Behaviour Fingerprint of the Bitcoin Market

Anqi Liu, Hossein Jahanshahloo, Jing Chen*, Arman Eshraghi

Cardiff University

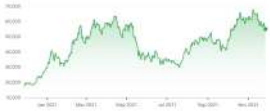
June 3, 2025

Overview

A currency, or an investment asset?

- As a crypto**currency** by design, the core function of Bitcoin should be, like any conventional currency, the purchasing power.
  - Broad categories of commodities;
  - Well managed by monetary and fiscal policy;
  - Strong regulations (e.g. taxes, foreign exchange controls, banking).
- However, there have been limited venues and sources that allow people to purchase goods and services using Bitcoin for various reasons.
- Also, the USD/BTC "exchange rate" is much more volatile than foreign exchange rates.

A currency, or an investment asset?



(a) USD/BTC

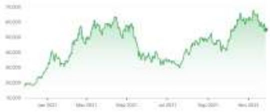
(b) GBP/USD

(c) MXN/USD

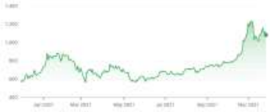
	Min.	Max.	Mean	Median	Std.	Skew.	Kurt.
USD/BTC	−14.81%	17.18%	0.32%	0.34%	4.22%	−0.13	4.48
GBP/USD	−1.37%	1.26%	0.00%	0.00%	0.44%	−0.19	3.29
MXN/USD	−2.58%	1.74%	0.00%	0.01%	0.67%	−0.30	3.54

Table: Return Statistics

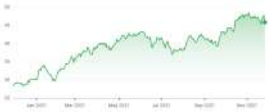
A currency, or an investment asset?



(a) USD/BTC



(b) TSLA



(c) BAC

	Min.	Max.	Mean	Median	Std.	Skew.	Kurt.
USD/BTC	−14.81%	17.18%	0.32%	0.34%	4.22%	−0.13	4.48
TSLA	−12.77%	17.93%	0.24%	0.24%	3.42%	0.34	6.49
BAC	−4.47%	6.06%	0.19%	0.12%	1.60%	0.00	3.38

Table: Return Statistics

Bitcoin as an investment asset?

- Most traders consider and use Bitcoins as a speculative asset.
  - High return;
  - High risk;
  - Fat-tails.
- However, the bitcoin market structure is not designed to make it similar to a stock market.
  - Average 10 minutes to mine a block vs. Microsecond-level response time;
  - 24/7 vs. Market closes overnight.
- Do the Bitcoin traders act just as stock traders?

Addresses in The Bitcoin Market

- A significant increase in the number of addresses since the beginning of 2018 – from $353,131,255$ addresses to $776,059,991$ addresses by the end of January 2021.
- Over 91.22% ($707,902,731$ out of $776,059,989$) addresses have no more than 2 transactions in total; majority of them have been inactive for years.
- The most active address (1HckjUpRGcrrRAtFaa-CAUaGjsPx9oYmLaZ) appears on 28 Oct 2017 (block #492078) and maintains an average $87,123.88$ transactions per month with in total $3,484,955$ transactions.

Bitcoin Market Microstructure

The literature grows very fast, in both technology and finance fields. We only briefly state a few literature that inspire us to do this study from the angle of market microstructure and you will see why so while we go along.

- Manahove and Urquhart (2021) found evidences that the Bitcoin market is populated by HFTs alike and substantial trading volumes come from them.

- The intraday price dynamics documented in Eross et al. (2019) have been used to develop trading strategies, e.g.
  - hedging tools (Urquhart and Zhang (2019));
  - portfolio formation (Platanakis and Urquhart (2019, 2020)).

- Non-HFTs Bitcoin trading involves strategies with "currency pairs". Duan et al. (2021) analysed various currency-against-Bitcoin price pairs and verified cross-market statistical arbitrage opportunities.

