Appendices for: Trading volume manipulation and competition among centralized crypto exchanges

Dan Amiram, Evgeny Lyandres, and Daniel Rabetti

Appendix A : Methods for detecting anomalous trading: Detailed descriptions

Examples of anomalous trading

Figure A.1 presents examples of trading activity that illustrate various patterns of anomalous trading.

[Insert Figure A.1 here]

Panel A depicts the evolution of trading volume (in the upper two figures) and ten-minute-aggregated number of trades (in the lower two figures) in the OmiseGo-Bitcoin pair (OMG-BTC) during 4,464 ten-minute intervals in January 2019 on two exchanges—OKX and Binance. The upper two figures suggest that while trading volume on Binance follows an arguably random pattern, there seem to be several structural breaks in trading volume on OKX. Similar structural breaks are evident in the number of trades on OKX, whereas they do not seem to be present on Binance.

Panel B presents trading volume and number of trades in one of the most important cryptocurrency pairs—ETH-BTC—on two exchanges: ZB (on the left) and Binance (on the right) in March 2019. While the plots of both trading volume and number of trades on Binance do not reveal unusual patterns, there are numerous spikes of trading activity on ZB, which are often two orders of magnitude larger than the typical trading volume and occur with a roughly constant frequency.

Benford-Law-based anomalous trading measure

Construction of Benford-Law-based MAD

Benford-Law-Based MAD (mean absolute deviation) statistic is calculated as the sum of the absolute difference between the empirical frequency of each digit, from 1 to 9, and the theoretical frequency found in Benford's Law, divided by the number of leading digits used (9). The construction of our Benfords' distance measure (Benford (1938)) follows Cong, Li, Tang, and Yang (2022). However, we use a different conformity statistic. While Cong et al. (2022) compare the differences in first-significant-digit distribution and Benford's Law at a particular moment in time using Chi-squared statistic, we use the mean-absolute-distance (MAD) of the differences in first-significant-digit distribution and Benford's Law across time. The reason for our choice is that ours is a panel data; and MAD has been used in the context of Benford-Law-based measure within panel data (e.g., Amiram, Bozanic, and Rouen (2015)).

Illustrations of Benford-Law-based measures

Panel A of Figure A.2 shows examples of observed series of first digits of trading volume over the course of one month (March 2019) in the ETH-BTC pair on four crypto exchanges—Binance, ZB, OKX, and Bibox.

[Insert Figure A.2 here]

In each of the four figures, the solid line depicts the actual distribution of leading digits of trading volume, whereas the dashed line depicts theoretical, Benford-Law-based distribution. The observed series for Binance seems to roughly approximate the theoretical Benford's Law. However, the other plots indicate frequencies of leading digits that deviate substantially from Benford's Law. For instance, the plot for ZB shows that the frequencies of digits 7-9 are over twice as large as those predicted by Benford's Law; the same is true for digit 5 on OKX and digits 2 and 3 on Bibox. The deviations from Bedford's Law are even more striking in Panel B of Figure A.2, which presents examples of frequencies of first digits of ten-minute-aggregated number of trades series.

E-divisive with medians (EDM)-based anomalous trading measure

Construction of EDM measure

EDM recursively partitions a time-series and uses a permutation test to determine change points. This non-parametric method has at least three advantages over parametric ones. First, EDM can detect both “mean shift" (sudden change) and “ramping" (gradual change) in data series. This seems especially relevant in our case, as illustrated in Figure A.1. Second, EDM uses moving median as opposed to mean, and, as a result, is robust to the presence of anomalies, i.e. irregular large trades. Finally, the non-parametric nature of EDM implies that the model adapts to the underlying distribution in the data.

We use the Twitter BreakoutDetection package for R as an implementation of the EDM algorithm.[1,2] To distinguish a short-term tick up or down from a change in the underlying signal, this package decomposes a time-series into a series of segments of one of three types: (a) steady state: the-time series follows a fixed mean (with random noise around the mean); (b) mean shift: the time-series jumps directly from one steady state to another; and (c) ramp up/down: the time-series transitions linearly from one steady state to another, over a fixed period. The breakout algorithm then returns a list of breakout points, or in other words, the number of times these state transitions are detected.

Measurement of EDM-based anomalous trading

Figure A.3 presents two examples of the performance of EDM algorithm in detecting structural breaks in our data.

