scientific reports

OPEN

Chapman-Kolmogorov test for estimating memory length of two coupled processes

H. Motahari[1], M. Ghanbarzadeh Noudehi[2], T. Jamali[1] & G. R. Jafari[1,3,4]✉

Real-world processes often display a prolonged memory, which extends beyond the single-step dependency characteristic of Markov processes. In addition, the current state of an empirical process is often not only influenced by its own past but also by the past states of other dependent processes. This study introduces the generalized version of the Chapman-Kolmogorov equation (CKE) to estimate the memory size in such scenarios. To assess the applicability of our approach, we generate coupled time series with predetermined memory lengths using the autoregressive model. The results show a high degree of accuracy in measuring memory lengths. Subsequently, we employ the generalized CKE to analyze cryptocurrency data as a real-world case study. Our results indicate that the past dynamics of cryptocurrencies significantly impact their current states, thereby highlighting interdependencies among them. The method proposed in this study can be also utilized in forecasting coupled time series.

The conventional method for estimating the current state of a process involves studying its past states since the present behavior of a process is not entirely independent of its past. To quantify the relationship between the current and preceding states, we employ a measure referred to as the Markov length[1–4]. One of the common approaches to computing the Markov length is the Chapman-Kolmogorov equation (CKE)[5,6]. In a study on rough surfaces researchers introduced three methods to calculate Markov length: the CK test, the Wilcoxon test, and a method based on the Langevin equation[7]. Among these, the CK method stands out as one of the most widely used due to its generalization capabilities, unlike the more specific applications of the other two. Given the growing significance of interdependent processes across various fields in recent decades, applying the generalized CK method to both synthetic and empirical data provides valuable insights into underlying dynamics. This approach has been applied to diverse fields such as turbulence[8,9], rough surface[1,7,10], financial markets[11–14], biology[15], signal processing[16], and earthquakes[17].

In the real world, finding a process in isolation from other processes is challenging. As a result, a process can be not only influenced by its own past but also by the past of other processes[18–20]. For example, historical oil prices directly impact specific commodities' costs and even some countries' GDPs[21]. Another instance is the interdependency between currency exchanges and the values of various cryptocurrencies[22,23]. In order to analyze the memory length of such examples, researchers have applied the CK test method across various domains. This approach is utilized in a study to analyze the roughness of surface height profiles, providing insights into the stochastic behavior of rough surfaces[7]. In another study researchers employed the CK test to evaluate heart rate beat-to-beat fluctuations in non-stationary data, effectively distinguishing healthy subjects from those with congestive heart failure[24]. This raises the question of how to measure the mutual memory effects between such dependent processes. Gao et al. worked on assessing the efficiency of stock markets in the US and China, concluding that while the US market largely conforms to the Efficient Market Hypothesis (EMH), the Chinese market shows deviations, especially in high-frequency data[25]. The method they used was Lempel-Ziv complexity and permutation entropy to quantify unpredictability and they also aimed to understand how temporal patterns evolve in financial systems, capturing deviations from randomness by analyzing complexity. Their study highlighted the characterization of long-range correlations in the Chinese stock market using the Hurst parameter, which quantifies memory effects in time series data. The Hurst exponent is commonly employed to measure persistence or anti-persistence in complex systems, offering insight into how past states influence future dynamics. This finding indicates that price movements in the Chinese market are not entirely random but are influenced by past events over extended time scales. While our approach does not explicitly

[1]Department of Physics, Shahid Beheshti University, Evin, Tehran 1983969411, Iran. [2]Department of Physics, Institute for Advanced Studies in Basic Sciences, 45195-1159 Zanjan, Iran. [3]Institute of Information Technology and Data Science, Irkutsk National Research Technical University, 83, Lermontova St., 664074 Irkutsk, Russia. [4]Centre for Communications Technology, London Metropolitan University, London N7 8DB, UK. ✉email: gjafari@gmail.com

utilize the Hurst parameter, it inherently accounts for memory effects through the structure of the dynamical equations governing the system. In this paper we intend to measure memory effects and inter-dependencies between cryptocurrency processes, focusing on how past states influence future dynamics. Our focus is on quantifying short-term memory effects using the generalized Chapman-Kolmogorov equation. Similarly, our approach highlight the importance of high-frequency data in revealing market inefficiencies.

This study aims to address this question by proposing a generalization of the Chapman-Kolmogorov equation to two coupled processes. In order to evaluate the effectiveness of the generalized CKE, we initially employ a toy model based on the autoregressive model, specifically designed for generating two coupled processes with predefined Markov lengths. By applying our method on toy model with the stochastic Langevin signals, we indeed investigated that the method can distinguish the Markov length for this stochastic data to ensure having the satisfying results for our toy model to be utilized for the real data as well. Then, we apply our approach to the cryptocurrency market data. We gather data from three different time frame samples and proceed to measure the Markov lengths for each time frame. Finally, we compare the obtained results for each respective time frame. It is worth mentioning that the Markov length can be affected by many different sources but regardless of whether what is the reason of constructing such a memory, this method can detect the length of the memory.

This paper is organized as follows. Section Method presents the methodology, which is divided into two subsections: (A) a review of the CKE and its application for calculating the Markov length of a single process, and (B) the introduction of the generalized version of CKE for analyzing two coupled processes and calculating their respective Markov lengths. In Sect. Results, we evaluate our proposed method through two experiments. Firstly, we test our approach using the synthetic data of two coupled autoregressive processes with known Markov lengths. Subsequently, we apply our method to empirical data from the financial cryptocurrency market. In Sect. Conclusion, we present our final arguments and the key findings of our study. Additionally, Appendix A provides details of the error estimation method utilized in finding Markov length.

Method

Chapman-Kolmogorov equation

Markov processes, renowned for their mathematical simplicity, are frequently employed to model dynamic systems[26]. This simplicity results from the assumption of memoryless transitions between the two successive states of the system. However, many real-world systems exhibit memory effects[27]. In such cases, a characteristic length scale known as Markov length may exist that describes the time scale up to which the influence of memory is significant while beyond which the memory effect becomes negligible[28,29]. Hence, for time scales larger than Markov length, a non-Markov process can be approximated by a Markov process and consequently should satisfy the prerequisites of Markov processes[30,31]. One prerequisite is the Chapman-Kolmogorov equation that relates the probability of transitioning from one state to another with intermediate transitions between those states[32]. CKE has proved itself as a suitable method for estimating Markov length[8,33].

Before proceeding further, it is important to clarify some fundamental concepts regarding stochastic processes. Consider a collection of identical systems, each producing a sample process denoted as $X(t)$, with "t" representing a parameter typically associated with time. A stochastic process is defined as the ensemble comprising all possible sample processes and is represented by $\{X(t)\}$. Each sample process is a specific instance or realization of the stochastic process. Hence, it is important to distinguish between a sample process and a stochastic process. At any given time t, each sample process has a unique value, while the stochastic process $\{X(t)\}$ is a set of random variables.

The probability that the stochastic process has values within the interval $(x, x+dx)$ at time t is expressed as $p(x, t)dx$, where $p(x, t)$ is referred to as the probability density function (PDF). It is defined as $\langle \delta(x - X(t)) \rangle$, where $\langle \cdots \rangle$ shows the ensemble average across all realizations of $\{X(t)\}$ and δ is the Dirac delta function[5]. In a similar manner, the conditional probability density $p(x_2, t_2|x_1, t_1)$, representing the likelihood that the stochastic process $\{X(t)\}$ traverses the interval $(x_2, x_2 + dx_2)$ at time t_2 after starting from x_1 at time t_1, is defined as $\langle \delta(x_2 - X(t_2)) \rangle \big|_{X(t_1)=x_1}$. It is crucial to note that although t and x are independent variables in the definition of probability densities, when dealing with a sample process, which is the practical scenario, the value of the process depends on time, and they are not independent variables. All subsequent equations introduced or derived pertain to stochastic processes, which are then applied to analyze sample processes.

