Bitcoin Transaction Behavior Modeling Based on Balance Data

***Abstract*—Do Bitcoin users' transaction behaviors follow some mechanism? This is a significant question. In this paper, we found that users' balance follow the log-normal distribution, but their transaction behaviors didn't follow Gibrat's proportional growth rule by calculating corresponding parameters. During the exploration process, we also found that there were mainly two kinds of bitcoin users who behave differently. Users who own only a bit of bitcoins, in the beginning, tend to buy more bitcoins, while users who own more bitcoins in the beginning tend to sell bitcoins.**

***Index Terms*—Balance distribution, log-normal distribution, Gibrat's proportional growth, transaction behavior.**

I. Introduction

The Bitcoin transaction network provides a chance for us to research Bitcoin users' behavior modes. If we define S_i as the cryptocurrency balance owned by the i^{th} user. Can the change of cryptocurrency balance (dS_i) be modeled by the stochastic equation $dS_i = {S_i}^{\alpha} \cdot \mu \cdot dt + {S_i}^{\alpha} \cdot \sigma \cdot dw_i$ equation? In the cryptocurrency world, the question of what the exact value of the exponent α of S_i in the above formula is also very important and deserves to be researched because it denotes the mechanism that leads to the scale-free distribution of bitcoin balance data.

Lots of papers have confirmed that the indegree and outdegree of Bitcoin transaction networks were distributed as power-law and this result could be explained by linear degree preferential attachment. When it comes to users' bitcoin balance (the number of bitcoins owned by each user), its formation mechanism is not linear preferential attachment according to [3] even if users' bitcoin balance distribution follows scale-free rules. [4] compared the constructed index "cumulative distribution function of rank function" to the corresponding theoretical one visually and concluded that the transaction of bitcoin follows sublinear preferential attachment. One shortcoming of this research is that they just took every address as one node, and didn't cluster these addresses to the user level. Another shortcoming is that they actually got the conclusions by only plotting but not by statistical methods, for which it is easy to get wrong conclusions [5]. Thus, it is necessary to give a deep insight into how users' bitcoin balance evolves, what is the mechanism behind it, and what mechanism leads to current bitcoin balance distribution.

In the following section, we first choose the proper bitcoin balance data and explore these empirical balance data to find out the basic facts; secondly, we analyze the mechanism that leads to the current balance distribution based on the Geometric Brownian Motion model (GBM); in the last part, we summarized and discuss this paper.

II. Data Description and Exploration

A. Data Description

We chose the bitcoin balance data on 2016-01-23. There is nothing special about this choice except that the bitcoin transaction network is more mature and relatively more stationary at that time compared with the earlier date. The log-log scale histogram in Fig. 1 indicates that the users' bitcoin balance is a heavy-tail distribution.

When it comes to the balance distribution, we need to distinguish two kinds of users. The first kind of users are those whose balances on 2016-01-23 are positive and who have transactions during the next period dt (it is 28 days in the right panel), which is called user group A. The other kinds of users are those whose balances on 2016-01-23 are positive, but who do not have transactions during the next period dt, which is called user group B. This differentiation is important and necessary, otherwise, data from users who have not transacted for a long time will affect the accuracy of our analysis, for example, a dead bitcoin address.

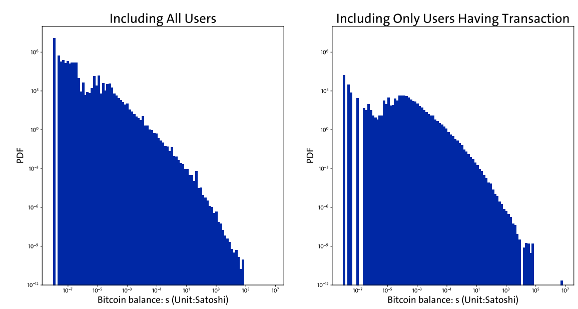


Fig. 1: Bitcoin balance distribution on 2016-01-23 (unit: satoshi, 1 bitcoin=10^8 satoshi). The left panel depicts the probability distribution function of all users (groups A and B), and the right panel depicts only users whose balances are positive and who have transactions (group A) during the next period dt starting from 2016-01-23 (it is 28 days in the right panel).

The left panel in Fig. 1 depicts the probability distribution function (pdf) with both kinds of users' balance data. The right panel in Fig. 1 depicts the pdf with the balance data of only those whose balances on 2016-01-23 are positive and who have transactions during the next period dt.

