Lempel-Ziv Complexity and Crises of Cryptocurrency Market

Vladimir Soloviev¹ Serhiy Semerikov¹ Victoria Solovieva²,*

¹Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
²Kryvyi Rih Economic Institute of Kyiv National Economic University named after Vadym Hetman

*Corresponding author. Email: vikasolovieva2027@gmail.com

ABSTRACT
The informational (Kolmogorov) measure of complexity in accordance with the Lempel-Ziv algorithm (LZC) is calculated for the logarithmic returns of daily Bitcoin/$ values. The calculations were carried out for a moving window with a variation in its size (50-250 days) in increments of one day in the framework of the implemented coarse graining procedure. It is shown that in both mono-and multi-scaling versions, LZC is sensitive to noticeable fluctuations in the Bitcoin price that occur as a result of critical events in the cryptocurrency market. In equilibrium, stable state, having a relatively low value, LZC rapidly increases immediately before the crisis, which proves the dominance of the chaotic component of the time series. The classification and periodization of crisis phenomena in the cryptocurrency market for the period 2010-2020 has been carried out. The results demonstrate the possibility of using the LZC measure as an indicator-precursor of crisis phenomena in the cryptocurrency market.

Keywords: information theory, time series, returns, complex systems, Kolmogorov complexity, entropy, Lempel-Ziv complexity, cryptocurrency, Bitcoin, crisis

1. INTRODUCTION
Signals coming from the real world often have very complex dynamics or arise as a result of the self-organized dynamics of multidimensional systems. We can find many examples from various fields. In physiology, we can mention the processes in the electrical propagation of the heart, which provides electrocardiograms (ECG), or the dynamics underlying epileptic electroencephalograms (EEG) [1, 2]. Financial or social systems are also illustrative examples of how “complexity” emerges in these systems [3, 4, 5]. In this case, an important problem arises, consisting in the ability to extract relevant information from these complex time series [6, 7, 8]. A quantitative assessment of the complexity in the various types and nature of time-series signals was studied using various methods, especially those from (1) information, (2) dynamical system or (3) complex networks theory.

The idea of the first approach is to measure the spread of the statistical distribution underlying the data, identify changes in this distribution, analyze the spectral content of signals, etc. Among commonly used toolboxes information theory occupies a special place [9, 10, 11]. The second approach is well suited for signals of deterministic origin, usually non-linear. The tools used come from the world of chaos, such as fractal measurements, Lyapunov exponents, among many others [6, 12], or from the concept of complexity in the Kolmogorov sense (for example, Lempel-Ziv complexity) [13, 14, 15] Recently it was proposed analyze time series using the “complexity-entropy plane”, demonstrating that the joint use of two quantities gives more detailed information about the series than each measure separately [16].

Lempel-Ziv complexity (LZC) is a classical measure that, for ergodic sources, relates the concepts of complexity (in the Kolmogorov-Chaitin sense), and entropy rate [17, 18]. For an ergodic dynamical process, the amount of new information gained per unit of time (entropy rate) can be estimated by measuring the capacity of this source to generate new patterns. Because of the simplicity of the LZC method, the entropy rate can be estimated from a single discrete sequence of measurements with a low computational cost [19].

Methods of the complex networks theory [20] proceed from the possibility of representing the time series in the form of a graph and the subsequent analysis of the spectral and topological properties of the latter from the standpoint of complexity theory [21].

In the conditions of volatile, environmentally dependent financial markets, it is important to take measures of complexity that allow the early stages to identify critical phenomena that manifest themselves in the form of crises and that cause significant damage to the global and country economies [21-24].

In this paper, we show that the LZC measure can be just such a measure of complexity, which is an early precursor of crisis phenomena in the cryptocurrency market.


1.1. Analysis of Previous Studies

Historically, the first LZC measure system studies for financial time series were conducted by S. Da Silva et al. [23-27]. They considered the deviation of LZC from that value for a random time series as a measure of an actual market efficiency in absolute [23-26] or relative [27] terms. Using this approach authors were able to detect decreases in efficiency rates of the major stocks listed on the Sao Paulo Stock Exchange in the aftermath of the 2008 financial crisis [26].