Here, we explain CKE and its application to calculate Markov length. In probability theory, the two-point joint probability density, denoted as $p(x_3, t_3; x_1, t_1)$, can be obtained by integrating the three-point joint probability density $p(x_3, t_3; x_2, t_2; x_1, t_1)$ over the variable x_2[34]. Respecting the definition of conditional probability, which states that $p(x_i, t_i|x_j, t_j) = p(x_i, t_i; x_j, t_j)/p(x_j, t_j)$, where $t_i > t_j$, it is straightforward to prove that[5]

$$p(x_3, t_3|x_1, t_1) = \int_{-\infty}^{\infty} p(x_3, t_3|x_2, t_2) p(x_2, t_2|x_1, t_1)\, dx_2. \tag{1}$$

The above equation is known as the Chapman-Kolmogorov equation[35–37]. It is worth noting that satisfying Eq. (1) is a necessary condition for a stochastic process to be classified as a Markov process. In other words, if a process fails to satisfy CKE, it cannot be considered a Markov process[32,38]. Exploiting this property, we can define Markov length, which represents a characteristic of systems with memory.

Consider an arbitrary stationary stochastic process $\{X(t)\}$ along with points (t_i, x_i), (t_k, x_k), and (t_j, x_j), where $t_i > t_k > t_j$. To assess the validity of CKE for these points, we introduce a parameter "S" that quantifies the difference between the two sides of Eq. (1). This parameter is defined as follows:

$$S(x_i, x_j; \tau) = p(x_i|x_j; \tau) - \int_{-\infty}^{\infty} p(x_i|x_k; \tau')p(x_k|x_j; \tau'')\, dx_k, \tag{2}$$

where $\tau = t_i - t_j$, $\tau' = t_i - t_k$ and $\tau'' = t_k - t_j$. If the stochastic process satisfies CKE, *S* would be zero; otherwise, *S* would be non-zero. To obtain a measure that captures the contribution of all possible values of x_i and x_j, we define a memory function as

$$\mathrm{M}(\tau) = \int_{-\infty}^{\infty} |S(x_i, x_j; \tau)|\, dx_i dx_j \tag{3}$$

By definition, the memory function M is a non-negative function that depends only on the temporal distance τ. This means that the memory function does not take into account any specific values of the stochastic process but rather focuses on the temporal relationship between the two points. This function exhibits different characteristics depending on the magnitude of τ. For sufficiently small temporal distances, it significantly deviates from zero, whereas for sufficiently large temporal distances, it converges toward zero. The specific temporal distance before and after which this transition occurs is called the Markov length. To calculate the Markov length, we compute M for various values of τ, and observe the specific values τ^* at which an abrupt change from non-zero to nearly zero values happens in M. The Markov length is then determined as $l_m := \tau^*$.

It is noteworthy to emphasize that finding the Markov length for empirical processes is prone to error due to the uncertainties in calculating the probabilities presented in Eq. (2). These uncertainties arise as a consequence of the limited length of real data. The error computation is significant in that it reflects the precision of measurements. Consequently, determining the cutoff point becomes essential as it enables us to specify the Markov length within the bounds of measurement accuracy. For detailed information regarding the computation of errors, please refer to Appendix A.

Generalized Chapman-Kolmogorov equation

In the previous section, we showed how to estimate Markov length for a stochastic process using CKE. Here, we generalize CKE to determine the *mutual Markov length* of two dependent processes.

Consider two dependent stochastic processes, denoted as $\{X(t)\}$ and $\{Y(t)\}$, which possess the ability to memorize each other's past states. In other words, the current state of each process is influenced not only by its own past behavior but also by the historical dynamics of the other process. For instance, Fig. 1 shows the schematic configuration of these processes with distinct memory characteristics. The arrows, together with the conditional probability density function, illustrate the dependence of each process at time *t* on its preceding steps, as well as the preceding steps of the other process. In such cases, it is possible to represent both processes as a unified vector stochastic process denoted as $\{\mathrm{R}(t)\}$, where $\mathrm{R}(t) := X(t)\mathrm{i} + Y(t)\mathrm{j}$, with and representing the unit vectors along the x-axis and y-axis, respectively. The Chapman-Kolmogorov equation for $\{\mathrm{R}(t)\}$ can then be expressed as follows[7]:

$$p(\mathrm{r}_i, t_i|\mathrm{r}_j, t_j) = \int_{-\infty}^{\infty} p(\mathrm{r}_i, t_i|\mathrm{r}_k, t_k)p(\mathrm{r}_k, t_k|\mathrm{r}_j, t_j)\, d\mathrm{r}_k, \tag{4}$$

where $d\mathrm{r}_k = dx_k dy_k$. Similar to the case of a single stochastic process, we can establish the definition of *S* for a stationary $\{\mathrm{R}(t)\}$ as follows:

$$\begin{aligned} &S(\mathrm{r}_i, \mathrm{r}_j; \tau) := \\ &p(\mathrm{r}_i|\mathrm{r}_j; \tau) - \int_{-\infty}^{\infty} p(\mathrm{r}_i|\mathrm{r}_k; \tau')p(\mathrm{r}_k|\mathrm{r}_j; \tau'')\, d\mathrm{r}_k. \end{aligned} \tag{5}$$

It should be noted that in the case of multiple processes, each process may possess its own Markov length and a mutual Markov length with respect to other processes. Consequently, for the two processes *X*(*t*) and *Y*(*t*), we can identify four distinct Markov lengths: $l_m^{x\leftarrow x}$, $l_m^{x\leftarrow y}$, $l_m^{y\leftarrow x}$, and $l_m^{y\leftarrow y}$. The arrowhead in the superscript indicates the direction of influence. For instance, the mutual Markov length $l_m^{y\leftarrow x}$ signifies that the past states of *X*(*t*) influence the present state of *Y*(*t*). To simplify notation, the superscript arrows will be omitted hereinafter, but the order should be kept in mind. The four Markov lengths can be determined through the following relations:

$$\begin{aligned} S(x_i, x_j; \tau) &= \int_{-\infty}^{\infty} S(\mathrm{r}_i, \mathrm{r}_j; \tau)\, dy_i dy_j, \\ S(x_i, y_j; \tau) &= \int_{-\infty}^{\infty} S(\mathrm{r}_i, \mathrm{r}_j; \tau)\, dy_i dx_j, \\ S(y_i, x_j; \tau) &= \int_{-\infty}^{\infty} S(\mathrm{r}_i, \mathrm{r}_j; \tau)\, dx_i dy_j, \\ S(y_i, y_j; \tau) &= \int_{-\infty}^{\infty} S(\mathrm{r}_i, \mathrm{r}_j; \tau)\, dx_i dx_j. \end{aligned} \tag{6}$$

