A Study of CUSUM Statistics on Bitcoin Transactions

Master's Thesis Presentation

Ivan E. Perez

Hunter College - CUNY

July 23, 2020

Presentation Outline

The Coinbase Exchange and the BTC-USD Limit Order Book

The Coinbase Exchange is a transparent exchange that publicly lists all market, limit, orders and the instantaneous state of the limit Order Book.

Types of Orders

1. **Limit Order**: A buy or sell order at a specified price in USD and size in BTC.
2. **Market Order**: A buy or sell order that is immediately filled at best price and size of the market making side.
3. **Canceled Order**: The removal of a limit order before execution, not directly observable in the Limit Order Book.

Table: Six recorded market order transactions collected from the Coinbase BTC-USD ticker channel using CoPrA

Timestamp (UTC)	Volume (BTC)	side	Price (USD)
03/19/2020 22:10:46	0.015439	buy	6252.08
03/19/2020 22:10:47	0.494871	buy	6255.91
03/19/2020 22:10:49	0.017539	buy	6254.72
03/19/2020 22:10:51	0.007641	buy	6254.73
03/19/2020 22:10:52	1.061969	sell	6254.72
03/19/2020 22:10:55	0.069624	buy	6255.90

Modeling Market Order Arrivals as a time-Homogeneous Compound Poisson Process (HCPP)

Definition: time-Homogeneous Compound Poisson Point Process

A stochastic process $\{Q(t)\}_{t \geq 0}$ is said to a compound Poisson process if it can be represented as:

$$Q(t) = \sum_{i=1}^{N(t)} Y_i, \qquad t \geq 0$$

where $\{N(t)\}_{t \geq 0}$ is a Poisson process, and $\{Y_i\}_{i \geq 1}$ is a family of i.i.d. RV's independent of $\{N(t)\}_{t \geq 0}$

We can apply Wald's equality to $Q(t)$:

$$\mathbb{E}[Q(t)] = \mathbb{E}\left[\sum_{i=1}^{N(t)} Y_i\right] = \mathbb{E}[N(t)]\mathbb{E}[Y_1] = \lambda t \mu, \tag{1}$$

In other words in a interval length t, the expected sum of arrivals $\lambda \mu t$.

Modeling Market Order Arrivals as a time-Homogeneous Compound Poisson Process (HCPP)

1. Modeling Market Order arrivals as a HCPP with message rate λ(msg/s), and mean order size, μ(BTC/msg).
2. Therefore, in an interval of length t, with $N(t)$ number of messages, and orders with mean μ, the expected total transaction volume in the interval is $\lambda\mu t$.
3. In order to estimate the baseline *message* arrival rate $\hat{\lambda}$, we count X messages in an interval of length t.

We model market orders using a HCPP where: Applying our HCPP construction, the messages are:

Timestamp (UTC)	Volume (BTC) $= Y_i$	t (s)	$N(t)$	$Q(t)$
03/19/2020 22:10:45	0.000000	0	0	0
03/19/2020 22:10:46	0.015439	1	1	0.015439
03/19/2020 22:10:47	0.494871	2	2	0.649261
03/19/2020 22:10:49	0.017539	4	3	0.666800
03/19/2020 22:10:51	0.007641	6	4	0.674441
03/19/2020 22:10:52	1.061969	7	5	1.736410
03/19/2020 22:10:55	0.069624	10	6	1.806034

Maximum Likelihood Estimation (MLE) of λ

Recall the probability mass function (PMF) of a discrete Poisson random variable X with mean λt:

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \qquad k = 0, 1, 2, \ldots \tag{2}$$

The joint distribution of $X_1, X_2, X_3, \ldots X_n$ is the product of their PMFs.

$$L(X_1, X_2, \ldots, X_n; \lambda) = \prod_{i=1}^{n} \frac{(\lambda t)^{X_i} e^{-\lambda t}}{X_i!}. \tag{3}$$