In [28] authors 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.

In paper [29] was proposed a generic methodology to estimate the Kolmogorov complexity of financial returns. Examples are given with simulated data that illustrate the advantages of our algorithmic method: among others, some regularities that cannot be detected with statistical methods can be revealed by compression tools. Applying compression algorithms to daily returns of the Dow Jones Industrial Average, the authors concluded on an extremely high Kolmogorov complexity and by doing so, proposed another empirical observation supporting the impossibility to outperform the market.

In [30], the structural complexity of time series describing returns on New York’s and Warsaw’s stock exchanges was studied using two estimates of the Shannon entropy rate based on the Lempel-Ziv and Context Tree Weighting algorithms. Such structural complexity of the time series can be used as a measure of the internal (modellless) predictability of the main pricing processes, and testing the hypothesis of an effective market.

Somewhat surprisingly, the results of [31], in which the authors computed the Lempel-Ziv complexity from two composite stock indices, the Shanghai stock exchange composite index (SSE) and the Dow Jones industrial average (DJIA), for both low-frequency (daily) and high-frequency (minute-to-minute) stock index data. The calculation results indicate that the US market is basically fully random and consistent with efficient market hypothesis (EMH), irrespective of whether low- or high-frequency stock index data are used. The Chinese market is also largely consistent with the EMH when low-frequency data are used. However, a completely different picture emerges when the high-frequency stock index data are used.

H. Cao and Y. Li [32] present a novel method for measuring the complexity of a time series by unraveling a chaotic attractor modeled on complex networks. The complexity index, which can potentially be exploited for prediction, has a similar meaning to the Lempel-Ziv complexity, and is an appropriate measure of a series’ complexity. The proposed method is used to research the complexity of the world’s major capital markets. The almost absent sensitivity of the LZC to fluctuations in the time series indicates most likely errors in the calculation algorithm during the transformation of the time series.

The complexity–entropy causality plane is employed in order to explore disorder and complexity in the space of cryptocurrencies [33]. They are found to exist on distinct planar locations in the representation space, ranging from structured to stochastic-like behavior.

Crashes and critical events that took place on this market as well as the reasons that led to them, did not go unheeded. A review of the main articles and the results of the study of the crisis conditions of the crypto market was made by us in recent works [22, 34]. A brief analysis of the problem indicates that to date, the Lempel-Ziv informational measure of the complexity has not been used to study the stability and behavior of the cryptocurrency market in a crisis. The aim of this work is an attempt to fill this gap.

1.2. Our Contribution

In this paper, for the first time, we use the Lempel-Ziv complexity measure to study the cryptocurrency market. Using the example of the most capitalized cryptocurrency - Bitcoin - we demonstrate the ability to identify dynamics of varying complexity. Particularly relevant is the identification of the characteristic behavior of Bitcoin during the crisis phases of market behavior. By observing the dynamics of the Lempel-Ziv measure, precursors of crisis phenomena can be constructed.

1.3. Paper Structure

The rest of the paper is organized as follows. Section 2 presents the basic concepts of informational (Kolmogorov) complexity, the implementation of the Lempel-Ziv algorithm for time series, describes the tools for calculating Lempel-Ziv complexity. Section 3 presents a description of the database of calculations, the classification and periodization of crisis phenomena in the cryptocurrency market, and the actual results of calculations of mono-and multi-scaling window measures LZC. Finally, Section 4 concludes the paper and presents direction for future research.