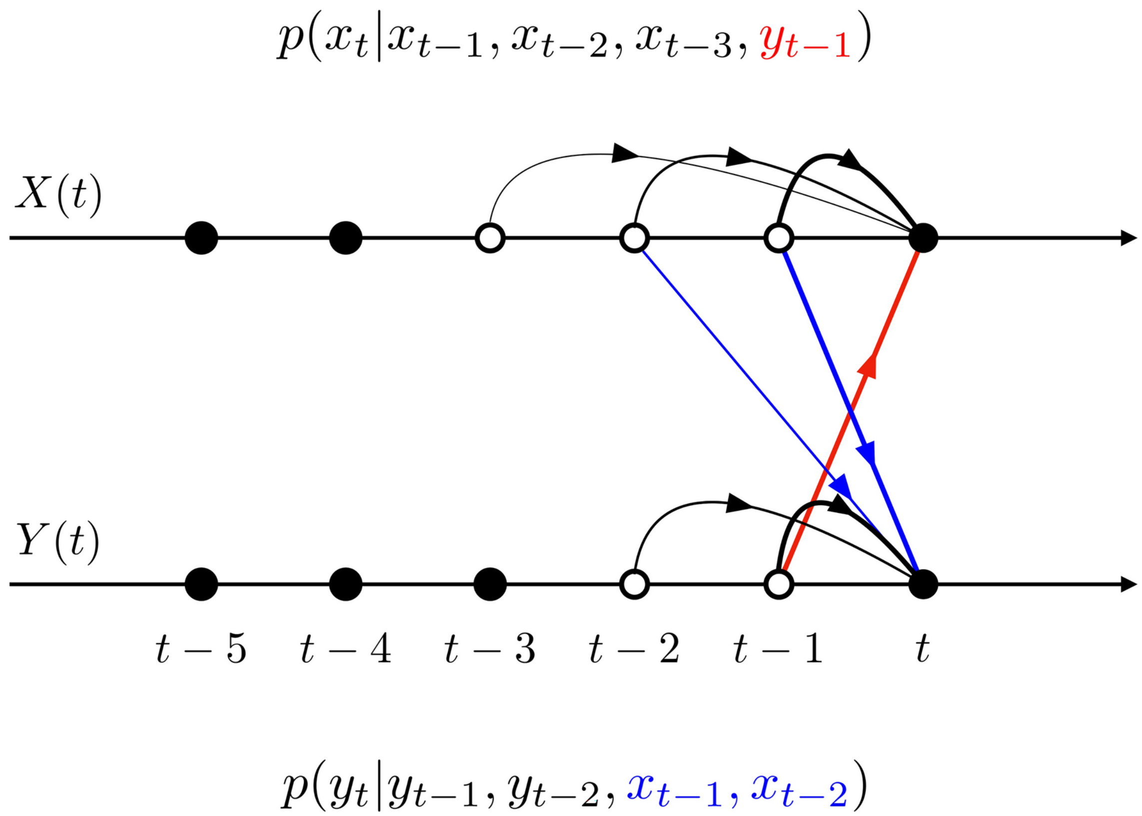


Fig. 1. Schematic representation of two interdependent processes, denoted as $X(t)$ and $Y(t)$, exhibiting distinct memory characteristics. The time steps for X and Y are indicated by dots. Unfilled dots represent the time steps that remain within the memory range of each process at time t. The Markov length, which refers to the number of time steps included in the memory, is three for X and two for Y. The connection between the latest time step of each process in the memory of the other is depicted by blue and red arrows. As illustrated, the Markov length of X with respect to Y is one, while the Markov length of Y with respect to X is two. The conditional probability density of each process reflects the dependence of that particular process at time t on its previous steps as well as the previous steps of the other process.

Similar to the memory function defined for a single process in Eq. (3), four memory functions are defined as follows:

$$
\begin{aligned}
\mathrm{M}^{xx}(\tau) &:= \int_{-\infty}^{\infty} |S(x_i, x_j; \tau)|\, dx_i dx_j, \\
\mathrm{M}^{xy}(\tau) &:= \int_{-\infty}^{\infty} |S(x_i, y_j; \tau)|\, dx_i dy_j, \\
\mathrm{M}^{yx}(\tau) &:= \int_{-\infty}^{\infty} |S(y_i, x_j; \tau)|\, dy_i dx_j, \\
\mathrm{M}^{yy}(\tau) &:= \int_{-\infty}^{\infty} |S(y_i, y_j; \tau)|\, dy_i dy_j.
\end{aligned}
\tag{7}
$$

Finally, the corresponding four Markov lengths are determined by looking for the time differences at which the memory functions change from non-zero to near-zero values.

Results

Synthetic data: Autoregressive model

In order to evaluate the generalized CKE for measuring Markov lengths, we apply our method to two synthetic time series with predetermined Markov lengths. To generate synthetic time series, we use a widely recognized statistical model known as the autoregressive (AR) model[39,40]. This model defines a dependent variable as a linear combination of its past values, and possibly a noise term. The autoregressive model of order p, denoted by $AR(p)$, is defined as:

$$x_t = \sum_{n=1}^{p} \phi_n x_{t-n} + \xi_t, \tag{8}$$

where x_t is the dependent variable at time t; x_{t-n} is the lagged value of the dependent variable; ϕ_n is the constant coefficient representing the relationship between the current value, x_t, and its lagged value, x_{t-n}; p indicates the number of past time steps related to the current state; ξ_t is a Gaussian noise with zero mean and constant variance σ_ξ^2. Using Eq. (8), we can generate time series that depends on the specific number of its previous steps.

Figure 2a shows a sample time series, x_t, for $p = 18$ and $\phi_n = 0.3/p$. Since the time series x_t is stationary, we use Eq. (3) to calculate the Markov length. Figure 2b shows that $\mathrm{M}(\tau)$ decreases to zero at $\tau = 19$ (taking the error into account). This abrupt transition implies that x_t depends on the 18 previous steps, which exactly agrees with $p = 18$. The red line in Fig. 2b represent the calculated error of the memory function, denoted as σ_{M}.

Inspired by the autoregressive model, we extend Eq. (8) to enable the generation of two dependent AR processes with specific Markov lengths. This allows us to assess the effectiveness of our generalized Chapman-Kolmogorov equation in determining the Markov lengths. Consider the following system of coupled equations governing the dynamics of the two processes x_t and y_t:

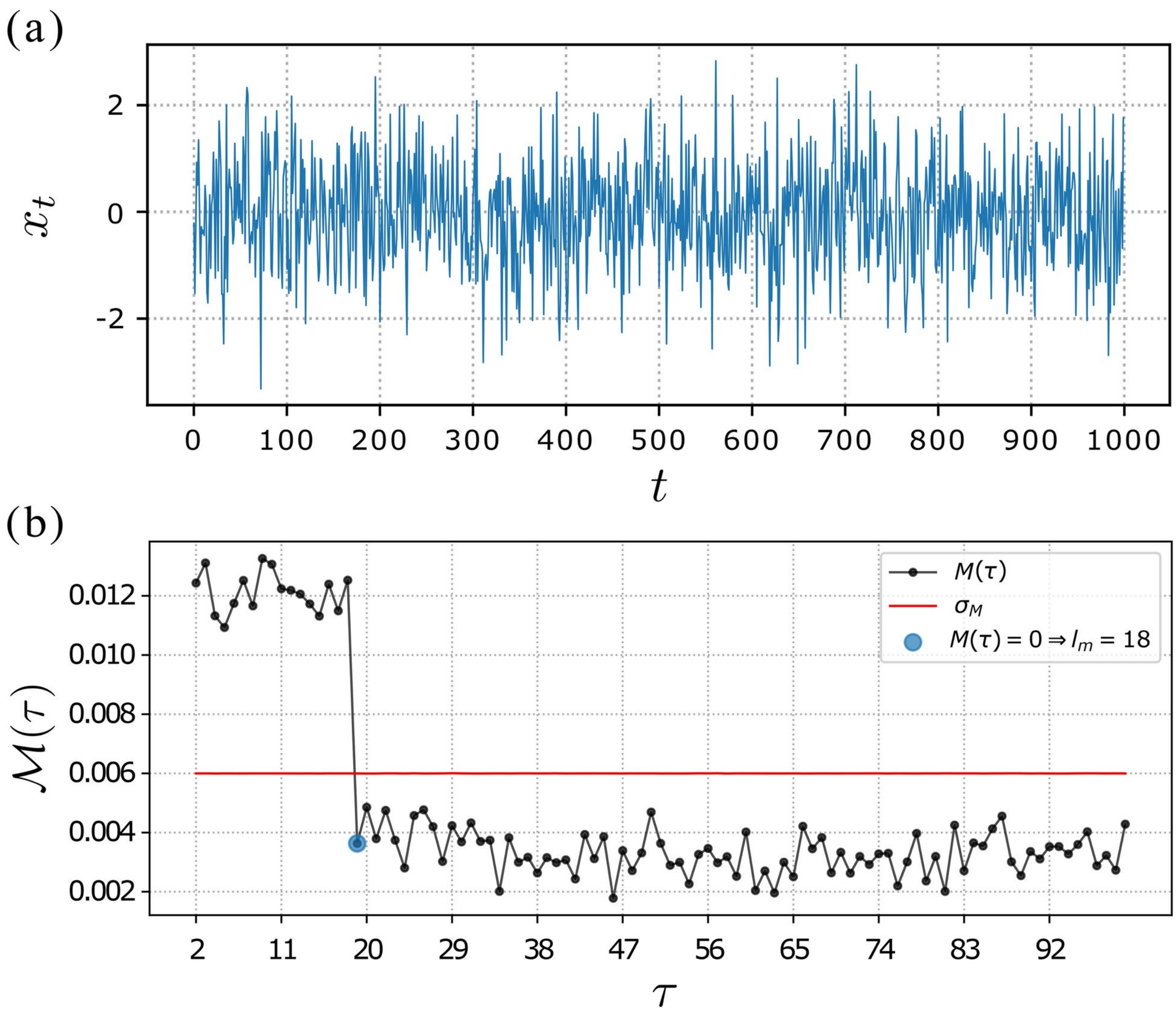


Fig. 2. (**a**) The figure presents a sample autoregressive process generated using Eq. (8) with $p = 18$, indicating that x_t is constructed to depend on the preceding 18 steps. (**b**) CKE is applied to the time series x_t. The plot depicts $\mathrm{M}(\tau)$ as a function of τ. $\mathrm{M}(\tau)$ reaches zero at $\tau = 19$ by considering the error σ_{M}. This zero-crossing indicates that the process depends on the preceding 18 steps, thereby establishing the Markov length $l_m = 18$.

$$x_t = \sum_{n=1}^{l_m^{xx}} \phi_n^{xx} x_{t-n} + \sum_{n=1}^{l_m^{xy}} \phi_n^{xy} y_{t-n} + \xi_t^x, \\ y_t = \sum_{n=1}^{l_m^{yx}} \phi_n^{yx} x_{t-n} + \sum_{n=1}^{l_m^{yy}} \phi_n^{yy} y_{t-n} + \xi_t^y, \tag{9}$$

where ϕ^{xx}, ϕ^{xy}, ϕ^{yx}, ϕ^{yy} are constant coefficients; ξ_t^x and ξ_t^y are stochastic Gaussian noises; l_m^{xx}, l_m^{xy}, l_m^{yx}, and l_m^{yy} are Markov lengths of *x* and *y* with respect to themselves and other process.

We generate two sample time series, x_t and y_t, using Eq. (9) for $l_m^{xx} = 20$, $l_m^{xy} = 10$, $l_m^{yx} = 15$, and $l_m^{yy} = 25$. The values of ϕ_n's used in this case are analogous to that employed in the single time series case. Figure 3 shows the memory function, $\mathrm{M}(\tau)$, for four Markov lengths. As observed, the generalized CKE demonstrates successful performance, as the estimated Markov lengths ,taking into account the error σ_{M}, are equal to the initially chosen values utilized to generate the two time series.

Empirical data: Cryptocurrencies

This section provides the outcomes obtained from applying our method to empirical financial market data. We specifically focus on analyzing two cryptocurrency markets: Bitcoin (XBT/USD) and Ethereum (ETH/USD), using the permanent contracts offered by the Bitmex exchange. The objective is to calculate Markov lengths for the price dynamics of these two cryptocurrencies. The data period is three months, from the beginning of May to the end of July, 2020. The final price of each contract is determined per second using the price of every single transaction; see Fig. 4.

The initial step preceding the estimation of the Markov lengths using the generalized CKE method is to calculate the logarithmic returns, denoted as R_t, based on the market price p_t. The logarithmic return is defined by the following relation:

$$R_{t+1} = \ln\left(\frac{p_{t+1}}{p_t}\right), \tag{10}$$

where p_t denotes the price at time *t*. Logarithmic returns are preferred over actual prices when analyzing financial markets. They offer several advantages, such as facilitating the interpretation and comparison of price changes across different assets and periods, which is done by changing the focus to relative price changes rather than absolute price levels. Moreover, they tend to be more stationary than raw price data, enhancing their usefulness for statistical analysis and modeling.

Because of the nature of financial markets, estimating Markov length on an annual or monthly basis is not advisable. However, we should take care of very short time intervals[41] as the number of data points decreases, and consequently, the density functions might not be accurate enough. Therefore, we have implemented our method on a weekly basis. Here, we separately present the results for three weeks. The price returns of XBT/USD and ETH/USD are represented by x_t and y_t, respectively.

Figure 5 illustrates the memory functions for the initial week spanning from May 27th to June 3rd, 2020. The memory functions, denoted as M^{xx} and M^{yy}, represent the influence of process history on itself, while M^{xy} and M^{yx} signify the impact of one process history on another. To determine the corresponding Markov length for each of the four memory functions, we identify the error threshold, σ_{M}, depicted in Fig. 5 by the red horizontal curves. We employ the same methodology for analyzing two next weeks: July 13th to July 20th, 2020, and July 22nd to July 29th, 2020, see Figs. 6 and 7.

The data show that the volatility of Bitcoin and Ethereum was at its greatest for the entire three months during the first sample week. As a result, the Markov lengths of the two coins are high (526 and 436 seconds, respectively), and the dependency duration between the two coins is also significant (approximately 300 seconds). Note that the dependency between the two coins is not symmetric. Ethereum exhibits a slightly higher level of dependence on Bitcoin compared to the dependence of Bitcoin on Ethereum, as depicted in Fig. 5. During the second sample week, which is characterized by minimal fluctuations in the overall data, the Markov lengths are observed to be relatively small, see Fig. 6. Finally, for the third sample week illustrated in Fig. 7, all Markov lengths expand further compared to the previous two samples, showing a longer memory effect during this week.

Conclusion

Estimating future events based on past events has always been challenging when dealing with non-deterministic and stochastic processes. The well-known mathematical model of the Markov process explains how processes transition from previous states. This particular class of processes helps to understand non-Markovian processes that rely on more than just the immediately preceding step. The length of this dependency is referred to as Markov length, which is calculated by applying the Chapman-Kolmogorov equation.