[Insert Figure A.3 here]

The first example, presented in Panel A, is number of trades in the OMG-BTC pair on the OKX exchange in January 2019. The series is characterised by a step-like shape. The EDM algorithm is able to correctly detect four structural breaks, highlighted by dashed vertical lines (where the timing of the breaks is identified by the algorithm). Importantly, due to its reliance on comparisons of medians (and not means), the algorithm also correctly ignores a temporary spike in trading around bin 3,800. The second example, depicted in Panel B, is trading volume in the ETH-BTC pair on Binance in May 2019. The EDM algorithm does not identify any breakouts in this case, consistent with no visible structural breaks in the trading volume plot.

Principal-components-based measure of anomalous trading

In constructing our principal components measures of anomalous trading, we are guided by Kaiser (1961), Kaiser (1970), Jollife (2002), Peres-Neto and Somers (2005), and Abdi and Williams (2010), who propose tests of the appropriateness of factor analysis based on the variance of the first principal

[1] See https://github.com/twitter/BreakoutDetection.

[2] The model is set at its standard form with the following parameters: method='multi', min.size = 6, beta=0.002, and degree=1. Adjusting beta down (up) makes the EMD estimate more (less) sensitive to breakouts in the distribution. For an example, see https://blog.twitter.com/engineering/en_us/a/2014/breakout-detection-in-the-wild We set beta at .002 following standard practice in the computer science literature (e.g., James and Matteson (2015), and James, Kejariwal, and Matteson (2016)).

components, correlations with underlying measures, and vector direction. All original measures of fake trade pass Bartlett's test of sphericity[3] and Kaiser-Meyer-Olkin test.[4]

Panel A of Figure A.4, which presents a biplot of the orthogonalization of MAD and EDM measures and their relation with the first two principal components for both volume-based and number-of-trades-based measures. The length of each vector (squared cosine) represents the respective variable in the first two principal components. The horizontal and vertical projections of a vector represent the variable in the first and second principal components, respectively. Panel B of the same figure, which presents correlations between the first and second principal components of MAD and EDM. Both Panel A and Panel B show that these correlations are quite high but far from perfect—ranging between 0.6 and 0.777, suggesting that there is additional information in principal-component-based measures.

[Insert Figure A.4 here]

Panel C of Figure A.4 presents fractions of variance explained by the first four principal components of the individual anomalous trading measures. Since the first principal component explains a significant fraction of variation in anomalous trading—56%, we concentrate on the first principal component as an aggregated measure of anomalous trading in the empirical analysis, in addition to MAD and EDM measures.

[3]Bartlett's sphericity test provides information about whether the correlations in the data are strong enough to use a dimension-reduction technique such as principal components. The test asks whether a correlation matrix is an identity matrix. If so, the variables are uncorrelated, and one cannot perform a principal components analysis to reduce the dimensionality of the data. More formally, Bartlett's sphericity test is a test of whether the data are a random sample from a multivariate normal distribution, the covariance matrix of which is a diagonal matrix. In our case, the test statistic is very high (38.16), leading to a rejection of the null hypothesis and supporting the use of a dimension-reduction technique (PCA).

[4]Kaiser-Meyer-Olkin (KMO) test is a measure of how suited the data are for factor analysis. The test measures sampling adequacy for each variable in the model and for the model as a whole. The statistic (ranging between zero and one) is a measure of the fraction of variance among variables that is common. The higher the proportion, the more suited the data are to factor analysis. As rule of thumb, factor analysis is appropriate for values between 0.8 and 1. In our sample, the statistic equals 0.871. This indicates significant overlap of information among our measures of anomalous volume supports the use of factor analysis.

Figure A.1. Examples of trading volume and number of trades series. The plot shows trading volume (in upper figures of Panels A and B) and number of trades (in the lower figures of Panels A and B), aggregated into 4,464 ten-minute bins throughout a one-month period. Panel A displays data for OMG-BTC pair on OKX (left panels) and Binance (right panels) in January 2019. Panel B displays data for ETH-BTC pair on ZB (left panels) and Binance (right panels) in March 2019.