2. LEMPEL-ZIV COMPLEXITY

Based on the different nature of the methods laid down in the basis of the formation of the measure of complexity, they pay particular demands to the time series that serve the input. For example, information requires stationarity of input data. At the same time, they have different sensitivity to such characteristics as determinism, stochasticity, causality and correlation.
2.1. The Concept of Kolmogorov Complexity

Let us begin with the well-known degree of complexity proposed by A. Kolmogorov [35]. The concept of Kolmogorov complexity (or, as they say, algorithmic entropy) emerged in the 1960s at the intersection of algorithm theory, information theory, and probability theory. A. Kolmogorov’s idea was to measure the amount of information contained in individual finite objects (rather than random variables, as in the Shannon theory of information). It turned out to be possible (though only to a limited extent). A. Kolmogorov proposed to measure the amount of information in finite objects using algorithm theory, defining the complexity of an object as the minimum length of the program that generates that object. This definition is based on the algorithmic information theory as well as algorithmic probability theory: an object is considered random if its complexity is close to maximum.

What is the Kolmogorov complexity and how to measure it? In practice, we often encounter programs that compress files (to save space in the archive). The most common are called zip, gzip, compress, rar, arj and others. Applying such program to some file (with text, data, program), we get its compressed version (which is usually shorter than the original file). After that, you can restore the original file using the paired program “decompressor”. Therefore, in the first approximation, the Kolmogorov complexity of a file can be described as the length of its compressed version. Thus, a file that has a regular structure and is well compressed has a small Kolmogorov complexity (compared to its length). On the contrary, a badly compressed file has a complexity close to its length.

Suppose we have a fixed method of description (decompressor) D. For this word x we consider all its descriptions, i.e. all words y for which D(y) = x. The length of the shortest of them is called the Kolmogorov complexity of the word x in this way of description D:

\[ K_{S_y}(x) = \min \{ l(y) \mid D(y) = x \}, \]

where l(y) denotes the length of the word. The index D emphasizes that the definition depends on the choice of method D. It can be shown that there are optimal methods of description. The better the description method, the shorter it is. Therefore, it is natural to make the following definition: the method D is no worse than the method D’ if \( K_{S_y}(x) \leq K_{S_{y’}}(x) + c \) for some c and all x.

Thus, according to Kolmogorov, the complexity of an object (for example, text is a sequence of characters) is the length of the minimum program that outputs the text, and entropy is the complexity that is divided by the length of the text. Unfortunately, this definition is purely speculative. There is no reliable way to uniquely identify this program. But there are algorithms that are actually just trying to calculate the Kolmogorov complexity of text and entropy.

2.1. Kolmogorov Complexity Estimation According to the Lempel-Ziv Algorithm

A universal (in the sense of applicability to different language systems) measure of complexity of the finite character sequence was suggested by Lempel and Ziv [28]. As part of their approach, the complexity of a sequence is estimated by the number of steps in the process that gives rise to it. Acceptable (editorial) operations are:

a) character generation (required at least for the synthesis of alphabet elements) and
b) copying the "finished" fragment from the prehistory (i.e. from the already synthesized part of the text).

Let be \( \Sigma \) a complete alphabet, \( S \) - text (a sequence of characters) composed of elements \( \Sigma \); \( S[i] \) - i-th text symbol; \( S[i : j] \) - a snippet of text from the i-th to j-th character inclusive (\( i < j \)); \( N = |S| \) - length of text S. Then the sequence synthesis scheme can be represented as a concatenation

\[ H(S) = S[1 : i_1]S[i_1 + 1 : i_2]...S[i_{m-1} + 1 : i_m]S[i_m + 1 : N], \]

where \( S[i_{m-1} + 1 : i_m] \) is the fragment S generated at the k-th step, and m = \( m_{\Sigma}(S) \) is the number of process steps. Of all the schemes of generation is chosen the minimum number of steps. Thus, the Lempel-Ziv complexity of the sequence S is

\[ c_{LZ}(S) = \min \{ m_{\Sigma}(S) \}. \]

The minimum number of steps is provided by the choice to copy at each step the longest prototype from the prehistory. If you mark by the position number \( j(k) \) from which the copying begins in step k, the length of the copy fragment

\[ l_{j(k)} = i_k - i_{k-1} - 1 = \max \{ l_j : S[i_{k-1} + 1 : i_k + l_j] = S[j : j + l_j - 1] \}, \]

and the k-th component of these complex decomposition can be written in the form

\[ S[i_{k-1} + 1 : i_k] = \begin{cases} S[j(k) : j(k) + l_{j(k)} - 1] & \text{if } j(k) \neq 0; \\ [S[i_{k-1} + 1]] & \text{if } j(k) = 0. \end{cases} \]

The case \( j(k) = 0 \) corresponds to a situation where a symbol is in the position \( i_{k-1} + 1 \) that has not previously been encountered. In doing so, we use a character generation operation.