In reality, it is rare to find a phenomenon that exists in isolation from all other phenomena, as various phenomena interact with each other in direct and indirect ways. The level of reliance between a process and other processes has reached a stage that sometimes exceeds the reliance on its own past. Certain phenomena can emerge and continue to influence other phenomena. For instance, increased oil prices today may alter the cost of various products in the future. In this research we have used real data with heavy-tailed probability distribution functions which is the characteristic of financial markets. Here, a question may raise that the dependent Markov length of coupled process are influenced by the probability distribution. For instance, in dependent financial

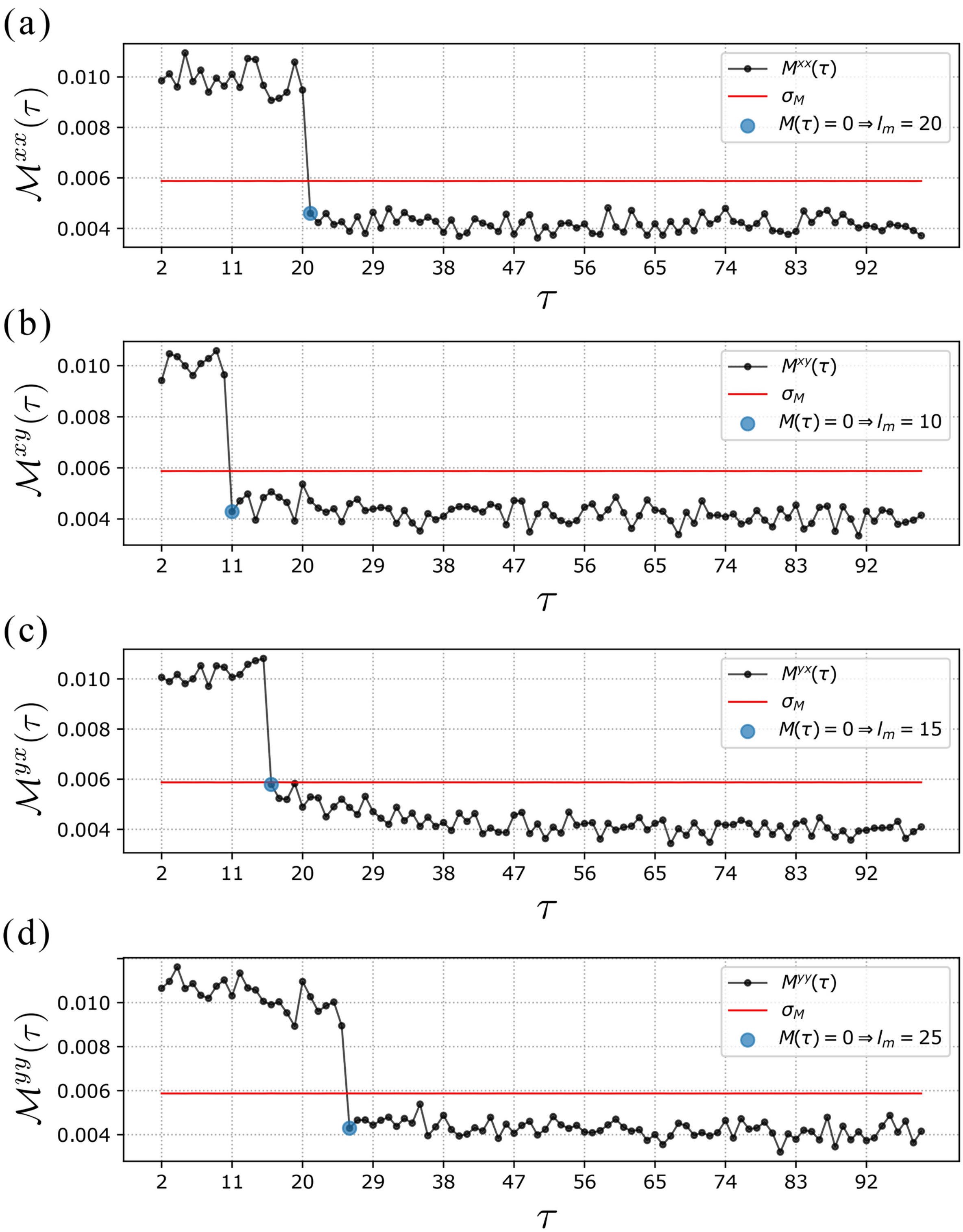


Fig. 3. CKE results for the generated data using Eq. 9 with the following Markov lengths: $l_m^{xx}=20$, $l_m^{xy}=10$, $l_m^{yx}=15$, and $l_m^{yy}=25$. The plot depicts the memory function $\mathrm{M}(\tau)$ as a function of τ, and by considering the error σ_{M}, we observe that the four Markov lengths obtained from CKE match the initial values with which the time series were generated.

markets, such as two competing companies, if the price of the product of a company has a small change, the others may not get affected but in case the volume of the change is significant, this would seriously influence the price of the other companies product. Thus, we anticipate that in high frequency range the markets behave independently, however, in large scale events dependencies emerge. In the Appendix of[42], it is explained that

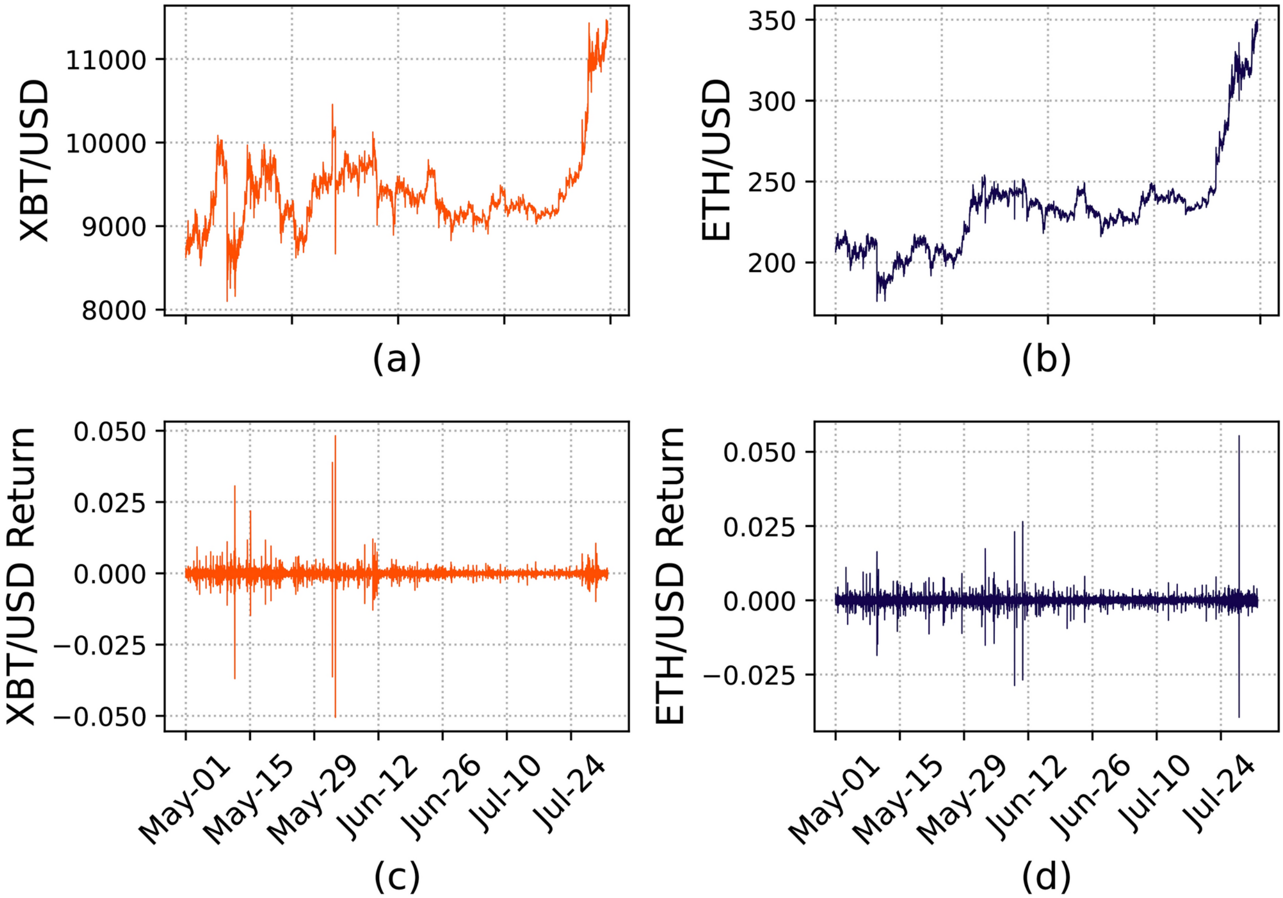


Fig. 4. (**a**) and (**b**) depict the coevolution of Bitcoin (XBT) and Ethereum (ETH) prices from May 1 to July 31, 2020. In contrast, (**c**) and (**d**) represent the logarithmic returns of XBT and ETH, respectively.

in order to investigate fat-tail data, we need to surrogate the data, meaning that we apply Fourier transform on the return data and then randomize the phases and finally apply inverse Fourier transform. This process convert fat-tail data to Gaussian data. Now, we can re-calculate the Markov length and the reported results show smaller Markov length.We have applied the surrogate method on our return data of real markets, and the results are shown in Fig. 8. This observation imply that after surrogate, the Markov length is significantly decreased. This indicates that the Markov length can be influenced when the market is low frequency. In low frequency state where we have heavy-tailed data and the significant events occur, the markets can be affected by the other markets state rather than their own history.