Motivation

- We find the Bitcoin market has many technical advantages in attracting investors, especially non-sophisticated ones. For example, easy entry, low costs, high return potential.
- With expansion of the Bitcoin market since 2017, we observe increasing pricing volatility and high speculative profitability in this innovative "financial asset".
- The lack of regulation and narrow understanding about this innovative system and fast growing trader community also cause irregularity in prices, liquidity and volume.

Motivation

- In the blockchain system, we can trace the "coinpath" of every Bitcoin, which tells the sent/received address-ids of every coin.
- The transparency of Bitcoin market data allows us to get a closer look at the trader behaviors than any other traditional, high-regulated financial markets.
  - A large population of "taster traders" in the Bitcoin market (more than 90%), which is uncommon for the traditional markets.
  - **Objective:** Whether there are similar trader types or compatible trading behaviours in the Bitcoin market as those in the traditional markets (e.g. fundamental traders, HFTs, market makers in stock markets).
- Further analysis about their impacts to market microstructure, price formation, etc. More importantly, our findings would be useful for regulators to govern the blockchain-based markets (e.g. indicators for price manipulation, cyber-crimes).

Key Findings

- We “learn” and identify five distinctive trader types, which are compared to the conventional trader categorizations: fundamental trader, technical traders, and liquidity providers.
- We analyse group-wide statistics and transaction time series analysis of each groups help establish a close comparison to the well-known conventional trader behaviours.
- We find interesting cases that clearly shows market making, technical trading, and fundamental investment like strategies by zooming into trading timeline of specific traders in each groups.

Bitcoin Trader Dataset

- Execution data in the bitcoin market includes
  - address ids of sent and received transactions;
  - block time of each transactions;
  - bitcoins sent and received in each transaction.

- We denote the time horizon of the data is from D_s to D_e, in total including
  - M blocks
  - N transactions
  - n addresses

- We create an address-level dataset using the execution data for trader classifications.

Bitcoin Trader Dataset

- $\{N^{(i)} : i = 1, 2, ..., n\}$ the number of transactions for address i.
- $\{M^{(i)} : i = 1, 2, ..., n\}$ the number of blocks that address i had transactions.

- $\{D_j^{(i)} : j = 1, 2, ..., N^{(i)}; i = 1, 2, ..., n\}$ the block date of jth transaction for address i.
- $\{T_j^{(i)} : j = 1, 2, ..., N^{(i)}; i = 1, 2, ..., n\}$ the block time of jth transaction for address i.

- $\{\Lambda_{D,k}^{(i)} : D_s \leq k \leq D_e; i = 1, 2, ..., n\}$ number of transactions of address i in day k.
- $\{\Lambda_{W,k}^{(i)} : D_s \leq k \leq D_e; i = 1, 2, ..., n\}$ number of transactions of address i in week k.
- $\{\Lambda_{M,k}^{(i)} : D_s \leq k \leq D_e; i = 1, 2, ..., n\}$ number of transactions of address i in month k.