Complex text analysis can be performed in two regime - segmentation and fragmentation. The first regime is discussed above. It gives an integrated view of the structure of the sequence as a whole and reduces it to disjoint but interconnected segments (without spaces). The other regime is to search for individual fragments
characterized by an abnormally low complexity, that is, a sufficiently high degree of structure. Such fragments are detected by calculating local complexity within variable length windows that slide along a sequence. Curves of change of local complexity along a sequence are called complex profiles. A set of profiles for different window sizes reveals the boundaries of anomalous fragments and their relationship.

We will use the Lempel-Ziv complexity for the time series, which is, for example, the daily values of the cryptocurrency price \( S(t) \). To investigate the dynamics of LZC and compare it with cryptocurrency price, we will find this measure of complexity for a fixed length (window) contract. To do this, we calculate the logarithmic returns \( \text{ret}(t) \) and turn them into a sequence of bits. The returns over some time scale \( \Delta t \) is defined as the forward changes in the logarithm of \( S(t) \):

\[
G(t) = \frac{\ln(S(t+\Delta t)) - \ln(S(t))}{\Delta t} \tag{1}
\]

Since different cryptocurrencies have different levels of variability (standard deviations), we will determine standardized returns

\[
g(t) = \frac{G(t) - \langle G \rangle}{\sigma} \tag{2}
\]

where \( \sigma = \sqrt{\langle G^2 \rangle - \langle G \rangle^2} \) is the standard deviation \( G \), and \( \langle \ldots \rangle \) denotes the average over the time period under study.

You can specify the number of states that are differentiated (calculus system). Yes, for two different states we have 0, 1, for three - 0, 1, 2, etc. In the case of three states, unlike the binary coding system, a certain threshold \( b \) is set and the states \( \text{ret} \) are coded as follows [24-26]:

\[
\text{ret} = \begin{cases} 0, & \text{if } \text{ret} < -b \\ 1, & \text{if } -b \leq \text{ret} \leq b \\ 2, & \text{if } \text{ret} > b
\end{cases} \tag{3}
\]

The algorithm performs two operations: (1) adds a new bit to an already existing sequence; (2) copies the already formed sequence. Algorithmic complexity is the number of such operations required to form a given sequence.

For a random sequence of lengths \( n \), the algorithmic complexity is calculated by expression

\[
\text{LZC}_r = n / \log(n)
\]

Then the relative algorithmic complexity is the ratio of the obtained complexity to the complexity of the random sequence \( \text{LZC} = \text{LZC}_r / \text{LZC}_r \).

Obviously, the classic indicators of algorithmic complexity are unacceptable and lead to erroneous conclusions. To overcome such difficulties, multiscale methods are used.

The idea of this group of methods includes two consecutive procedures: 1) coarse graining (“granulation”) of the initial time series – the averaging of data on non-intersecting segments, the size of which (the window of averaging) increased by one when switching to the next largest scale; 2) computing at each of the scales a definite (still mono scale) complexity indicator. The process of “rough splitting” consists in the averaging of series sequences in a series of non-intersecting windows, and the size of which – increases in the transition from scale to scale [34]. Each element of the “granular” time series is in accordance with the expression:

\[
y_j^{(r)} = 1 / \tau \sum_{t=(j-1)2^r +1}^{j2^r} x_t, \quad 1 \leq j \leq N / \tau, \tag{4}
\]

where \( \tau \) characterizes the scale factor. The length of each “granular” row depends on the size of the window and is even \( N / \tau \). For a scale equal to 1, the “granular” series is exactly identical to the original one.