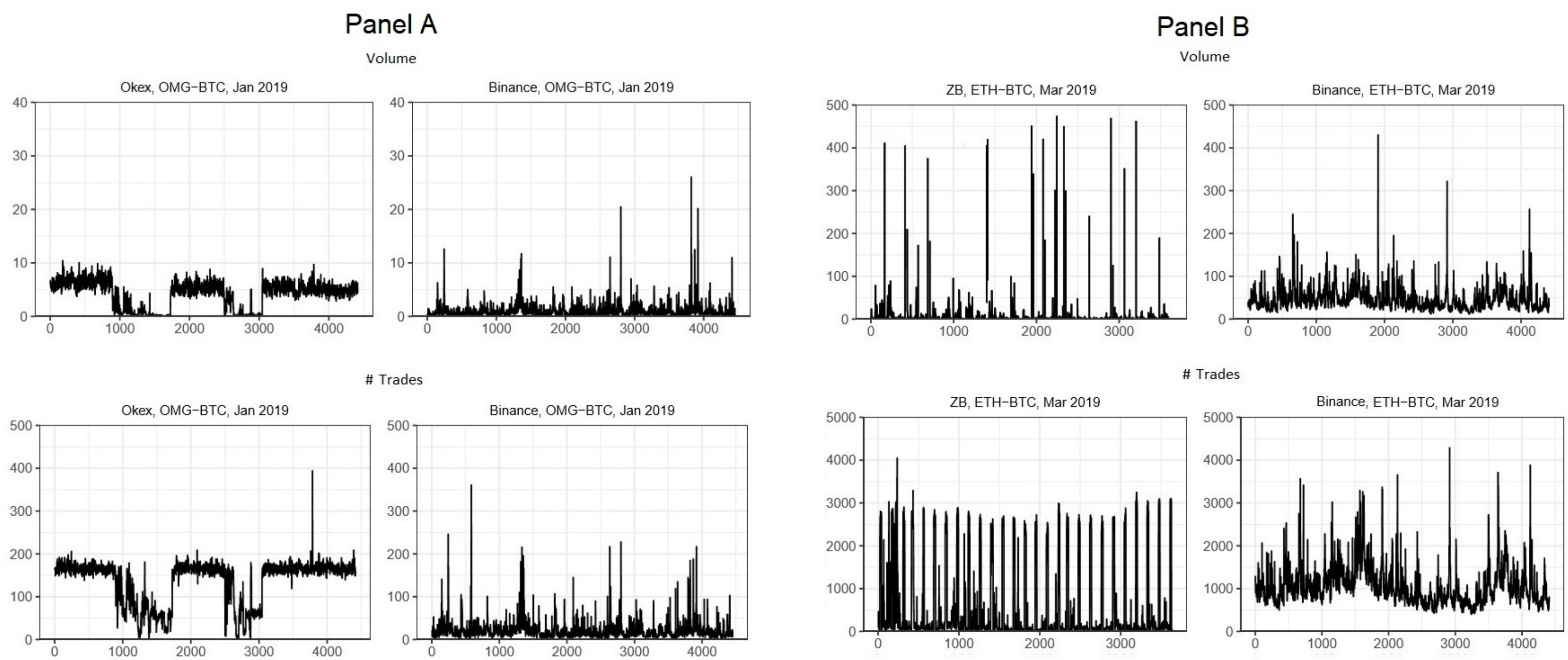


Figure A.2. Examples of deviations from Benford's Law. Blue solid curves represent empirical frequencies of leading digits of trading volume (in Panel A) and ten-minute aggregates of number of trades (in Panel B). Red dotted curves represent frequencies of leading digits under Benford's Law. The frequencies are computed for ETH-BTC pair in March 2019 on four exchanges: Binance, ZB, OKX, and Bibox.