Bitcoin Trader Features Summary

	Feature	Equation
1:	Number of transactions per block (in logarithm scale).	$CntTransB_i = \frac{N^{(i)}}{M^{(i)}}$
2:	Average number of transactions per day (excluding zero transaction days).	$CntTransD_i = \frac{N^{(i)}}{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{D,k})}$
3:	Average number of transactions per week (excluding zero transaction weeks).	$CntTransW_i = \frac{N^{(i)}}{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{W,k})}$
4:	Relative standard deviation of the number of transactions per week (excluding transaction weeks).	$StdCntTransW_i = \frac{\sqrt{\frac{\sum_k \left(\Lambda^{(i)}_{W,k} - CntTransW_i\right)^2}{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{W,k})}}}{CntTransW_i}$
5:	Median of transaction time intervals (in weeks).	$MedIntv_i = \frac{\mathrm{Med}_j(\Delta T^{(i)}_j)}{60 \cdot 24 \cdot 7}$
6:	Average transaction time intervals (in weeks).	$MeanIntv_i = \frac{1}{N^{(i)}-1} \sum_{j=1}^{N^{(i)}-1} \frac{\Delta T^{(i)}_j}{60 \cdot 24 \cdot 7}$
7:	Relative range of transaction intervals.	$RngIntv_i = \frac{\max_j \Delta T^{(i)}_j - \min_j \Delta T^{(i)}_j}{MeanIntv_i}$
8:	Relative standard deviation of transaction intervals.	$StdIntv_i = \frac{\sqrt{\frac{\sum_j \left(\Delta T^{(i)}_j - MeanIntv_i\right)^2}{N^{(i)}-1}}}{MeanIntv_i}$
9:	Active time ratio.	$ActTime2L_i = 100 \times \frac{D^{(i)}_{N^{(i)}} - D^{(i)}_1}{H^{(i)}}$
10:	Transaction days to life time ratio.	$TransD2L_i = 100 \times \frac{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{D,k})}{H^{(i)}}$
11:	Transaction weeks to life time ratio.	$TransW2L_i = 100 \times \frac{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{W,k})}{H^{(i)}/7}$
12:	Transaction months to life time ratio.	$TransM2L_i = 100 \times \frac{\sum_k \mathbf{1}_{>0}(\Lambda^{(i)}_{M,k})}{H^{(i)}/30}$

Table 1: Definition of address-level trading features.

Bitcoin Trader Features Quick Grasp

1. Trading frequency features (1- 4): different transaction counts and standard deviation;
2. Waiting time features (5-8): different transaction intervals to reflect consistency of their trading habits;
3. Active trading ratios features (9-12): different life time ratio to reflect the general activeness in this space.

The K-Means Algorithm

- Unsupervised, K clusters.

- Distance

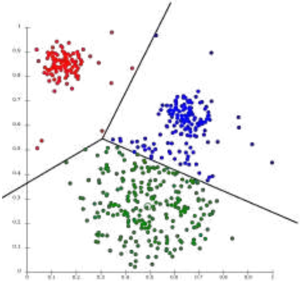
$$\mathcal{D}(X, Y) = \sqrt{\sum_{k=1}^{m} (x_k - y_k)^2},$$

where $(x_1, x_2, ..., x_m)$ and $(y_1, y_2, ..., y_m)$ are the feature vectors of X and Y respectively.

- Clustering centroid

$$O_k = \frac{1}{\sum_i \mathbf{1}_{\mathcal{C}_k}(i)} \sum_i \mathbf{1}_{\mathcal{C}_k}(X_i) X_i, \; k = 1, 2, ..., K$$

The K-Means Algorithm



Algorithm 1: K-means clustering algorithm

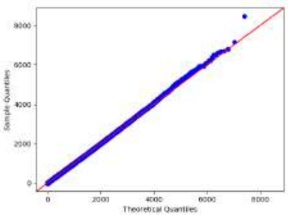
Data: $X = \{X_i : i = 1, 2, ..., n\}$
Result: $\mathcal{C} = \{C_k : k = 1, 2, ..., K\}$
Input : The number of classes K, the termination condition $\epsilon = 1e-8$
Output: $\{\mathbf{1}_{C_k}(i) : i = 1, 2, ..., n \text{ and } k = 1, 2, ..., K\}$