Taking the log of the likelihood function,

$$l(X_1, X_2, \ldots, X_n; \lambda) = \sum_{i=1}^{n} \left(X_i \log \lambda t - \lambda t - \log X_i! \right). \tag{4}$$

Differentiate with respect to λ and set equal to 0 to afford

$$\frac{1}{\lambda} \sum_{i=1}^{n} X_i - nt = 0. \tag{5}$$

MLE of rate of Market Order Arrival, $\lambda t \mu$

Solving for λ and recognizing that the running count of messages is $N(t)$, our MLE estimator is

$$\hat{\lambda} = \frac{1}{nt}\sum_{i=1}^{n} X_i = \frac{N(nt)}{nt}, \qquad n = 1, 2, \ldots \tag{6}$$

For the size of the market orders, recognize that their size is time invariant, but is indexed by $N(t)$. We take their mean at time nt to be:

$$\frac{1}{N(nt)}\sum_{i=1}^{N(nt)} Y_i = \hat{\mu}. \tag{7}$$

To assess whether the data obtained in 60 non-overlapping minute volume intervals follows a HCPP model in line with MLE rate $\hat{\lambda}\hat{\mu}$ we perform a χ^2-test with hypotheses:

$$H_0 : Q(t) \sim \text{compound Poisson with mean } \hat{\lambda}\hat{\mu}t.$$
$$H_1 : Q(t) \not\sim \text{compound Poisson with mean } \hat{\lambda}\hat{\mu}t.$$

χ^2-test Results for HCPP of Market Order Arrival Rate

Table: First and last entries of the χ^2 table to test the volume of market orders in 60 non-overlapping 1-minute intervals against an expected constant rate left, $\hat{\lambda}\hat{\mu} = 5.8483$ (BTC/min), right, $\hat{\lambda}\hat{\mu} = 2.7089$ (BTC/min).

	O	E	$\frac{(O-E)^2}{E}$
X_1	0.0346	5.8483	5.7793
X_2	2.4703	5.8483	1.9511
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{59}	1.3989	5.8483	3.3852
X_{60}	10.1915	5.8483	3.2253

	O	E	$\frac{(O-E)^2}{E}$
X_1	4.3675	2.7089	1.0155
X_2	5.1747	2.7089	2.2445
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{29}	4.2100	2.7089	0.8317
X_{30}	1.6771	2.7089	0.3931

Table: χ^2-test Results

Table	df	χ^2_{crit}	χ^2_{calc}	Reject H_0 ?
Left	59	77.931	352.01	Yes
Right	29	42.56	41.50	No

Real Time Detection of Instantaneous Changes in the Rate of Arrival of Market Orders

Suppose at some unknown time, $\nu, \nu = 1, 2, \ldots \leq N(t)$, when our sequence of market order arrivals, $Y_1, Y_2, Y_3, \ldots, Y_{N(t)}$ goes from a base rate $\lambda_0\mu$ to an increased rate of arrival $\lambda_1\mu$. To identify a stopping time τ, for the detection of the change $\lambda_0\mu \to \lambda_1\mu$, we use the log SPRT, $\log \mathcal{L}(Y_1, \ldots, Y_{N(t)}) = \ell(Y_1, \ldots, Y_{N(t)})$, with hypotheses

$$H_0 : Y_1, \ldots, Y_{N(t)} \sim \text{compound Poisson w/ rate } \lambda_0\mu,$$
$$H_1 : Y_1, \ldots, Y_{N(t)} \sim \text{compound Poisson w/ rate } \lambda_1\mu.$$

Our test becomes

$$\ell(Y_1, \ldots, Y_{N(t)}) \geq h \Rightarrow \text{Reject } H_0,$$
$$\ell(Y_1, \ldots, Y_{N(t)}) \leq 0 \Rightarrow \text{Fail to Reject } H_0,$$
$$0 \leq \ell(Y_1, \ldots, Y_{N(t)}) \leq h \Rightarrow \text{Continue sampling.}$$