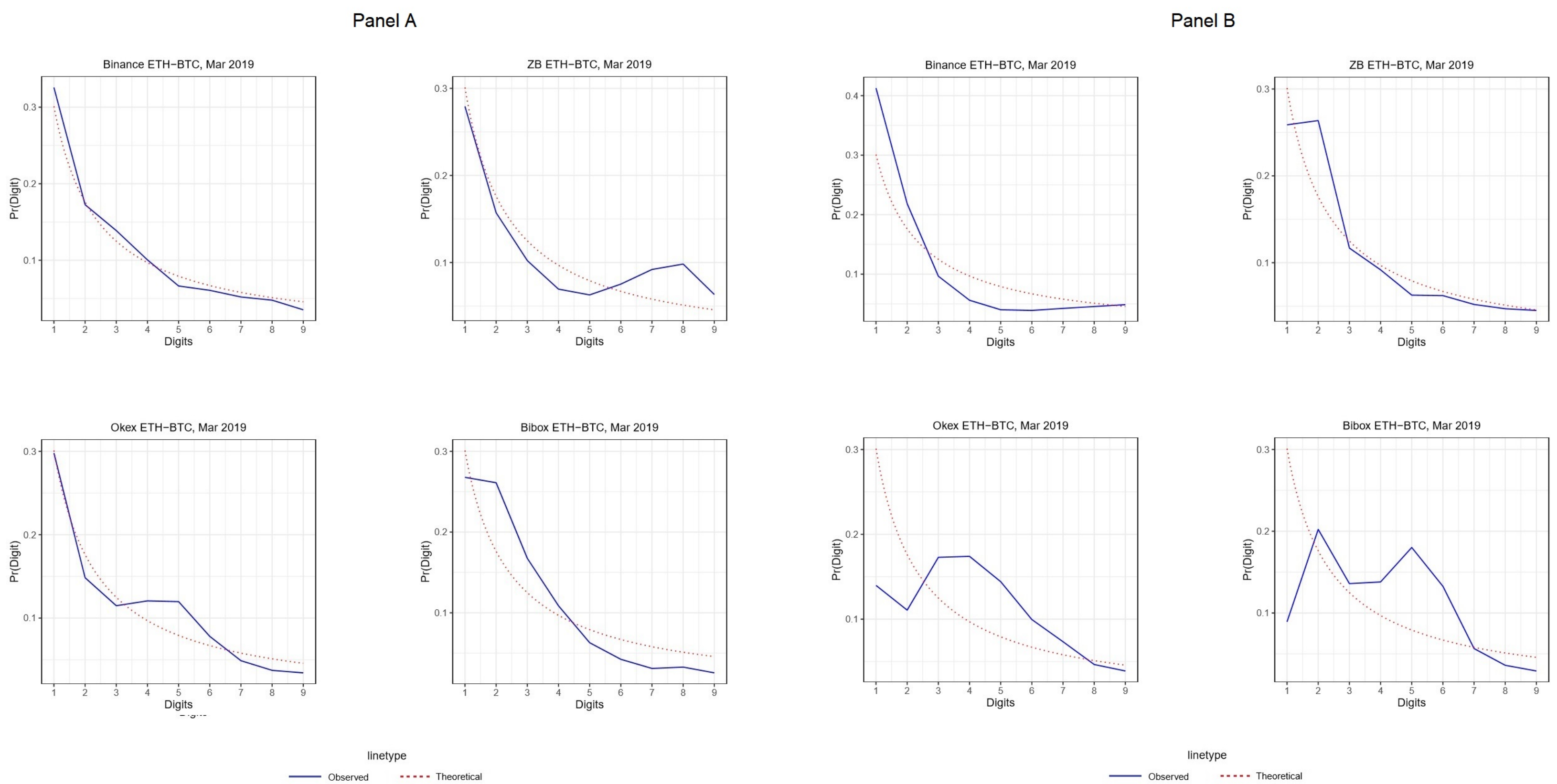
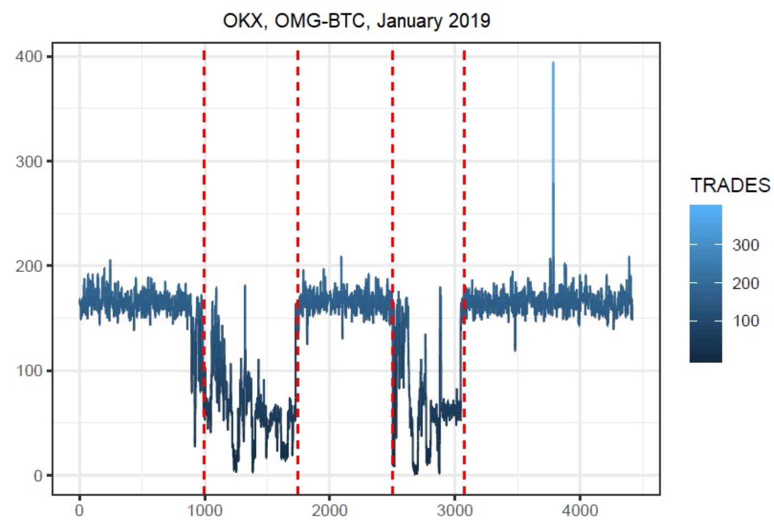


Figure A.3. Examples of application of EDM measure. The plots in both panels represent the number of trades series, aggregated into 4,464 ten-minutes bins. Panel A presents trading in OMG-BTC pair in January 2019 on OKX. Panel B presents trading in ETH-BTC pair on Binance in May 2019. Dashed red lines depicts structural breaks as detected by the EDM algorithm.