Then applying the python power-law package developed by [3], we fitted the bitcoin balance data on 2016-01-23 to the power-law and log-normal distribution and found that the log-normal distribution fits the data better.

Fig. 2 and statistical test in Fig. 3 also confirmed that the log-normal distribution is better than the power-law in fitting the balance distribution data.

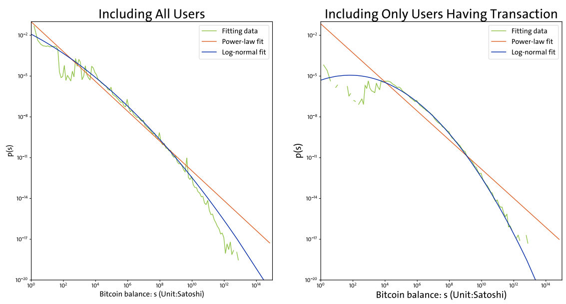


Fig. 2: Bitcoin balance fitting (unit: satoshi). The left panel depicts the fitting results with the data from all users on 2016-01-23 (groups A and B), and the right panel depicts the fitting result with data from those users whose balances are positive and who have transactions (group A) during the next period dt starting from 2016-01-23 (it is 28 days in the right panel).

We also compared the fitting results between the log-normal distribution and power-law distribution by gradually increasing the minimum fitting data. The first minimum fitting data we choose is 10^{-8} bitcoin, and the increasing step is 1 bitcoin. That means that the minimum fitting data in second time is 1 bitcoin, then 2 bitcoins the third time, and so on. The result shows that the log-normal function is always a better fitting than the power law in most cases. Even if the data is generated by power-law distribution, we can't still refuse the hypothesis that the data is from a log-normal distribution only if its variance is huge just by fitting the data using the package developed by [3]. So, it is not enough to get the conclusion that our empirical data comes from power-law distribution or log-normal distribution by this statistic package.

The uniformly most powerful unbiased (UMPU) Wilks test as suggested by [9] can be used to distinguish power-law distribution and log-normal distribution. This method comes from the idea that exponentiality can be tested against normal distribution [10] [11] using the saddle point approximation method and the idea that power-law distribution and log-normal distribution can be transferred to exponential distribution and normal distribution after taking log calculation to bitcoin balance data, respectively. The null hypothesis for this test is that the data is distributed as a power-law, and its alternative hypothesis is that the data is distributed as log-normal. The test is performed as follows: Firstly, we choose a threshold for the bitcoin balance; secondly, the UPMU Wilks test is performed for bitcoin balance whose value is larger than the threshold by computing the p-value. Though the Monte Carlo method can also be used to calculate the p-value, it is very time-consuming here because we have millions of data. As shown in Fig. 3, we can reject the null hypothesis and accept the alternative hypothesis in almost all regions of bitcoin balance except regions that include only tens of the largest value of bitcoin balance. However, the proportion of tens of the largest value of bitcoin to a total number of bitcoins in our specific time-point is no more than 5%.

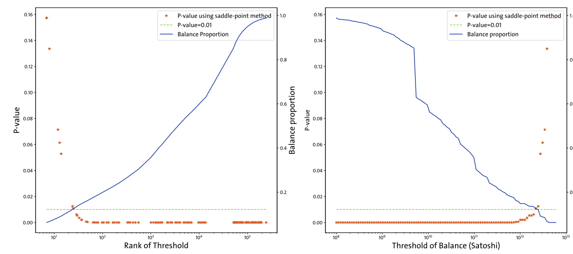


Fig. 3: Left panel is the p-value versus the rank of the chosen threshold of bitcoin balance on 2016-01-23. After the threshold is chosen, all bitcoin balance value above this threshold is sorted inversely, namely the largest value of bitcoin balance rank one, then the second largest value of bitcoin rank two. The ranking proceeds until the chosen threshold. The right panel is the p-value versus the threshold of bitcoin balance.

B. Data Exploration

Now that we know that the balance distribution may be log-normal, a natural question is what is the mechanism behind the transaction that leads to the log-normal distribution?