Derivation of the log SPRT process, $u_t(\lambda_0, \mu)$

Starting from the SPRT, $\mathcal{L}(Y_1, \ldots, Y_{N(t)})$ we derive the log SPRT process $u_t(\lambda_0, \mu)$. We begin from

$$\mathcal{L}(Y_1, \ldots, Y_{N(t)}) = \frac{L(Y_1, \ldots, Y_{N(t)}; (\lambda_1, \mu))}{L(Y_1, \ldots, Y_{N(t)}; (\lambda_0, \mu))} \tag{8}$$

$$= \prod_{i=1}^{N(t)} \frac{e^{\lambda_1 \mu t}(\lambda_1 \mu t)^{Y_i}}{Y_i!} \frac{Y_i!}{e^{\lambda_0 \mu t}(\lambda_0 \mu t)^{Y_i}} \tag{9}$$

$$= e^{(\lambda_0 \mu - \lambda_1 \mu)t} \prod_{i=1}^{N(t)} \left(\frac{\lambda_1 \mu}{\lambda_0 \mu}\right)^{Y_i} \tag{10}$$

$$= e^{-(\lambda_0 \mu - \lambda_1 \mu)t} \left(\frac{\lambda_1 \mu}{\lambda_0 \mu}\right)^{Q(t)} \tag{11}$$

$$\ell(Y_1, \ldots, Y_{N(t)}) = Q(t) \log \left[\frac{\lambda_1 \mu}{\lambda_0 \mu}\right] - (\lambda_0 \mu - \lambda_1 \mu)t. \tag{12}$$

Our stopping time, τ, can be concisely expressed as

$$\tau = \inf\{t \geq 0; u_t(\lambda, \mu) \notin (0, h)\}. \tag{13}$$

Deriving the CUSUM Stochastic Process, $y_t(\lambda_0, \mu)$

1. Let $Y_1, Y_2, \dots Y_{N(t)}$, concisely described by the process $Q(t)$, be a sequence of Market Orders.
2. Suppose $Y_1, \dots, Y_{\nu-1}$, is compound Poisson distributed with parameters $\lambda_0\mu$, and
3. $Y_\nu, \dots, Y_{Nt}, N(t) \geq \nu$ is compound Poisson distributed with an increased parameter $\lambda_1\mu$.

From the log SPRT construction, we make a decision to *reject* H_0 when $u_t \geq h$, and *fail to reject* H_0 when $u_t \leq 0$.
The CUSUM stochastic process, $y_t(\lambda, \mu)$, arises when we attempt to test composite hypotheses of the SPRT with a shifting index:

$$
\begin{aligned}
&H_0 : \ell(Y_1, \dots, Y_{N(t)}) < h, H_1 : \ell(Y_1, \dots, Y_{N(t)}) > h, \\
&H_2 : \ell(Y_2, \dots, Y_{N(t)}) > h, \ell(Y_3, \dots, Y_{N(t)}) > h, \dots, \ell(Y_\nu, \dots, Y_{N(t)}) > h.
\end{aligned}
$$

To *minimize* the delay between when the change happens, ν, and the stopping time τ. We must *maximize* the difference between $\ell(Y_1, \dots, Y_{N(t)})$ and $\ell(Y_1, \dots, Y_\nu)$. Before ν, $\ell(Q(t))$ reaches a new minimum upon arrival of a new Y. Therefore:

$$
\max_{1 \leq \nu \leq N(t)} [\ell(Y_1, \dots, Y_{N(t)}) - \ell(Y_1, \dots, Y_{\nu-1})] = \ell(Q(t)) - \min_{1 \leq \nu \leq N(t)} [\ell(Q(t))] \tag{14}
$$