The coarse graining procedure for scales 2 and 3 is shown in Figure 1.

![Figure 1: Coarse-graining procedure diagram.](image)

To find the LZC measure of the time series, sliding time windows were considered; the index for every window was calculated and then the average was obtained.

### 3. DATA AND ANALYSIS

At the moment, there are various research works on what crises are and how to classify such interruptions in the market of cryptocurrencies. Taking into account the experience of previous researchers, we have created our classification of such leaps and falls [22], relying on Bitcoin time series during the entire period (16.07.2010 – 31.12.2019) of verifiable fixed daily values of the Bitcoin price [https://finance.yahoo.com/cryptocurrencies](https://finance.yahoo.com/cryptocurrencies). The author's version of the cryptocurrency crisis periodization is shown in Table.
Table List of Bitcoin major crises since June 2011

<table>
<thead>
<tr>
<th>№</th>
<th>Name</th>
<th>Days in correction</th>
<th>Bitcoin High Price, $</th>
<th>Bitcoin Low Price, $</th>
<th>Decline, %</th>
<th>Decline, $</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>07.06.2011-10.06.2011</td>
<td>4</td>
<td>29.60</td>
<td>14.65</td>
<td>50</td>
<td>15.05</td>
</tr>
<tr>
<td>2</td>
<td>15.01.2012-16.02.2012</td>
<td>33</td>
<td>7.00</td>
<td>4.27</td>
<td>39</td>
<td>2.73</td>
</tr>
<tr>
<td>3</td>
<td>15.08.2012-18.08.2012</td>
<td>4</td>
<td>13.50</td>
<td>8.00</td>
<td>40</td>
<td>5.50</td>
</tr>
<tr>
<td>4</td>
<td>08.04.2013-15.04.2013</td>
<td>8</td>
<td>230.00</td>
<td>68.36</td>
<td>70</td>
<td>161.64</td>
</tr>
<tr>
<td>6</td>
<td>05.02.2014-25.02.2014</td>
<td>21</td>
<td>904.52</td>
<td>135.77</td>
<td>85</td>
<td>768.75</td>
</tr>
<tr>
<td>7</td>
<td>12.11.2014-14.01.2015</td>
<td>64</td>
<td>432.02</td>
<td>164.91</td>
<td>62</td>
<td>267.11</td>
</tr>
<tr>
<td>8</td>
<td>26.01.2015-31.01.2015</td>
<td>5</td>
<td>269.18</td>
<td>218.51</td>
<td>20</td>
<td>50.67</td>
</tr>
<tr>
<td>9</td>
<td>09.11.2015-11.11.2015</td>
<td>3</td>
<td>380.22</td>
<td>304.70</td>
<td>20</td>
<td>75.52</td>
</tr>
<tr>
<td>10</td>
<td>18.06.2016-21.06.2016</td>
<td>4</td>
<td>761.03</td>
<td>590.55</td>
<td>22</td>
<td>170.48</td>
</tr>
<tr>
<td>11</td>
<td>04.01.2017-11.01.2017</td>
<td>8</td>
<td>1135.41</td>
<td>785.42</td>
<td>30</td>
<td>349.99</td>
</tr>
<tr>
<td>12</td>
<td>03.03.2017-24.03.2017</td>
<td>22</td>
<td>1283.30</td>
<td>939.70</td>
<td>27</td>
<td>343.60</td>
</tr>
<tr>
<td>13</td>
<td>10.06.2017-15.07.2017</td>
<td>36</td>
<td>2973.44</td>
<td>1914.08</td>
<td>36</td>
<td>1059.36</td>
</tr>
<tr>
<td>14</td>
<td>31.08.2017-13.09.2017</td>
<td>13</td>
<td>4921.85</td>
<td>3243.08</td>
<td>34</td>
<td>1678.77</td>
</tr>
<tr>
<td>17</td>
<td>09.07.2019-16.07.2019</td>
<td>7</td>
<td>12567.02</td>
<td>9423.44</td>
<td>25</td>
<td>3143.58</td>
</tr>
</tbody>
</table>

Figure 2 shows the dynamics of the daily values of the MTC price for the study period, and the inset shows the 3rd and 4th of the crises presented in the Table.