$e \leftarrow 1e8$ // initialize a large distance between centroids

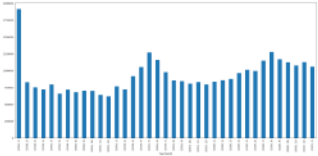
for $i \leftarrow 1$ **to** n **by** 1 **do**
  $I \leftarrow \mathrm{RandInt}(1, K)$ // randomly select a class
  $$\mathbf{1}_{C_j}(i) = \begin{cases} 0, & \text{if } j = I \\ 1, & \text{otherwise} \end{cases}, j = 1, 2, ..., K$$
end
/* iterate over all classes */
for $k \leftarrow 1$ **to** K **by** 1 **do**
  $O_k = \frac{1}{\sum_i \mathbf{1}_{C_k}(i)} \sum_{X_i \in C_k} X_i$
end
/* loop until the centroids of newly developed clusters stay the same */
while $e \geq \epsilon$ **do**
  for $i \leftarrow 1$ **to** n **by** 1 **do**
    /* iterate over all observations */
    $\tilde{I} = \arg\min_k \mathcal{D}(X_i, O_k)$
    // update to the closest class
    $$\mathbf{1}_{C_j}(i) = \begin{cases} 0, & \text{if } j = \tilde{I} \\ 1, & \text{otherwise} \end{cases}, j = 1, 2, ..., K$$
  end
  /* update all class centroids */
  for $k \leftarrow 1$ **to** K **by** 1 **do**
    $\tilde{O}_k = \frac{1}{\sum_i \mathbf{1}_{C_k}(i)} \sum_i \mathbf{1}_{C_k}(i) X_i$
  end
  $e = \mathcal{D}(O_k, \tilde{O}_k)$ // update the distance between centroids
  $O = \tilde{O}$ // update centroids
end

Data

- Address-level dataset: all accounts joined the Bitcoin market before 2021-01-31 and traded during 2018-01-01 to 2021-01-31.
- Exclude the addresses that had less than 3 trades or shorter than 1 minute active time or traded in the market in less than 3 different weeks.
- $6,108,128$ addresses are used in the classification experiment.



(a) QQ plot: block arrivals vs. exponential distribution (intensity=9.33 minutes).



(b) New trader arrival.

Five Trader Classes (FINAL)

- Class #1 “taster traders”;
- Class #2 “fundamental traders”;
- Class #3 “technical traders”;
- Class #4 “market makers”;
- Class #5 “high-frequency trader”.

Five Trader Classes (FINAL)

- $5,215,499$ (out of $6,108,128$) addresses have the same classification throughout all runs.
- Take the consistent clustering results $\longrightarrow$ centroids.

Class	Num. of addresses	*CntTransB*	*CntTransD*	*CntTransW*	*StdTransW*
1	2,089,694	1.07	2.00	4.28	0.53
2	2,101,244	1.04	1.71	3.01	0.49
3	1,023,726	1.05	1.85	4.66	0.57
4	566	2.08	70.38	389.43	0.82
5	269	13.95	560.35	2569.68	0.85

Table: Trading frequency features

Five Trader Classes (FINAL)

- 5,215,499 (out of 6,108,128) addresses have the same classification throughout all runs.
- Take the consistent clustering results $\longrightarrow$ centroids.

Class	Num. of addresses	*MedIntv* (days)	*MeanIntv* (days)	*RngIntv*	*StdIntv*
1	2,089,694	1.57	7.27	8.92	1.87
2	2,101,244	3.21	22.36	13.72	2.33
3	1,023,726	2.24	4.86	13.04	1.74
4	566	0.02	0.14	1893.16	27.01
5	269	0.01	0.11	366.16	8.89

Table: Waiting time features

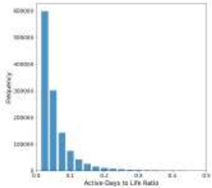
Five Trader Classes (FINAL)

- $5,215,499$ (out of $6,108,128$) addresses have the same classification throughout all runs.
- Take the consistent clustering results $\longrightarrow$ centroids.

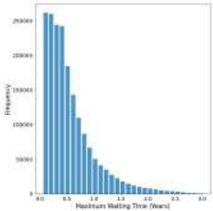
Class	Num. of addresses	*ActTime2L* (%)	*TransD2L* (%)	*TransW2L* (%)	*TransM2L* (%)
1	2,089,694	20.07	3.07	10.24	19.38
2	2,101,244	72.65	4.63	17.84	39.56
3	1,023,726	88.37	21.24	61.90	88.53
4	566	72.35	34.13	40.68	48.11
5	269	31.24	24.61	28.04	33.05

Table: Active trading features

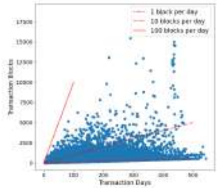
Fundamental traders trading specs.