Gibrat's proportional growth law is an important tool in explaining the forming mechanism of power-law distribution if we change Gibrat's proportional growth equation a bit [3], and especially in Zipf's distribution when taking other mechanism into consideration, such as birth process and death process. However, we also understand that the probability density function (pdf) of the solution of standard Gibrat's proportional growth is log-normal distribution which is better than the power-law distribution in fitting our bitcoin balance data. At the same time, our test confirms that the bitcoin balance distribution function is fitted well by log-normal distribution. We also checked the distribution of bitcoin balance on 2019-01-19 which is almost three years later than 2016-01-23, and we find that the distribution is also log-normal. Can Gibratt's proportional growth law be the mechanism to explain our data? To answer this question, we need to investigate our data first and then check whether the exponent in Gibrat's proportional growth equation is one or neither.

We first depict the scatter plot of bitcoin balance data (s) versus bitcoin balance change (ds) data to get a comprehensive impression of the holistic data distribution. Fig. 4 is the scatter plot of bitcoin balance versus bitcoin balance change within a half year. Four sub-scatter plots with different scales are depicted so that it is easy for us to look closely into the data.

There exist different clusters in the data. Hopkin statistics (smaller than 10^{-4} in the condition of 100 random samples) test (the null hypothesis is that there is only one cluster

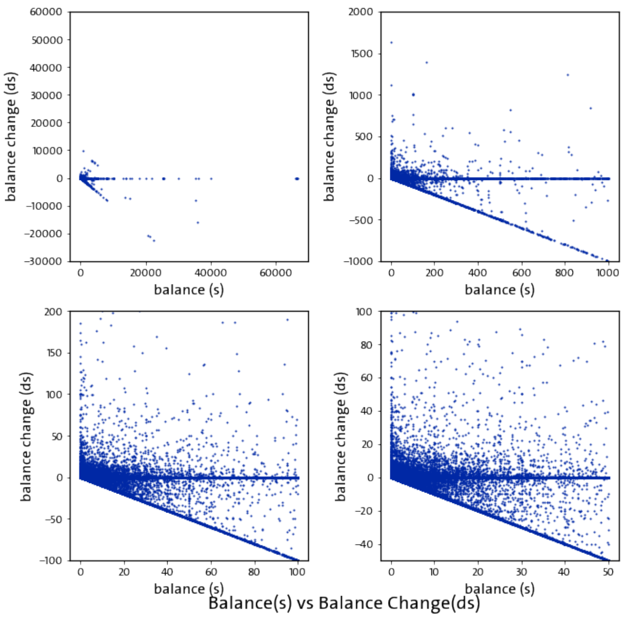


Fig. 4: Bitcoin balance and balance change on 2016-01-23, the time interval (dt) is half a year.

in the data; the alternative hypothesis is that there is more than one cluster in the data) also confirms the existence of multiple clusters. There are lots of methods for clustering data, such as distance-based K-means method, and probability-based Gaussian Mixture models. Both methods work well if data points are spared in a circle shape. Gaussian Mixture models also assume that it is Gaussian distributed in all dimensions of data points which is not the case in our data. It can also be found that there are three straight lines above Fig. 4, the vertical line, the horizontal line, and the diagonal line that correspond to different situations. The vertical line corresponds to those users who don't own or own only a small number of bitcoins on 2016-01-23 but got lots of bitcoins by trading before 2016-02-20 (28 days later after 2016-01-23). The horizontal line corresponds to those users who own bitcoins and the number of bitcoins didn't change in the time interval between 2016-01-23 and 2016-02-20. The diagonal line corresponds to those users who owned bitcoins on 2016-01-23 but sold them all before 2016-02-20.

As shown in Fig. 4, the log-normal function fits better, especially in the tail if data on the horizontal line is deleted compared with the case that all data points are reserved.

Based on our data exploration, we think that we should delete the data point on the horizontal line because these corresponding users didn't take part in trading activities in our specific time span.

III. Mechanism Detection

Now, we explore the mechanism behind bitcoin distribution. As before, we still define S_i as the cryptocurrency balance owned by the i^{th} user. The change of cryptocurrency balance (dS_i) can be modeled as follows if they follow the Geometric Brownian Motion (GBM) mechanism:

$$dS_i = {S_i}^{\alpha} \cdot \mu \cdot dt + {S_i}^{\alpha} \cdot \sigma \cdot dw_i \tag{1}$$

where dt is time interval, dS_i is the balance change of the i^{th} user during dt, w_i is the Brownian Motion, μ and σ is the drift and volatility, respectively. α is the exponent we will focus on. We can get the following equation by taking the expectation and variance on both sides of the equation 1:

$$\begin{cases} E(dS_i) = {S_i}^{\alpha} \cdot \mu \cdot dt \\ \sigma(dS_i) = {S_i}^{\alpha} \cdot \sigma \cdot \sqrt{dt} \end{cases} \tag{2}$$