CUSUM up and down definitions & application for our work

Definition: 2-CUSUM Construction

Let $\{Q(t)\}_{t\geq 0}$ be a compound Poisson process with mean $\lambda_0 \mu t$, where $\lambda_1^+ = (1+\epsilon)\lambda_0$, $\lambda_1^- = (1-\epsilon)\lambda_0$, $\{h, \lambda_0, \epsilon, \mu\} \in \mathbb{R}^+$ define the following processes:

$$u_t^+(\lambda_0, \mu) = Q(t) \log \left[\frac{\lambda_1^+, \mu}{\lambda_0, \mu} \right] - t\epsilon; \, m_t^+(\lambda_0, \mu) = \inf_{0 \leq s \leq t} u_s^+(\lambda_0, \mu) \tag{15}$$

$$u_t^-(\lambda_0, \mu) = Q(t) \log \left[\frac{\lambda_1^- \mu}{\lambda_0 \mu} \right] + t\epsilon; \, m_t^-(\lambda_0, \mu) = \inf_{0 \leq s \leq t} u_s^-(\lambda_0, \mu) \tag{16}$$

$$y_t^+(\lambda_0, \mu) = u_t^+(\lambda_0, \mu) - m_t^+(\lambda_0, \mu), \tag{17}$$

$$y_t^-(\lambda_0, \mu) = u_t^-(\lambda_0, \mu) - m_t^-(\lambda_0, \mu), \tag{18}$$

$$\tau = \inf\{t \geq 0; y_t^+(\lambda_0, \mu) \vee y_t^-(\lambda_0, \mu) \geq h\}, \leftarrow \text{the CUSUM stopping time.} \tag{19}$$

Classification of Stopping times (alarms), τ

To identify shift to an increased(decreased) rate, $\lambda_0 \to \lambda_1^+$; $(\lambda_0 \to \lambda_1^-)$ we distinguish τ:

$\tau^+ = \inf\{t \geq 0; y_t^+(\lambda_0, \mu) \geq h\}$, an upward alarm is declared.

$\tau^- = \inf\{t \geq 0; y_t^-(\lambda_0, \mu) \geq h\}$, a downward alarm is declared.

Implementation of 2-CUSUM Construction - Short Experiments

Applying 2-CUSUM construction to our sequence of market orders, with an initial estimate for λ_0 we display CUSUM-up and CUSUM-down.

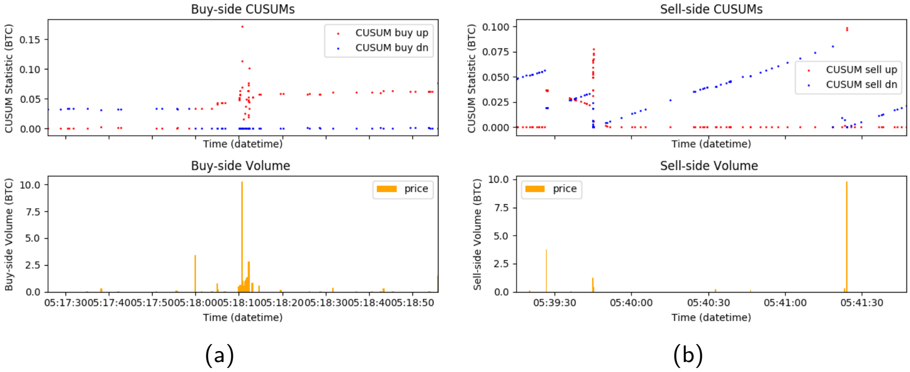


Figure: Two-sided CUSUM processes for buy(a) and sell(b) sides, under set with last transaction volume (BTC)

Results show *short bursts of high transaction volume* that trigger many upward alarms in *quick succession*.

Using a Sequence of CUSUM alarms to define Active/Inactive Periods

Structure of the set of market Orders $\mathcal{O}$

Let the set of market orders, $\mathcal{O}$, be a tuple of random variables ordered by arrival time, $t_i, i \in \mathbb{N}_0$. The i^{th} element in $\mathcal{O}$, (Y_i, t_i) is an ordered pair containing transaction volume as the random variable, Y_i, and its arrival time, t_i. They are broken up into the set of Active periods, $\mathcal{A}$, and Inactive periods, $\mathcal{I} = \{\mathcal{A}^c \in \mathcal{O}\}$.