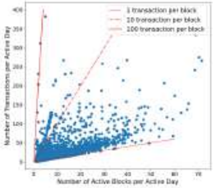
(a) Hist. of active days %.



(b) Hist. of maximum waiting time interval.

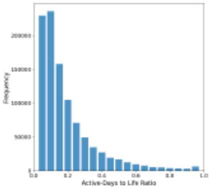


(c) Active blocks vs. Active days.

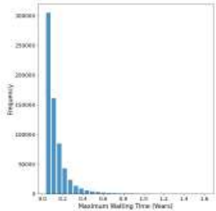


(d) Number of transactions vs. Active blocks.

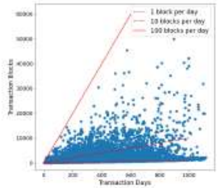
Technical traders trading specs.



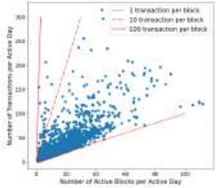
(a) Hist. of active days %.



(b) Hist. of maximum waiting time interval.

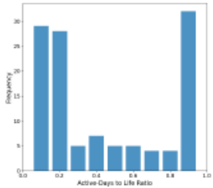


(c) Active blocks vs. Active days.

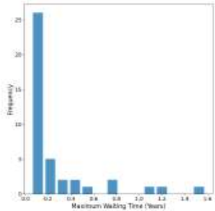


(d) Number of transactions vs. Active blocks.

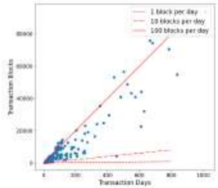
Market makers trading specs.



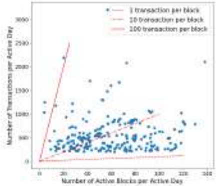
(a) Hist. of active days %.



(b) Hist. of maximum waiting time interval.

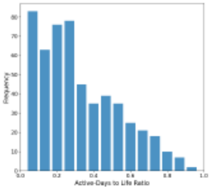


(c) Active blocks vs. Active days.

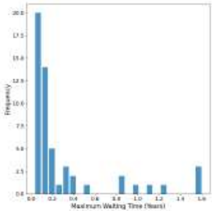


(d) Number of transactions vs. Active blocks.

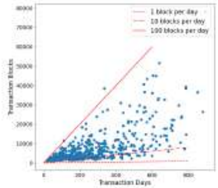
High-frequency traders trading specs.



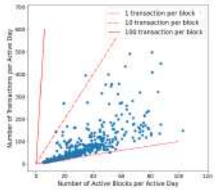
(a) Hist. of active days %.



(b) Hist. of maximum waiting time interval.



(c) Active blocks vs. Active days.



(d) Number of transactions vs. Active blocks.

A case of fundamental traders (in Class #2)

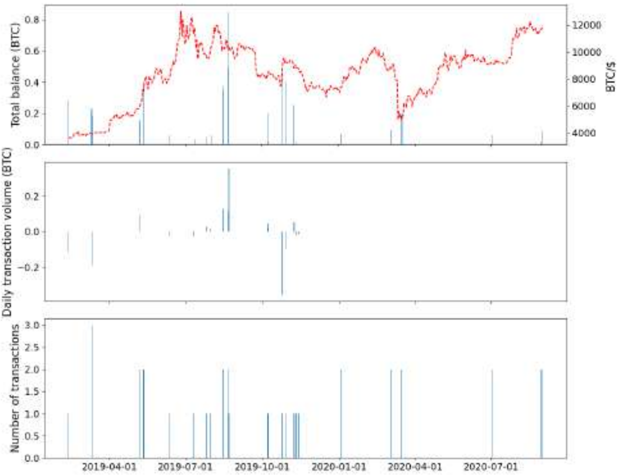


Figure: 1BKsv1JRKgap6ySprBrNDAiWS3wxz9EzJn

A case of technical traders (in Class #3)