To plot the equation 2, we need to bin the bitcoin balance first. The whole process includes three main steps:

- At first, the range of bitcoin balance is split as n ($n = 300$) consecutive bins;
- Then, we classify the bitcoin balance data (dS_i) and corresponding bitcoin balance change data (dS_i) according to bins that we choose. After classifying, we delete those bins where the number of data (dS_i) is less than 50 and the corresponding bitcoin balance change data (dS_i).
- At last, we calculate the average and standard deviation of dS_i in each bin.

Because the bitcoin balance distribution is scale-free, there are no data or only a few data points in lots of bins that correspond to large bitcoin balances, and most data are located in bins that correspond to several small balances. So, we think it would be a good choice to apply exponential bins and we got 167 data points which were shown in Fig. 5.

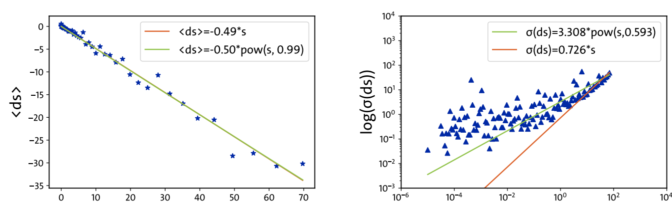


Fig. 5: The left panel is bitcoin balance versus the average of bitcoin balance change; the right panel is bitcoin balance versus standard variation of bitcoin balance change.

We calculated the fitting results based on equation 2 first. Because there are both negative and positive values in the average of a bitcoin balance change, it is not possible to use a log-log scale coordinate system to show the fitting equation 2. So, we still use a constant scale coordinate system to show our data and fitting results. As shown in the left panel of Fig. 5, the red straight line corresponds to the case of proportional growth (exponent α is set to 1 in equation 2), we only need to calculate the value of $\mu \cdot dt$. The green line corresponds to the case that both exponent α and $\mu \cdot dt$ were calculated by fitting. By comparing visually and making regressions, it seems that the exponent α is 1 in equation 2 can be accepted.

In the right panel of Fig. 5, the relationship between $\sigma(dS_i)$ and S_i is shown. The red line corresponds to the case of proportional growth (exponent α is set to 1 in equation 2). We fit the data points into the model $\sigma(dS_i) = S_i \cdot \sigma \cdot \sqrt{dt}$ and get the red line. The green line corresponds to the case in which we calculated the exponent α by making a regression. The exponent α is 0.739 After fitting the data points to the model $\sigma(dS_i) = {S_i}^{\alpha} \cdot \sigma \cdot \sqrt{dt}$. The exponent value we get from the first equation in 2 is very different from the exponent value we get by fitting the second equation in 2. Does this result denote that the exponent α in the volatility term is different from the exponent α in the drift term?

Because the absolute value of bitcoin balance change varies a lot for different users, we turn to research the ratio of balance change to balance. We can get the following equation 3 by dividing S_i in both sides of equation 2:

$$\begin{cases} E(\frac{dS_i}{S_i}) = {S_i}^{\alpha-1} \cdot \mu \cdot dt \\ \sigma(\frac{dS_i}{S_i}) = {S_i}^{\alpha-1} \cdot \sigma \cdot \sqrt{dt} \end{cases} \tag{3}$$

Based on equation 3, there will be no relationship between $E(\frac{dS_i}{S_i})$, $\sigma(\frac{dS_i}{S_i})$ and S_i if $\alpha = 1$. The only difference between our current calculation and previous ones is that we need to calculate the average and standard variance of $\frac{dS_i}{S_i}$ in each bin, now, but not dS_i.