Defintion: Active/Inactive Periods

An Active Period, $A_n, n = 1, 2, \ldots \in \mathcal{A}$ is also a tuple of observations, (Y_i, t_i), that contain at least 2 or more upward alarms. Inactive Periods, $I_n, n = 1, 2, \ldots \in \mathcal{I}$, are the remaining orders.

Active Period Start and End Points

An Active periods begins on τ_1^+, and continues if there is a second up-alarm, τ_2^+ shortly thereafter. Let $T = \tau_j^+ - \tau_{j-1}^+, j = 1, 2, \ldots$.

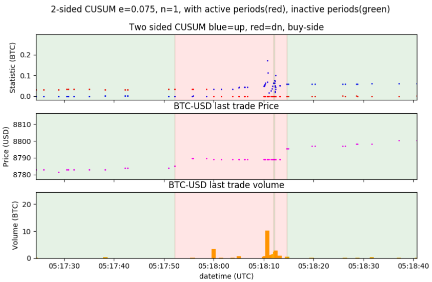
An Active period ends on the detection of a down-alarm, τ^- or if T seconds have passed since the last up-alarm, τ_j.

Active Period Start Compensation

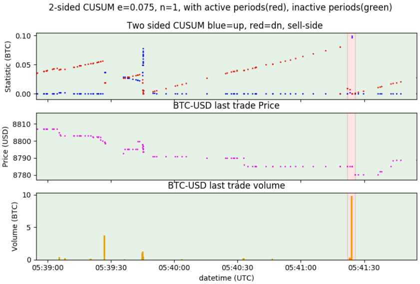
Let t_0 be the time y_t^+ was last reinitialized. Let $D = \tau_1^+ - t_0$, represent time between last reset and the first alarm. In order to roughly capture ν, we shift the beginning of the active period backwards by $\min[D, T]$

Defining Active/Inactive Periods - Short Experiments

Applying our definition for Active(red) and Inactive(green) periods to results from the 2-CUSUM construction.



(a)



(b)

Figure: Two-sided CUSUM processes for buy(a) and sell(b) sides, under set with last transaction volume (BTC)

Results show that sometimes there are larger swings in BTC-USD Price(magenta) during Active periods.

Defining Price-Swings as Momentum

Defintion: Input Period $\mathcal{S}$

Let $\mathcal{S} = \{(Y_i, t_i), i \in \mathbb{N}_0\}$.
$\mathcal{S}$ is the tuple collection of observations corresponding to A_m or I_m.

Definition: Momentum, $M : \mathcal{S} \to \mathbb{R}^+$

$M(\mathcal{S}) = (\max_{n \in \mathcal{S}} Y_n - \min_{n \in \mathcal{S}} X_n)/(t_0 - t_{|\mathcal{S}|})$
Momentum (in BTC/s) is defined to be the difference between the largest and smallest transaction price (BTC) divided by the length of the interval in seconds.

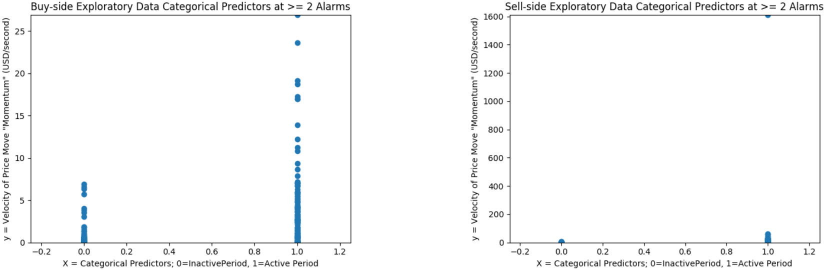
Linear Model

By encoding Active periods as 1, and Inactive periods $= 0$. The set of m momentum observations becomes $\mathcal{M} = \{(M_i, E_i), i = \mathbb{N}_0 \leq m\}$. The linear model

$$M_m = \beta_0 + \beta_1 E_m$$

examines the level effect β_1 on M_m from categorical predictor E_m.