Panel A



Panel B

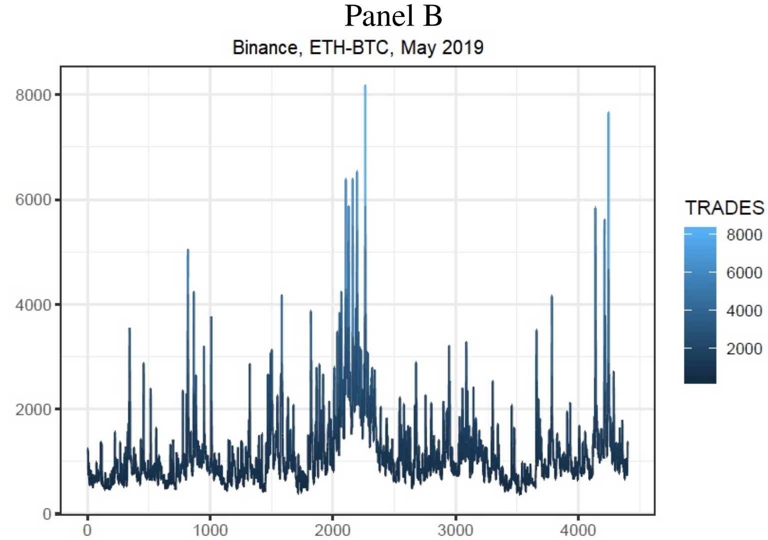
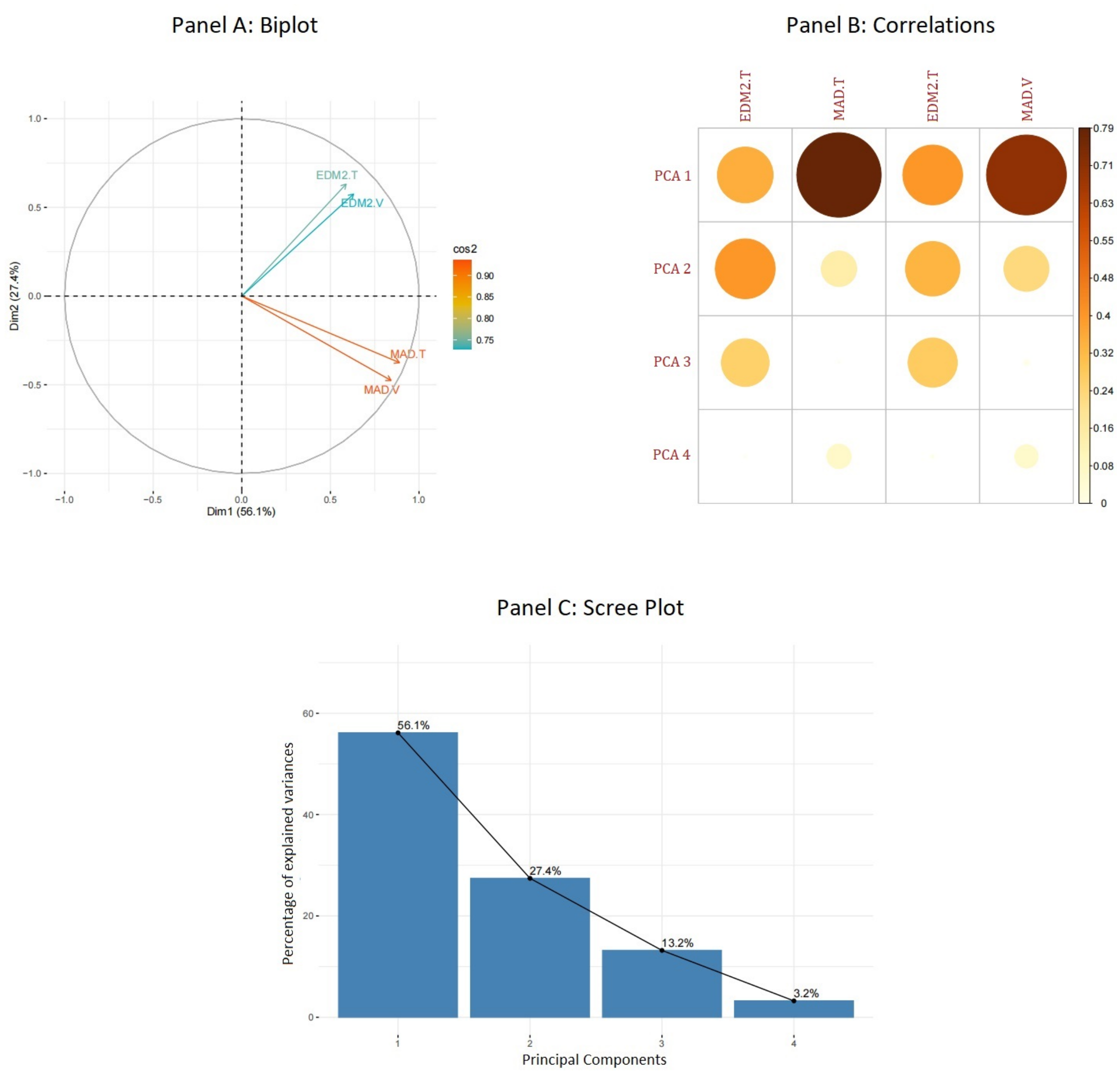