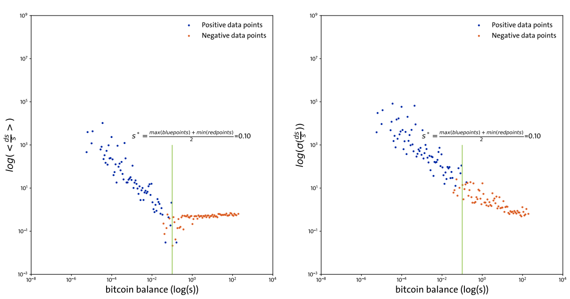


Fig. 6: The left panel is $E(\frac{dS_i}{S_i})$ versus bin value s, the right panel is the standard variance $\sigma(\frac{dS_i}{S_i})$ versus bin value s. The coordinate is a log-log scale. The blue points correspond to these bins whose $E(\frac{dS_i}{S_i})$ is positive and red points correspond to these bins whose $E(\frac{dS_i}{S_i})$ is negative. Because the log function can't be applied to a negative value, every negative value of $E(\frac{dS_i}{S_i})$ needs to be multiplied by minus one, first. The green line corresponds to the average of the largest balance value of those blue points and the smallest balance value of those red points. Note: The starting time is 2016-01-23, and the time interval (Δt) is 112 days (almost four months).

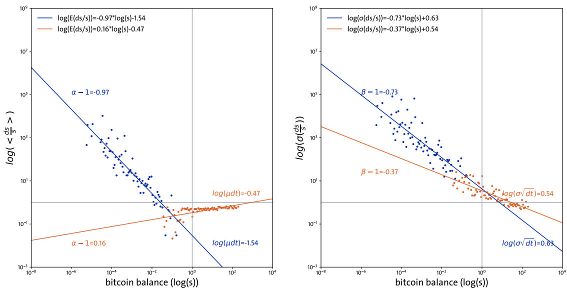
As Fig. 6 shows, surprisingly, there are two different modes (blue points and red points) for the growth of users' Bitcoin balance. For users whose bitcoin balance is less than a specific value (blue points), the average of $\frac{dS_i}{S_i}$ is positive, namely $\mu > 0$ in equation 3. The left panel of Fig. 6 shows that there is a linear relationship between $E(\frac{dS_i}{S_i})$ and the bin value s, and the slope of this linear line is negative which means that $\alpha < 1$ for those blue points. At the same time, the right panel of Fig. 6 shows that there is also a negatively correlated relationship between $\sigma(\frac{dS_i}{S_i})$ and S_i, which denotes again $\alpha < 1$ for those blue points. By contrast, for users whose bitcoin balance is larger than a specific value (red points), the average of $\frac{dS_i}{S_i}$ ($E(\frac{dS_i}{S_i})$) is negative, which means that $\mu < 0$ in equation 3 for red points. The line in the left panel of Fig. 6 is almost horizontal, but upward a bit actually, which means that α is larger than or close to one for red points. However, the right panel of Fig. 6 shows that the linear line is not exactly horizontal for red points, which means that $\alpha < 1$ for the volatility term. These analyses show that there exist two different balance growth models for Bitcoin users, which are as follows:

$$\begin{cases} dS_i = {S_i}^{\alpha^{11}_{<1}} \cdot \mu_{>0} \cdot dt + {S_i}^{\alpha^{12}_{<1}} \cdot \sigma dw_i if S_i < S^* \\ dS_i = {S_i}^{\alpha^{21}_{>1}} \cdot \mu_{<0} \cdot dt + {S_i}^{\alpha^{22}_{<1}} \cdot \sigma dw_i if S_i > S^*, \end{cases} \tag{4}$$

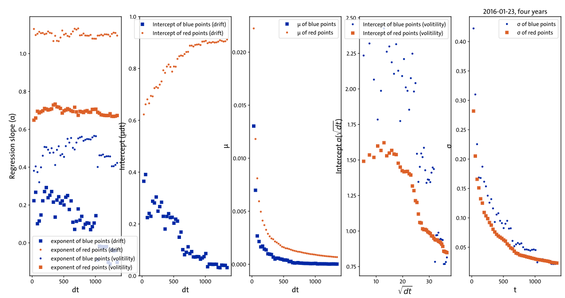
where subscript < 1, $= 1$, > 0, and < 0 denote that the corresponding value is smaller than one, equal to one, larger than zero, and smaller than zero, respectively. For example, $\alpha^{11}_{<1} < 1$, $\mu_{>0} > 0$. S^* is a threshold value. That means that for users whose bitcoin balance value is smaller than S^*, their balances grow according to the first model of equation 4. Because the corresponding exponent $\alpha < 1$ and $\mu > 0$, we can't get an analytical solution for this model. So, we can't predict how the Bitcoin balance of these users will change on average. For users whose bitcoin balance value is larger than S^*, their balances grow according to the second model of equation 4.