Such observations led us to generalize the Chapman-Kolmogorov equations and suggest a method to estimate the duration of the mutual past dependency, which we call the mutual Markov length. Once we applied our method to the synthetic time series produced by the auto-regressive model, we proceeded to employ cryptography as real data-sets. We tested these data-sets to determine the specified mutual Markov length. The approach proposed in this study demonstrates its capability to identify the mutual dependency between processes based on the past impact of one process on the current state of another.

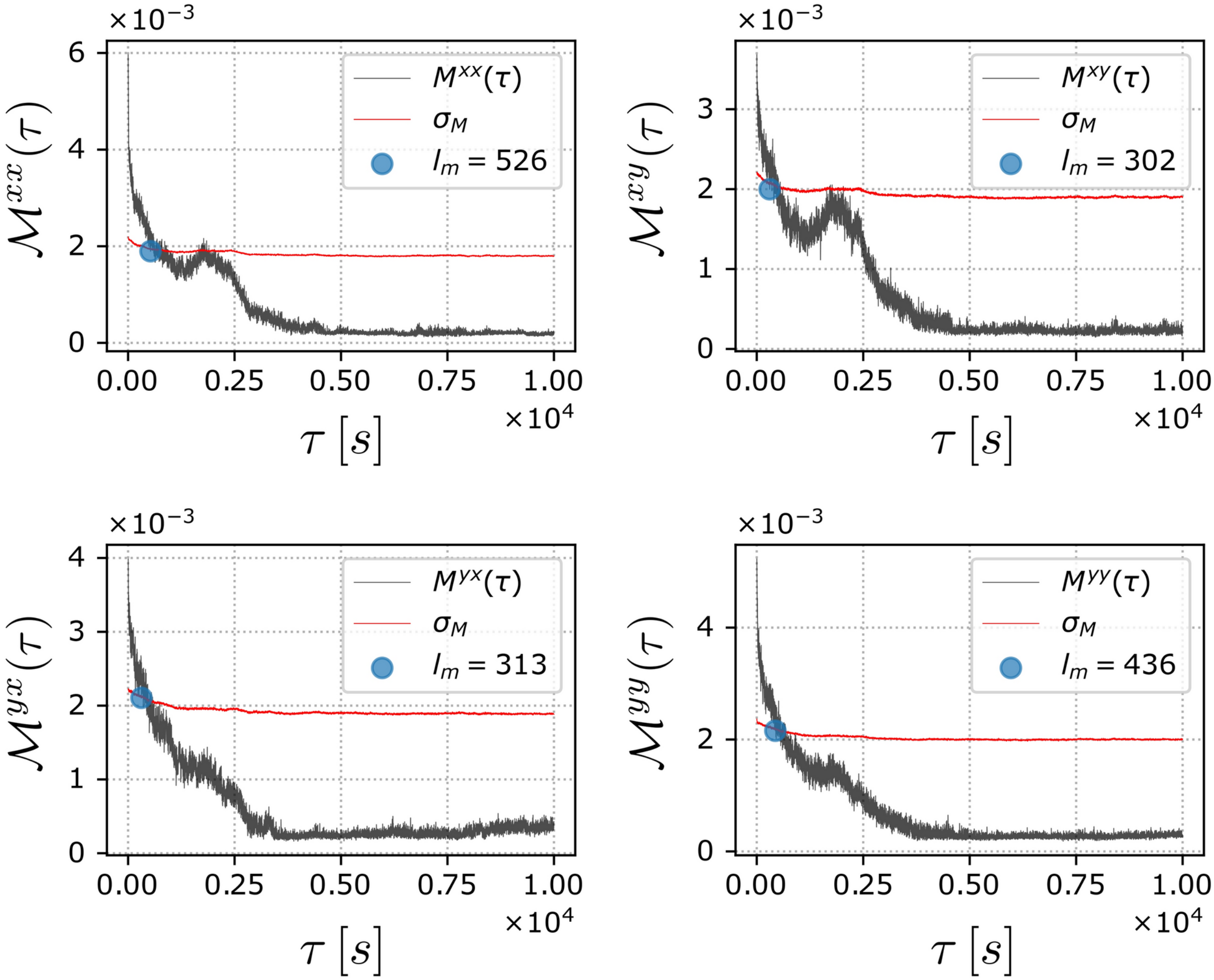


Fig. 5. The CKE results are presented for the first week, which extends from May 27th to June 3rd, 2020. During this interval, price fluctuations of Bitcoin reached their maximum compared to the other weekly intervals. The superscripts *x* and *y* represent Bitcoin and Ethereum, respectively. As an example, in the top left figure, M^{xx} signifies the CKE result for Bitcoin, indicating the influence of its past on itself.