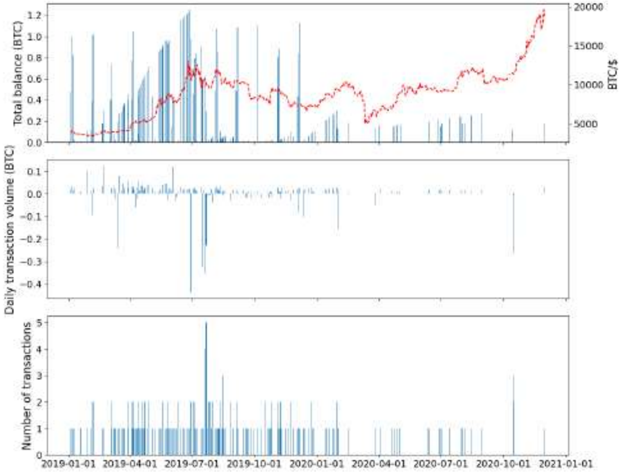


Figure: 3LxwRbT1nAWWdRK8CgNWTZqghubs6vftXg

A case of market makers (in Class #4)

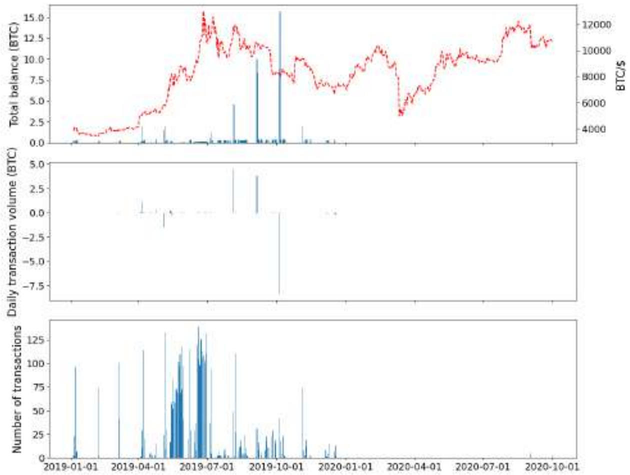


Figure: 138hQapNmwkRHtC6DTb7H4tQbjMmj49ej2

A case of high-frequency traders (in Class #5)

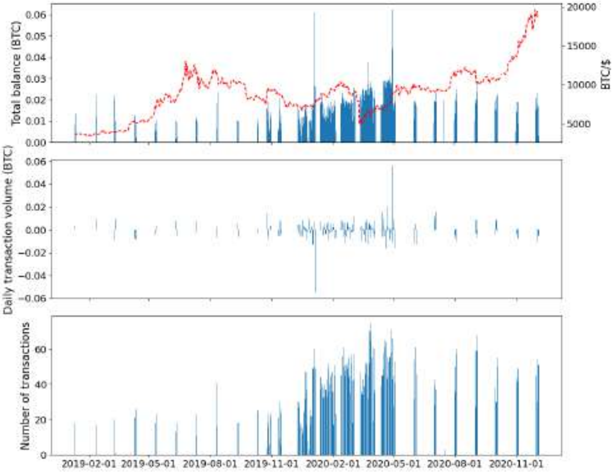


Figure: 3K1kmJbfHm135wfpxgcNEYRtvXjZ2AT7y8

Extended Trader Classification using SVM

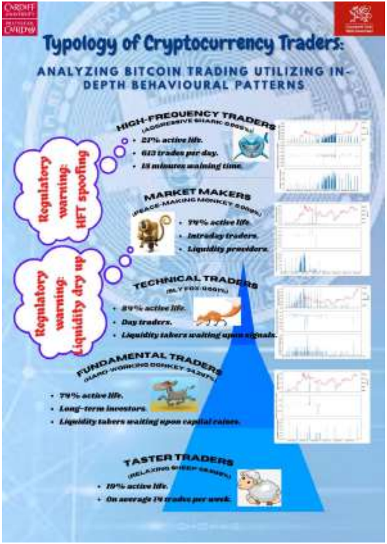


Figure: SVM Trader Classification

THANK YOU!!!