We now research whether this type of growth model is stable by changing the time interval dt. Our researching target include the exponent α, drift parameter μ, volatility parameter σ. For every time interval dt, we calculated the average ($E(\frac{dS_i}{S_i})$) and the variance ($\sigma(\frac{dS_i}{S_i})$) of $\frac{dS_i}{S_i}$ in each bin of bitcoin balance. Then, we got these target parameters (α, μ, σ) by making a linear regression between the average ($E(\frac{dS_i}{S_i})$), variance ($\sigma(\frac{dS_i}{S_i})$) and balance S, respectively.

As shown in panel (b) of Fig. 7, for the first stochastic equation ($dS_i = {S_i}^{\alpha^{11}_{<1}} \cdot \mu_{>0} \cdot dt + {S_i}^{\alpha^{12}_{<1}} \cdot \sigma dw_i if S_i < S^*$) in equation 4, the value of exponent α in both drift term and volatility term fluctuate a lot, both value of exponent α are less than 1 and are different from each other. The exponent α in the drift term is negatively correlated with the time interval (dt), while the exponent α in the volatility term seems constant despite of much fluctuation. The second figure in panel (b) of Fig. 7 shows that the drift term parameter μ is a monotonically decreased function with time interval (dt). The third figure in panel (b) denotes that $\sigma \cdot \sqrt{dt}$ fluctuates a lot with time interval (dt) but is negatively correlated with time interval (dt). So, the parameter σ in the volatility term of the first equation of equation 4 should be a monotonically decreased function of time interval (dt). By analyzing, the exact formula of the equation should be:



(a) Parameter fitting



(b) Parameter with dt

Fig. 7: Panel (a) depicts the relationship between exponent α in two different equations, two different terms (drift and volatility), and time interval dt. Panel (b) depicts the relationship between the intercept term ($\mu \cdot dt$) of drift term in two different equations, the intercept term ($\sigma \cdot \sqrt{dt}$) of volatility term in two different equations and time interval dt. Note: The time interval (dt) changes from 1 month to 24 months. The unit of the x-axis in panel (b) is day.

$$dS_i = S_i^{\alpha(dt)^{+11}_{<1}} \cdot \mu(dt)^{-}{}_{>0} \cdot dt + S_i^{\alpha^{12}_{<1}} \cdot \sigma(dt)^{-} dw_i if S_i < S^*; \tag{5}$$

where $\alpha(dt)^{+11}_{<1}$ denotes that exponent α in drift term is a monotonically increased function with time interval (dt) and smaller than 1; $\mu(dt)^{-}{}_{>0}$ denotes that μ is a monotonically decreased function with time interval (dt) and larger than zero; $\alpha^{12}_{<1}$ denotes that exponent α in volatility term is constant and smaller than 1; $\sigma(dt)^{-}$ denotes that σ is a monotonically decreased function with time interval (dt).

Now, we focus on analyzing the second formula of equation 4 ($dS_i = S_i^{\alpha^{21}_{>1}} \cdot \mu_{<0} \cdot dt + S_i^{\alpha^{22}_{<1}} \cdot \sigma dw_i if S_i > S^*$), where $\mu < 0$. When $\mu < 0$, we let $\mu \cdot dt$ multiply -1 so that we can take a log to it. As shown in panel (b) of Fig. 6, the value of exponent α in both drift term and volatility term is nearly constant with time interval. However, the exponent α in the drift term is larger than one and different from that in the volatility term which is smaller than one. The third figure in panel (b) of Fig. 6 shows that the drift term parameter $\mu \cdot dt$ is positive and increases with dt. After multiplying -1, $\mu \cdot dt$ is negative and decreases with dt, which means that users who own lots of bitcoins have the trend to sell their bitcoins with time flying.