Figure A.4. Principal components results. The three panels show details on the derivation of the principal components measure. Panel A shows the biplot of the original fake measures, MAD and EDM. The extension .V stands for a measure based on volume series. The extension .T stands for a measure based on trades series. The vector direction and location show their correlations with the first two principal components of the combined measures. Panel B shows the correlations of the four original measures with the four dimensions of the derived principal components measure. The size of the circle indicates the contribution of the original measure to the dimension of the principal components. The color of the circle indicates the strength of the correlation. Panel C shows the scree plot of the four first dimensions of the principal components measure. The height of the bars indicates the size of the explained variation of that principal component dimension to the original measures of fake trading.



Appendix B: Variable definitions and additional empirical results

Table B.1. Variable description

Variable	Type	Description
Market		
CMC Listed: Coverage	Continuous (0, 1]	The monthly fraction of market capitalization of cryptocurrencies covered in our sample to the aggregate crypto market capitalization from www.CoinMarketCap.com
Currencies	Integer	Total number of cryptocurrencies (tokens and coins) listed on at least one exchange in a given month
Currencies: Entry	Integer	Total number of cryptocurrencies listed on at least one exchange in a given month and not listed on any exchange in the previous month
Currencies: Exit	Integer	Total number of cryptocurrencies not listed on any exchange in a given month and listed on at least one exchange in the previous month
Currency pairs	Integer	Total number of distinct currency pairs listed on at least one exchange in a given month
Currency pairs: Entry	Integer	Total number of currency pairs listed on at least one exchange in a given month and not listed on any exchange in the previous month
Currency pairs: Exit	Integer	Total number of currency pairs not listed on any exchange in a given month and listed on at least one exchange in the previous month
Volume: $U.S. (MM)	Continuous $(0, \infty)$	Aggregate reported volume of trading in all currency pairs on all exchanges in a given month in billions of $U.S.
# Trades (M)	Integer	Aggregate reported number of trades in all currency pairs on all exchanges in a given month in millions
Exchanges	Integer	Number of exchanges listing at least one currency pair in a given month
Exchanges: Entry	Integer	Number of exchanges listing at least one currency pair in a given month and not listing any currency pairs in the previous month
Exchanges: Exit	Integer	Number of exchanges not listing any currency pairs in a given month and listing at least one currency pair in the previous month
HHI Exchanges: # Currencies	Continuous (0, 1]	Sum of squared number of currencies listed on each exchange divided by the squared sum of number of currencies listed on all exchanges in a given month