Figure 2 The dynamics of the daily values of the BTC price. The inset shows the 3rd and 4th of the crises presented in the Table.

Figure 3 shows the dynamics of standardized logarithmic returns, and the limits of the normal distribution (±3σ) are indicated by dashed lines.

Figure 3 The dynamics of standardized logarithmic returns of the BTC price.

Obviously, the crisis on the cryptocurrency market responds to noticeable fluctuations in standardized returns. Therefore, it is logical to choose σ as the value for the threshold value b in (3).

Figure 4 shows the dependence of the LZC on the scale. The absence of LZC fluctuations at scales exceeding 40
allows us to confine ourselves to this magnitude of the scale when calculating the multiscale measure.

Figure 4 Scale-dependent measure LZC

For a scale interval of 1–40, the dynamics of the multiscale measure LZC is shown in Figure 5.

Figure 5 Time dependence of a scale-dependent measure LZC

Calculations of measures of complexity were carried out both for the entire time series, and for a fragment of the time series localizing the crisis. In the latter case, fragments of time series of the same length with fixed points of the onset of crisis events were selected and the results of calculations of complexity measures were compared to verify the universality of the indicators. Comparing the dynamics of the BTC time series and the corresponding LZC measures, we can judge the characteristic changes in the dynamics of the behavior of complexity with changes in the cryptocurrency. If this or that measure of complexity behaves in a definite way for all periods of crises, for example, decreases or increases during the pre-crisis period, then it can serve as an indicator or precursor of such a crisis phenomenon. Figure 6 presents the results of calculations of mono- ($LZC_m$) and multiscaling ($LZC_{m40}$) LZC measures. Arrows indicate crises documented by the above Table. The calculations were performed for a sliding window of 100 days in increments of 1 day.

Figure 6 Comparative dynamics of BTC price fluctuations and mono- and multi-scaling LZC measures. The arrows indicate the periods of the onset of crises in accordance with their numbers in the Table.

The data in Fig. 6 indicate that the LZC measure, both in the case of mono-scale ($m$) and averaged over the scales from 1 to 40 ($m_{40}$) for all eighteen detected crises (see Table) in the immediate vicinity of the crisis point, is noticeably reduced. As the results of calculations showed, the choice of the size of a moving window is important: in the case of large windows, points of crises of different times can fall into the window, distorting the influence of each of the crises. When choosing small windows, the results fluctuate greatly, which makes it difficult to determine the actual point of the crisis. The used window size of a length of 100 days turned out to be optimal from the point of view of separation of crises and fixing the LZC measure as an indicator. Since the LZC measure begins to decrease even before the actual crisis point, it can be called an indicator-predictor of crisis phenomena in the cryptocurrency market.

4. CONCLUSION

Complexity theory and its use for the classification of complex dynamic systems, the study of critical phenomena, the prevention and prediction of crisis phenomena are of significant scientific and applied interest. Information measures of complexity due to their initial validity and transparency, ease of implementation and interpretation of the results occupy a prominent place among the tools for the quantitative analysis of complex systems. The Lempel-Ziv complexity measure was previously used to quantify the complexity of financial assets. In this paper, it was first used: (a) for the cryptocurrency market, (b) in the mono and multiscale versions, and (c) for the
construction of an indicator-precursor of crisis phenomena in the cryptocurrency market.

Obviously, in the future we should expect improved forecasts on the way of combining the Lempel-Ziv measure with other constructs, for example, various entropy, recurrence, irreversibility measures, and others that comprehensively reflect many facets of the complexity phenomenon.

REFERENCES


