stock market. *European Physical Journal B*, 50(1-2):8, 2005.
- [30] A. Dionisio, R. Menezes, and D. Mendes. Entropy and uncertainty analysis in financial markets. *Time*, 1937(1937):1–9, 2007.
- [31] S.-H. Chen and C.-W. Tan. Measuring randomness by rissanen’s stochastic complexity: Applications to the financial data. Proceedings of the Sixth International Conference on Information Processing and Management of Uncert, Granada, Spain, 1996.
- [32] S.-H. Chen and C.-W. Tan. Estimating the complexity function of financial time series: An estimation based on predictive stochastic complexity. *Journal of Management and Economics*, 3(3), 1999.
- [33] S. Azhar, G.J. Badros, A. Glodjo, M.-Y. Kao, and J.H. Reif. Data compression techniques for stock market prediction. In *Data Compression Conference, 1994. DCC '94. Proceedings*, pages 72–82, mar 1994.

- [34] A. Shmilovici, Y. Kahiri, I. Ben-Gal, and S. Hauser. Using a stochastic complexity measure to check the efficient market hypothesis. *Computational Economics*, 22(2):273–284, October 2003.
- [35] A. Shmilovici, Y. Kahiri, I. Ben-Gal, and S. Hauser. Measuring the efficiency of the intraday forex market with a universal data compression algorithm. *Computational Economics*, 33:131–154, 2009.
- [36] S. Da Silva, R. Matsushita, and R. Giglio. The relative efficiency of stockmarkets. *Economics Bulletin*, 7(6):1–12, 2008.
- [37] R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, and S. Da Silva. Algorithmic complexity theory and the relative efficiency of financial markets. *EPL (Europhysics Letters)*, 84(4):48005, 2008.
- [38] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, 1976.
- [39] E. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.
- [40] O. Brandouy, J.-P. Delahaye, and L. Ma. Algorithmic complexity of financial time series. Technical report, University of Lille 1, 2009.
- [41] M. Li and P. Vitányi. *An Introduction of Kolmogorov Complexity and its Implications*. Springer-Verlag, 2 edition, 1997.