Table B.1. Variable description – continued

Variable	Type	Description
Exchanges		
Age	Integer	The difference in months between current month and the first month any currency pair was listed on an exchange
Currency pairs	Integer	Number of currency pairs listed on an exchange in a given month
Currency pairs: Entry	Integer	Number of currency pairs that are listed on an exchange in a given month that were not listed on the exchange in the previous month
Currency pairs: Exit	Integer	Number of currency pairs that are not listed on an exchange in a given month that were listed on the exchange in the previous month
AML	Indicator	Equals one if an exchange implemented an AML policy and provides detailed information about conformity with accepted international AML procedures
KYC	Indicator	Equals one if there are evidence that the exchange provides clear guidelines, requires documents, and verifies sources of clients' funds
Crypto-friendly location	Indicator	Equals one if an exchange is located in Singapore, Russia, Estonia, Malta, Luxembourg or Switzerland
Traffic (MM)	Integer	Number of monthly visits (in billions) to an exchange's website, as reported by www.similarweb.com
Twitter retweets	Integer	Number of Twitter retweets of tweets by an exchange's official Twitter account in a given month
Github commits	Integer	The amount of code reviews (commits) at the exchange main Github repository in a given month
Africa, Asia (ex-China), China, Central and South America, North America, Eastern Europe, Western Europe (ex Islands), and Europe: Offshore	Binary	Indicator of hedquarters in a country belonging to a respective region
Currency pairs		
Listed on # exchanges:	Integer	The number of exchanges on which a currency pair is listed in a given month
log (Listed on # exchanges):	Continuous $(0, \infty)$	Natural logarithm of the number of exchanges on which a currency pair is listed in a given month
Age of listing on any exchange	Integer	Difference in months between current month and the first month a currency pair was listed on any exchange
Age of listing on a given exchange	Integer	Difference in months between current month and the first month a currency pair was listed on a particular exchange
Token	Indicator	Equals one for tokens and zero for coins

Table B.1. Variable description – continued

Variable	Type	Description
Quality measures		
MAD: Volume	Continuous (0, 1)	Mean absolute distance between the frequency of leading digits of the trading volume series in a currency pair on a given exchange in a given month and the frequency of leading digits given by Benford's Law
MAD: # Trades	Continuous (0, 1)	Mean absolute distance between the frequency of leading digits of the number of trades series aggregated at the ten-minute frequency in a currency pair on a given exchange in a given month and the frequency of leading digits given by Benford's Law
EDM: Volume	Integer	The number of structural breaks in series of trading volume within ten-minutes intervals of a currency pair on a given exchange in a given month identified by E-Divisive with Medians algorithm
EDM: # Trades	Integer	The number of structural breaks in series of number of trades within ten-minutes intervals involving a currency pair on a given exchange in a given month identified by E-Divisive with Medians algorithm
Principal components		
PC1: Volume	Continuous	The first principle component of Mad: Volume, and EDM: Volume
PC1: # Trades	Continuous	The first principle component of Mad: # Trades, and EDM: # Trades
PC1: Both	Continuous	The first principle component of Mad: Volume, EDM: Volume, Mad: # Trades, and EDM: # Trades

Table B.2. Exchange location. Area of main operations and headquarter country are from www.CoinGecko.com and www.Cointelligence.com.

Exchange	Main operations region	Headquarter country
ACX	Asia	Australia
Allcoin	North America	Canada
AnyBits	Western Europe	Ireland
BeQuant	Western Europe	Malta
Bibox	Asia	China
BigONE	Africa	Seychelles
Binance	Western Europe	Malta
Bitbank	Asia	Japan
BitBay	Eastern Europe	Poland
Bitfinex	Western Europe	BVI
bitFlyer	Asia	Japan
BitForex	Africa	Seychelles
Bitibu	Asia	Turkey
Bitlish	Western Europe	UK
BitMEX	Africa	Seychelles
Bitso	Others America	Mexico
Bitstamp	Western Europe	UK
Bittrex	North America	USA
Bit-Z	Asia	Hong Kong
BTC-Alpha	Western Europe	UK
BTC-e	Eastern Europe	Russia
BtcTurk	Asia	Turkey
bx.in.th	Asia	Thailand
CEX.IO	Western Europe	UK
Cobinhood	Asia	Taiwan
Coinbase	North America	USA
CoinEx	Western Europe	UK
CoinMate	Eastern Europe	Slovakia
Ethfinex	Western Europe	BVI
EXX	Asia	China
FTX	Others America	Bahamas
Gatecoin	Asia	Hong Kong
Gemini	North America	USA
HitBTC	Asia	Hong Kong
Huobi	Asia	China
Kraken	North America	USA
KuCoin	Africa	Seychelles
LocalBitcoins	Western Europe	Finland
OkCoin	Asia	China
OKX	Asia	Belize
Poloniex	North America	USA
Quoine	Asia	Japan
SouthXchange	Others America	Argentina
TheRockTrading	Western Europe	Italy
TideBit	Asia	Hong Kong
Tidex	Eastern Europe	Russia
UPbit	Asia	South Korea
Yobit	Eastern Europe	Russia
Zaif	Asia	Japan
ZB	Asia	China