The fourth figure in panel (b) of Fig. 6 denotes that $\sigma \cdot \sqrt{dt}$ decreases with $\sqrt{dt}$, so the parameter σ in volatility term of the second equation in equation 4 should be some monotonically decreased function of time interval dt. By analyzing, the exact formula of the equation it should be:

$$dS_i = S_i^{\alpha^{21}_{>1}} \cdot \mu_{<0} \cdot dt + S_i^{\alpha^{22}_{<1}} \cdot \sigma(dt)^{-} dw_i if S_i > S^*, \tag{6}$$

where $\sigma(dt)^{-}$ denotes that σ is a monotonically decreased function with time interval dt. Other parameters (α, μ) in this equation are constant. The model will be:

$$\begin{cases} dS_i = S_i^{\alpha(dt)^{+11}_{<1}} \cdot \mu(dt)^{-}{}_{>0} \cdot dt + S_i^{\alpha^{12}_{<1}} \cdot \sigma(dt)^{-} dw_i if S_i < S^* \\ dS_i = S_i^{\alpha^{21}_{>1}} \cdot \mu_{<0} \cdot dt + S_i^{\alpha^{22}_{<1}} \cdot \sigma(dt)^{-} dw_i if S_i > S^*. \end{cases} \tag{7}$$

IV. Think Oppositely

Are the exponents of S in equation 1 one for users whose balances are large enough? We can also explore this question from the opposite way, namely, we assume that the exponent α is one now and we want to find facts that are not consistent with our assumption if it is not equal to one.

If exponent α is one, then equation 1 would be written as:

$$dS_i = S_i \cdot \mu \cdot dt + S_i \cdot \sigma \cdot dw_i. \tag{8}$$

Namely, $\frac{dS_i}{S_i} = \mu \cdot dt + \sigma \cdot dw_i$. Because of $dw_i \sim N(0, dt)$, then $\frac{dS_i}{S_i}$ will follow the same distribution for all users and not be related to users' initial wealth S_i.

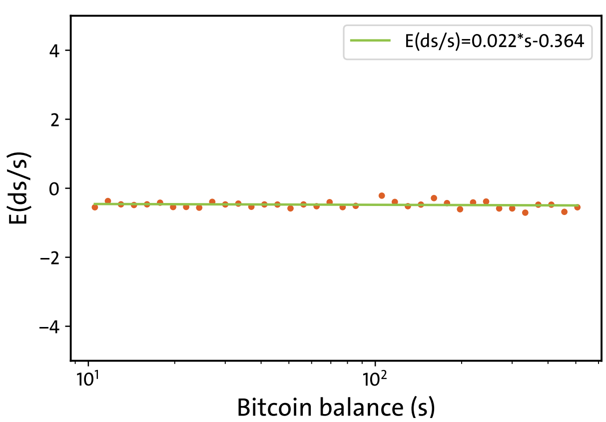


Fig. 8: Users' bitcoin balance versus balance change divided by bitcoin balance. Users include those whose balances are larger than 10 bitcoins and who have transactions during dt. If exponent α is one, there will no correlation between $\frac{dS_i}{S_i}$ and S_i.

As shown in Fig. 8, there is a positive relationship between $\frac{<dS_i>}{S_i}$ and S_i, and the slope is very small. Though it is weak, it can still help to make us confirm that the exponent α is a bit larger than one in equation 2, which is consistent with the result in the first figure of the right panel of Fig.7 (red circle points). We also apply this method to research the relationship between $\sigma(dS_i)$ and S_i and get Fig. 9.

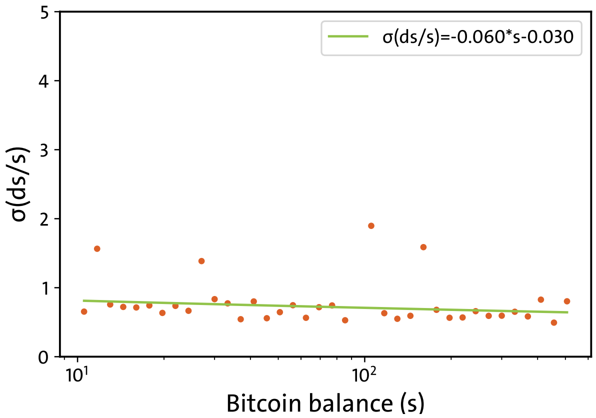


Fig. 9: Users' bitcoin balance versus standard variance of balance change divided by bitcoin balance. Users include those whose balances are larger than 10 bitcoins and who have transactions during dt. If exponent α is one, there will be no correlation between $\frac{\sigma(dS_i)}{S_i}$ and S_i.

As shown in scatter Fig. 9, $\frac{\sigma(dS_i)}{S_i}$ is negatively correlated to S_i, which denotes that the exponent of S in the volatility term is a bit less than one when S_i is large enough (it is 10 bitcoins in our case). This result is also consistent with the result in the first figure of the right panel of Fig. 7 (red square points).