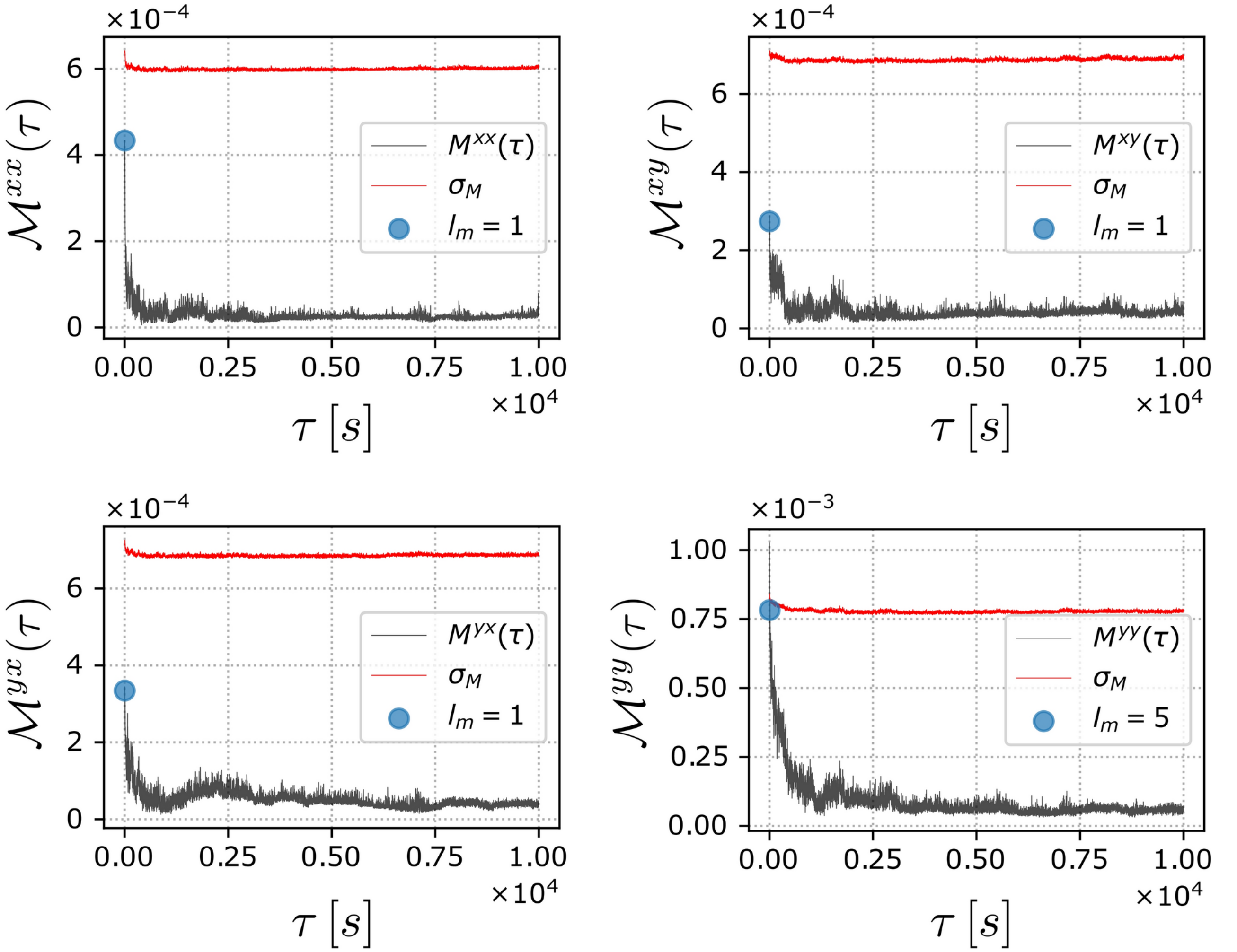


Fig. 6. The CKE results are presented for the second week, which extends from July 13th to July 20th, 2020. During this interval, price fluctuations of Bitcoin reached their minimum compared to the other weekly intervals.

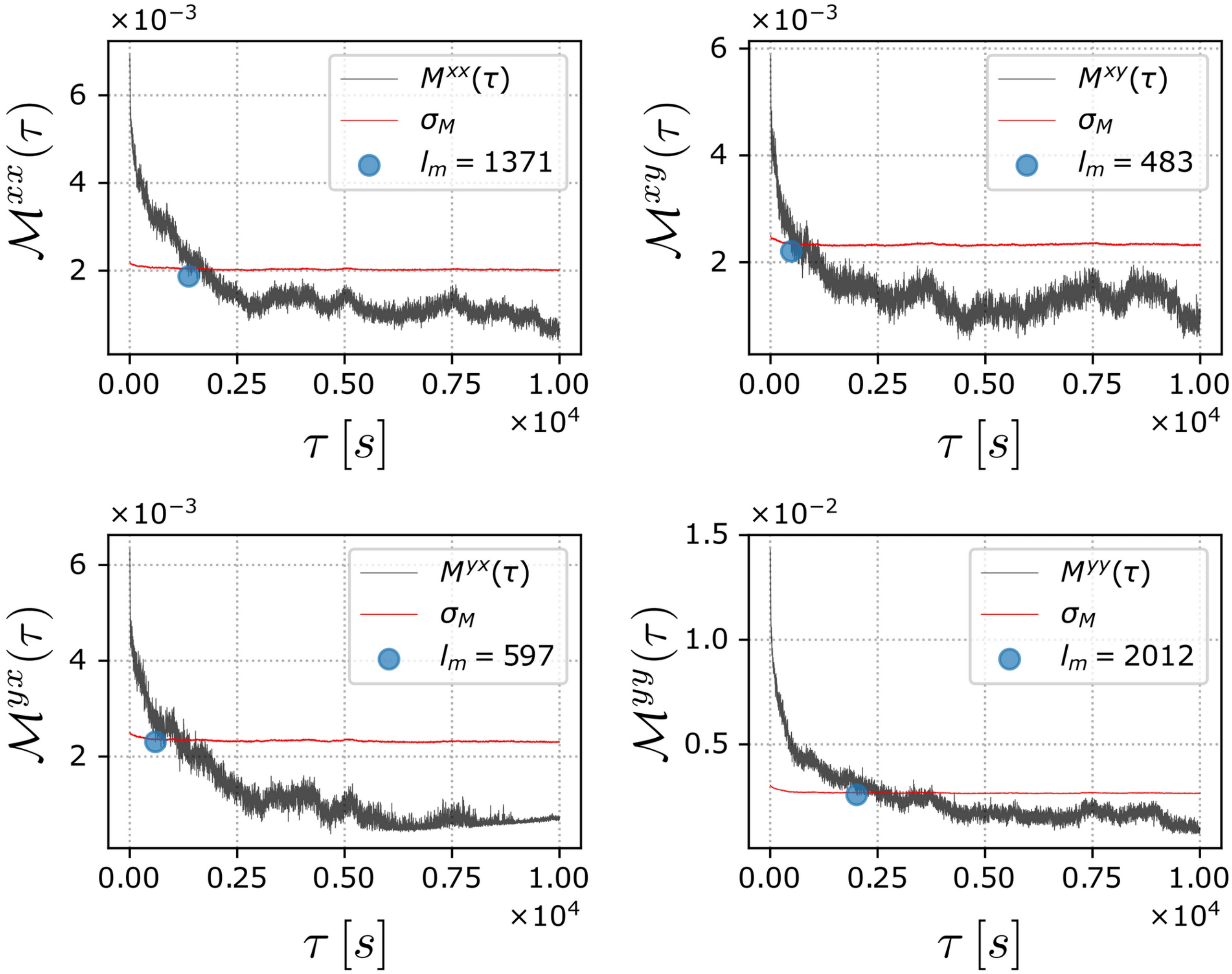
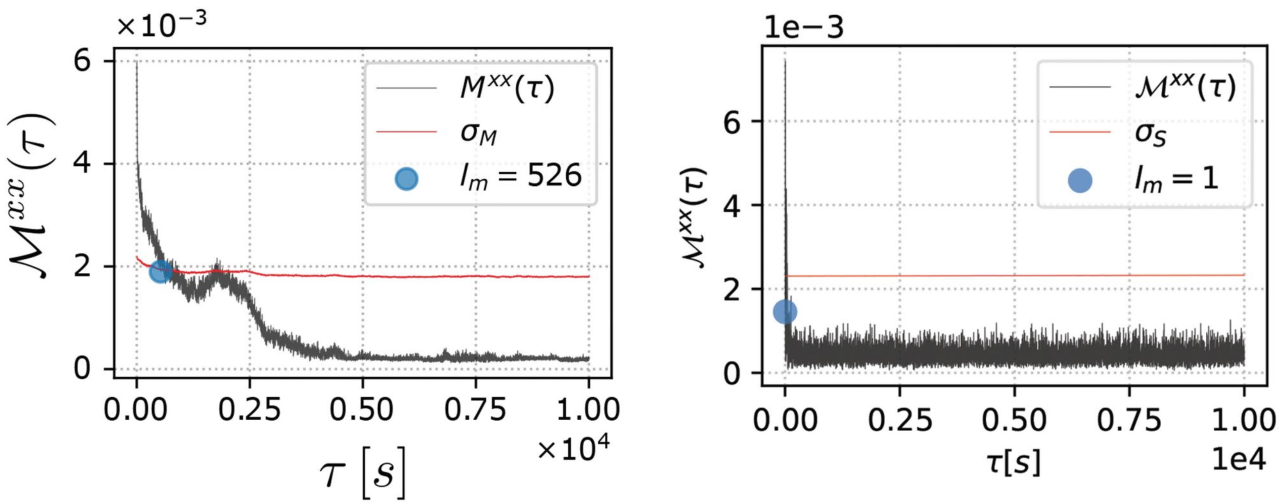


Fig. 7. The CKE results are presented for the third sample week, covering the period from July 22nd to July 29th, 2020. During this interval, the price fluctuation of Bitcoin was near its three-month mean value.