Table B.3. CMC Adjusted Volume: Exchanges average CMC (www.coinmarketcap.com) adjusted ranking drop between the three months before and three months after the CMC index implementation.

CMC Rank Change			
Exchange	**Top Losers**	**Exchange**	**Top Winners**
Bibox	1	Binance	1
Bittrex	2	Bitstamp	2
Kucoin	3	Coinbase	3
Bitfinex	4	Poloniex	4
Cex-io	5	Bitforex	5

Table B.4. Average measures of anomalous trading within three months prior to CoinMarketCap index introduction.

Rank	Exchange	PCA	MAD	EDM
Below-median PCA				
1	KuCoin	-0.21	0.12	0.09
2	ZB	-0.36	-0.07	0.03
3	Tidex	-0.45	-0.18	-0.33
4	BTC-Alpha	-0.50	-0.35	-0.28
5	HitBTC	-0.57	-0.48	-0.22
6	Binance	-0.57	-0.19	-0.29
7	Coinbase	-0.60	-0.33	-0.47
8	TheRockTrading	-0.67	-0.13	-0.51
9	Bitbank	-0.70	-0.27	-0.52
10	Poloniex	-0.75	-0.36	-0.50
11	Bittrex	-0.81	-0.46	-0.40
12	bitFlyer	-0.82	-0.47	-0.34
13	Bitstamp	-0.82	-0.39	-0.55
14	Kraken	-0.84	-0.44	-0.48
15	Bitso	-0.87	-0.61	-0.36
16	UPbit	-0.88	-0.54	-0.44
17	Yobit	-0.92	-0.63	-0.58
18	CEX.IO	-0.94	-0.50	-0.49
19	Quoine	-0.94	-0.55	-0.62
20	Bitfinex	-0.94	-0.54	-0.53
21	SouthXchange	-1.04	-0.65	-0.63
22	BtcTurk	-1.05	-0.70	-0.46
23	OkCoin	-1.05	-0.72	-0.55
24	BitBay	-1.06	-0.69	-0.59
25	Gemini	-1.07	-0.66	-0.57
26	CoinMate	-1.10	-0.70	-0.58
27	Zaif	-1.12	-0.70	-0.65
28	TideBit	-1.22	-0.74	-0.71
Above-median PCA				
1	Bit-Z	0.68	0.82	0.28
2	OKX	0.80	1.13	0.37
3	BitForex	0.82	0.42	0.95
4	BeQuant	1.01	-0.24	2.46
5	EXX	1.19	0.79	0.92
6	Huobi	1.31	2.02	-0.19
7	CoinEx	2.16	2.62	1.59
8	Bibox	2.33	2.14	2.94
9	BigONE	3.17	3.22	3.20

References

Abdi, H. and L. J. Williams (2010). Principal component analysis. *Interdisciplinary reviews: computational statistics 2*(4), 433–459.

Amiram, D., Z. Bozanic, and E. Rouen (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies 20*(4), 1540–1593.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society 78*(4), 551–72.

Cong, W., X. Li, K. Tang, and Y. Yang (2022). Crypto wash trading. *Management Science 69*(11), 6427–6454.

James, N. A., A. Kejariwal, and D. S. Matteson (2016). Leveraging cloud data to mitigate user experience from 'breaking bad'. *IEEE International Conference on Big Data*, 3499–3508.

James, N. A. and D. S. Matteson (2015). ecp: An R package for nonparametric multiple change point analysis of multivariate data. *Journal of Statistical Software 62*(7), 1–25.

Jollife, I. T. (2002). Principal component analysis. 2nd Edition. Springer. New York.

Kaiser, H. F. (1961). A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology 14*, 1–2.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika 35*(4), 401–415.

Peres-Neto, Pedro R., D. A. J. and K. M. Somers (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis 49*(4), 974–997.