24h Experiments - 2 alarms

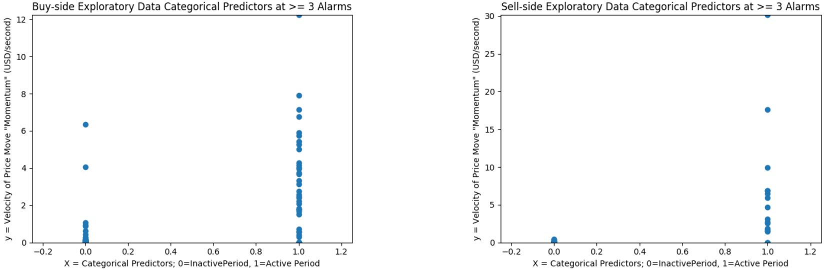


(a) (b)

Figure: Categorical data for active(1), and inactive(0) periods against momentum for buy(a), and sell(b) side momentum experiments with at least 2 consecutive up alarms

Summary 2-alarms, buy-side				
Groups	Count	Sum	Average	VAR.S
0	124	200.986	0.5422	1.621
1	123	2944.891	3.986	23.942
Summary 2-alarms, sell-side				
Groups	Count	Sum	Average	VAR.S
0	91	151.812	0.479	1.668
1	90	2.551 E 06	25.682	2.834 E 04

24h Experiments - 3 alarms



(a) (b)

Figure: Categorical data for active(1), and inactive(0) periods against momentum for buy(a), and sell(b) side momentum experiments with at least 2 consecutive up alarms

Summary 3-alarms, buy-side				
Groups	Count	Sum	Average	VAR.S
0	41	53.204	0.417	1.299
1	40	281.868	2.933	7.047

Summary 3-alarms, sell-side				
Groups	Count	Sum	Average	VAR.S
0	20	0.350	0.115	0.018
1	19	0.989	5.550	51.631

One-Way ANOVA of Active & Inactive Periods

The One-Way ANOVA tests the Hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. for the two groups of observed Momentums.

Table: ANOVA Table for Buy/Sell side experiments with at least 2 and 3 alarms.

cat.	Source of Variation	SS	df	MSS	F	F-crit
buy 2-alarms	Between Groups	732.445	1	732	57.043	3.877
	Within Groups	3145.878	245	12.840		
	Total	3878.322				
	Source of Variation	SS	df	MSS	F	F-crit
sell 2-alarms	Between Groups	2.874 E 04	1	0.479	2.017	3.894
	Within Groups	2.551 E 06	179	25.682		
	Total	2.580 E 06				
	Source of Variation	SS	df	MSS	F	F-crit
buy 3-alarms	Between Groups	128.216	1	2.874 E 04	30.226	3.962
	Within Groups	335.109	79	1.425 E 04		
	Total	463.325				
	Source of Variation	SS	df	MSS	F	F-crit
sell 3-alarms	Between Groups	277.266	1	128.216	10.454	4.105
	Within Groups	981.339	37	4.242		
	Total	1258.604				

Conclusions and Future Work

Conclusions:

1. Rates of market order arrivals can be estimated locally using a compound poisson process
2. CUSUM algorithms can be used to detect changes to increased(decreased) rates of market order arrival.
3. Sequences of alarms can be used to identify short periods of with large price swings.
4. These short period have are known statistically significant effect sizes.

Future Work, Developing new detection schemes that

1. do not rely on a backward looking algorithm,
2. reduce the rate of one off up-alarms,
3. have the added flexibility for different distributions of order sizes.

bibliography

beamer doesn't support bibtex Ihave to figure this one out

current progress on future works