V. Summary and Discussion

In this paper, we explore the transaction mechanism of bitcoin users. Firstly, we explored users' balance distribution and found that the log-normal distribution function can better fit the balance. Secondly, we explored whether bitcoin users' transaction behavior follows Gibrat's proportional growth rule and found that their transaction behaviors didn't follow Gibrat's proportional growth rule. During this part, we also find that there exist two kinds of different users, users who have many bitcoins in the beginning but tend to sell their bitcoins, and users who have few bitcoins in the beginning but tend to buy more bitcoins in the future.

By analyzing the balance data from users who owned more bitcoins in the beginning, we found that the exponent of S in the drift term is almost constant and slightly larger than one, and the exponent of S in the volatility term is also almost constant and slightly smaller than one, which was shown in the second equation in equation 7. The analysis in Section IV also confirms this analysis. The third and fifth figure in the right panel of Fig. 6 denotes that μ and σ are not constant, but decrease with dt. By analyzing the balance data from users who owned few bitcoins in the beginning, we found that the exponents of S in the drift term and volatility term are smaller than one and are not constant. μ is positive which means that the users in this group tend to buy bitcoins in the transaction.

Though we tried to explore the behavior modes in bitcoin transactions and get some useful insights, we didn't find a satisfying model that can model the transaction behavior in Bitcoin. The mechanism of transaction behavior is still a significant research topic and deserves more attention in the future.

References

[1] PL Krapivsky, S Redner, F Leyvraz, Connectivity of growing random networks, Physical review letters 85 (21), 4629 (2000).

[2] PL Krapivsky, S Redner, Organization of growing random networks, Physical Review E 63 (6), 066123 (2001).

[3] Alstott J, Bullmore E, Plenz D (2014) powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions. PLoS ONE 9(1): e85777. https://doi.org/10.1371/journal.pone.0085777.

[4] Clauset, Aaron, et al. "Power-Law Distributions in Empirical Data." SIAM Review, vol. 51, no. 4, Society for Industrial and Applied Mathematics, 2009, pp. 661–703, http://www.jstor.org/stable/25662336.

[5] Sheridan, P., Onodera, T. A Preferential Attachment Paradox: How Preferential Attachment Combines with Growth to Produce Networks with Log-normal In-degree Distributions. Sci Rep 8, 2811 (2018). https://doi.org/10.1038/s41598-018-21133-2

[6] Aspembitova A, Feng L, Melnikov V, Chew LY (2019) Fitness preferential attachment as a driving mechanism in bitcoin transaction network. PLoS ONE 14(8): e0219346. https://doi.org/10.1371/journal.pone.0219346

[7] Kondor D, Po sfai M, Csabai I, Vattay G (2014) Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network. PLoS ONE 9(2): e86197. doi:10.1371/journal.pone.0086197

[8] Maillart T, Sornette D, Spaeth S, von Krogh G. Empirical tests of Zipf's law mechanism in open source Linux distribution. Phys Rev Lett. 2008 Nov 21;101(21):218701. doi: 10.1103/PhysRevLett.101.218701. Epub 2008 Nov 19. PMID: 19113459.

[9] Malevergne Y, Pisarenko V, Sornette D. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. Phys Rev E Stat Nonlin Soft Matter Phys. 2011 Mar;83(3 Pt 2):036111. doi: 10.1103/PhysRevE.83.036111. Epub 2011 Mar 22. PMID: 21517562.

[10] del Castillo, Joan, Pedro Puig. "The Best Test of Exponentiality against Singly Truncated Normal Alternatives." Journal of the American Statistical Association, vol. 94, no. 446, [American Statistical Association, Taylor & Francis, Ltd.], 1999, pp. 529–32, https://doi.org/10.2307/2670173.

[11] Gatto, R. and Rao Jammalamadaka, S. (2002). A saddlepoint approximation for testing exponentiality against some increasing failure rate alternatives. Statistics Probability Letters (Vol. 58).

[12] Didier Sornette, Rama Cont. Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws. Journal de Physique I, EDP Sciences, 1997, 7 (3), pp.431-444. ff10.1051/jp1:1997169ff. ffjpa-00247337f

[13] A. Saichev and D. Sornette, Zipfs law and maximum sustainable growth, Journal of Economic Dynamics and Control 37 (6), 1195-1212 (2013).