(a) (b)

Fig. 8. The CKE results are presented for the first week data (**a**) and for the the surrogate of it (**b**), for $\mathrm{M}^{xx}(\tau)$ memory function, which shows a significant decrease on the Markov length.

Data availability

Data analyzed during the current study is publicly available in a repository: https://public.bitmex.com/

Received: 27 February 2024; Accepted: 26 February 2025
Published online: 04 April 2025

References

1. Jafari, G. et al. Stochastic analysis and regeneration of rough surfaces. *Phys. Rev. Lett.* **91**, 226101 (2003).
2. Fazeli, S., Shirazi, A. & Jafari, G. Probing rough surfaces: Markovian versus non-markovian processes. *New J. Phys.* **10**, 083020 (2008).
3. Stroe-Kunold, E., Stadnytska, T., Werner, J. & Braun, S. Estimating long-range dependence in time series: An evaluation of estimators implemented in r. *Behav. Res. Methods* **41**, 909–923 (2009).
4. Anvari, M., Tabar, M. R. R., Peinke, J. & Lehnertz, K. Disentangling the stochastic behavior of complex time series. *Sci. Rep.* **6**, 35435 (2016).
5. Risken, H. *Fokker-planck equation* (Springer, 1996).
6. Van Kampen, N. G. *Stochastic processes in physics and chemistry* Vol. 1 (Elsevier, 1992).
7. Waechter, M., Riess, F., Schimmel, T., Wendt, U. & Peinke, J. Stochastic analysis of different rough surfaces. *Eur. Phys. J. B-Condensed Matter Complex Syst.* **41**, 259–277 (2004).
8. Friedrich, R., Peinke, J., Sahimi, M. & Tabar, M. R. R. Approaching complexity by stochastic methods: From biological systems to turbulence. *Phys. Rep.* **506**, 87–162 (2011).
9. Peinke, J. et al. Markov property of lagrangian turbulence. *Bulletin of the American Physical Society* (2022).
10. Jannesar, M., Sadeghi, A., Meyer, E. & Jafari, G. R. A langevin equation that governs the irregular stick-slip nano-scale friction. *Sci. Rep.* **9**, 12505. https://doi.org/10.1038/s41598-019-48345-4 (2019).
11. Cao, W., Demazeau, Y., Cao, L. & Zhu, W. Financial crisis and global market couplings. In: *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 1–10 (IEEE, 2015).
12. Islam, I. & Verick, S. The great recession of 2008–09: Causes, consequences and policy responses. *From the Great Recession to labour market recovery: Issues, evidence and policy options* 19–52 (2011).
13. Farahpour, F. et al. A langevin equation for the rates of currency exchange based on the markov analysis. *Physica A* **385**, 601–608 (2007).
14. Ghasemi, F. et al. Markov analysis and kramers-moyal expansion of nonstationary stochastic processes with application to the fluctuations in the oil price. *Phys. Rev. E* **75**, 060102 (2007).
15. Bressloff, P. C. *Stochastic processes in cell biology* Vol. 41 (Springer, 2014).
16. Dougherty, E. R. *Random processes for image and signal processing* (SPIE Optical Engineering Press Bellingham, 1999).
17. Manshour, P. et al. Turbulencelike behavior of seismic time series. *Phys. Rev. Lett.* **102**, 014101 (2009).
18. Steele, J. M. *Stochastic calculus and financial applications* Vol. 1 (Springer, 2001).
19. Shreve, S. E. & Karatzas, I. *Methods of mathematical finance* (Springer, 2001).
20. Ardalankia, J., Askari, J., Sheykhali, S., Haven, E. & Jafari, G. R. Mapping coupled time-series onto a complex network. *Europhys. Lett.* **132**, 58002 (2021).
21. Arezki, M. R. *et al.Oil prices and the global economy* (International Monetary Fund, 2017).
22. Chemkha, R., BenSaïda, A. & Ghorbel, A. Connectedness between cryptocurrencies and foreign exchange markets: Implication for risk management. *J. Multinatl. Financ. Manag.* **59**, 100666 (2021).
23. Mallick, S. K. & Mallik, D. A. A study on the relationship between crypto-currencies and official indian foreign exchange rates. *Mater. Today Proc.* **80**, 3786–3793 (2023).
24. Ghasemi, F., Sahimi, M., Peinke, J. & Tabar, M. R. R. Analysis of non-stationary data for heart-rate fluctuations in terms of drift and diffusion coefficients. *arXiv preprint physics/0603130* .
25. Gao, J., Hou, Y., Fan, F. & Liu, F. Complexity changes in the us and china's stock markets: Differences, causes, and wider social implications. *Entropy* **22**, 75 (2020).
26. Doering, C. R. Modeling complex systems: Stochastic processes, stochastic differential equations, and fokker-planck equations. In *1990 Lectures in Complex Systems*, 3–52 (CRC Press, 2018).
27. Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D. & Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.* **5**, 4630 (2014).
28. Bühlmann, P. & Wyner, A. J. Variable length markov chains. *Ann. Stat.* **27**, 480–513 (1999).
29. Gagniuc, P. A. *Markov chains: from theory to implementation and experimentation* (John Wiley & Sons, 2017).
30. van Kampen, N. G. Remarks on non-markov processes. *Braz. J. Phys.* **28**, 90–96 (1998).
31. Ciobanu, G. Analyzing non-markovian systems by using a stochastic process calculus and a probabilistic model checker. *Mathematics* **11**, 302 (2023).
32. Gardiner, C. W. *Handbook of stochastic methods* Vol. 3 (Springer, 1985).
33. Jafari, G. R., Sahimi, M., Rasaei, M. R. & Tabar, M. R. R. Analysis of porosity distribution of large-scale porous media and their reconstruction by langevin equation. *Phys. Rev. E* **83**, 026309 (2011).
34. Durrett, R. *Probability: theory and examples* Vol. 49 (Cambridge University Press, 2019).
35. Park, K. I. & Park, M. *Fundamentals of probability and stochastic processes with applications to communications* (Springer, 2018).
36. Pavliotis, G. A. *Stochastic processes and applications: Diffusion processes, the Fokker-Planck and Langevin equations* Vol. 60 (Springer, 2014).
37. Haken, H. & Mayer-Kress, G. Chapman-kolmogorov equation and path integrals for discrete chaos in presence of noise. *Zeitschrift für Physik B Condensed Matter* **43**, 185–187 (1981).
38. Gillespie, D. T. *Markov processes: An introduction for physical scientists* (Elsevier, 1991).
39. Gandhi, V. Chapter 2-interfacing brain and machine. *Brain-Computer Interfacing for Assistive Robotics* 7–63 (2015).
40. Shumway, R. H. & Stoffer, D. S. *Time series analysis and its applications: With R examples* (Springer, India, 2017).
41. Nawroth, A. P. & Peinke, J. Medium and small-scale analysis of financial data. *Physica A* **382**, 193–198 (2007).
42. Hedayatifar, L., Vahabi, M. & Jafari, G. Coupling detrended fluctuation analysis for analyzing coupled nonstationary signals. *Phys. Rev. E* **84**, 021138 (2011).

Author contributions

H. M conducted the calculations and analyses, deriving the results. M. GH initialized the draft, engaged in result discussions, and contributed to finalizing the draft. T. J and GR.J were instrumental in conceiving the initial concept, actively participated in result discussions, and contributed to finalizing the draft. GR.J provided supervision throughout the project. All authors reviewed the manuscript.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92238-8.

Correspondence and requests for materials should be addressed to G.R.J.