

自己励起型点過程モデルによるビットコイン為替市場の 高頻度データ分析

慶應義塾大学理工学部管理工学科 3 年

中妻照雄研究会

岡田知樹

概要

近年、金融高頻度データ分析における研究が盛んに行われている。その中でも、点過程モデルが高頻度データを解析するための有用なクラスとして数多くの論文で利用されている。点過程モデルの中でも、自己励起性を持つ点過程である Hawkes 過程は、データのクラスター性などをモデル化できるという点で非常に強力であり、株、為替の約定や板情報を分析した多くの先行研究が存在する。近年の金融業界の革新的な動きとして、FinTech がにわかに騒がれ始めているが、本論文ではその一部である仮想通貨ビットコインに着目し、高頻度データ分析を行った。具体的には、ビットコインの為替市場の約定データを一変量の Hawkes 過程に当てはめて約定系列をモデル化することを試みた。このとき、モデルの当てはまりの良さの検証法として、時間変更を施した系列に対しての QQ プロットと、バーコードプロットという 2 種類のグラフを作成し検証を行った。結果として、ビットコイン為替市場の特性として、株や一般的な為替市場に比べて、外因的影響よりも、自己励起的作用によって場が盛り上がることが多いということが確認できた。また、モデルのそのものの当てはまりの良さに関しては、一変量の Hawkes 過程では分布の裾の部分で当てはまりが悪く、より良いモデルにするには、Hawkes 過程の次元の増加や非線形化などを検討する必要があるという結論に至った。

Keywords : ビットコイン、高頻度データ、点過程、Hawkes 過程、時間変更

目次

1	はじめに	2
2	Hawkes 過程の理論的背景	5
3	パラメータ推定	8
4	実証分析	10
5	結論	16

1 はじめに

2016年8月31日の税制改正要望において、金融庁はビットコインなどの仮想通貨の購入に消費税が適用されるかどうかを明確にするよう求めた。ビットコインなどで買い物を行うときにも消費税が発生することから二重課税ではないかという指摘があったからである。これにより、今後2017年4月を目処に仮想通貨への消費税はなくなる見込みである*1。

最近の国のこのような動きから分かるように、仮想通貨に対する社会の関心は高まってきている。その一因を担っているものとして、金融業界の革新的なムーブメントとされる「FinTech」がある。FinTechとはFinanceとTechnologyを掛け合わせた造語で、おおよそ金融業界のIT革命を意味する。背景には、ITやネットワークが発展していくにつれて、消費者がより優れたユーザーインターフェースを金融サービスにも求めるようになってきたという事情がある。そうした流れの中で、巧妙な暗号システムにより信頼性が高く、安価な手数料で瞬時に海外送金などを可能にする理想の通貨として、仮想通貨は評価を高めてきている。実は、今のFinTechブーム以前から仮想通貨は存在し多くの注目を集めていた。今流通しているビットコインなどの仮想通貨の起源は2008年のとある論文にある。当時、Satoshi Nakamotoと名乗る人物が暗号通貨に関する論文[1]をインターネット上に投稿し、一部で大きな話題となった。その内容に関心を持ったネット上のエンジニアら有志たちによって、論文発表からわずか3か月足らずで仮想通貨「ビットコイン」(Bitcoin)が開発され、そのおよそ1年後の2010年2月にビットコインが両替できる最初の取引所が設立された。最初こそほぼ価値がないに等しいほど為替レートが低かったが、メディアに注目されたことをきっかけに急速に価格が急上昇する。2011年4月16日時点で1ビットコイン当たり87円だった終値が、およそ2か月後の2011年6月12日には1ビットコイン当たり1,489円と大幅な値上がりを見せた。これがいわゆる最初のビットコインバブルである。その後、特定の国の外貨の信用が下がれば、相対的にビットコインの価値が上がったり、Mt.Goxなど大手取引所でのビットコイン流出事件*2で価値を下げたりと、激しく為替レートは変化していき、その値動きの激しさは「ビットコインのボラティリティは他の外貨の為替レートの10倍である」と評されるほどであった。今ではその激しい値動きなどから投機目的での需要も増し、また前述した昨今のFinTechブームもあって、ビットコインをはじめとする仮想通貨の需要は日に日に高まっている。

実際に、ビットコインと円の両替所として国内トップシェアを誇るBitFlyerでは日々多くのBTC/JPY為替取引が行われている。図1はBitFlyerにおける2016年10月2日の1日分の約定データを1分間隔で集計したものである。この日の合計の約定回数は8178回であり、1番多い時間では1分あたり60回以上約定していることが分かる。また、グラフから見てとれる特徴とし

*1 2016年現在、G7の中で仮想通貨に課税をしているのは日本だけである。

*2 2014年2月、世界最大のビットコイン両替所であったMt.Goxは744,408ビットコイン(当時、日本円にして200億円以上)の預かり資産のほとんどをハッキング攻撃によって失ったとして日本の裁判所に対して破産申請を行った。しかし、のちの調査でこの事件は外部からのサイバー攻撃によるものではなく、内部の取引所の不正操作によるものとされ、社長のマルク・カルプレス容疑者が横領の疑いで逮捕されるなど、取引所自体の問題でありビットコインそのものの信頼性は全くの無関係であることが明らかになる。

て、約定が多く起こる時間帯とあまり起こらない時間帯が存在することが分かる。本論文では、このようなビットコイン為替市場のクラスター性を持った約定系列をモデリングすることを考える。

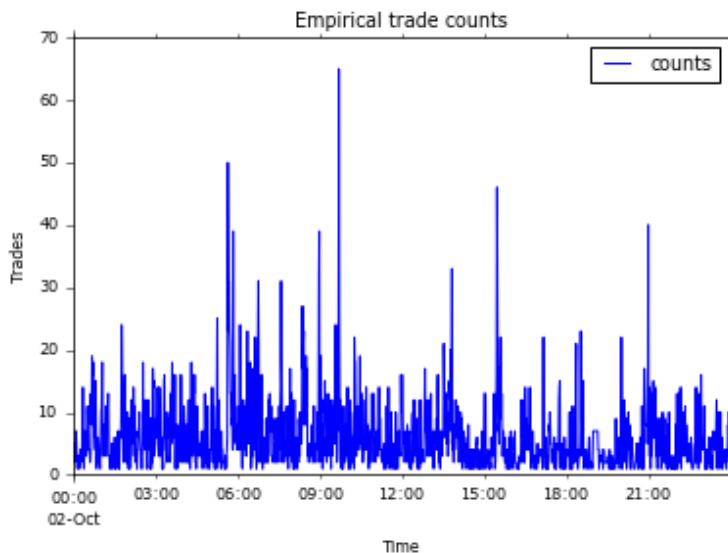


図 1: BitFlyer における 2016 年 10 月 2 日の BTC/JPY 為替取引の約定回数と時間の関係

今回、一定期間のすべての約定データを取り扱うが、このような非常に時間間隔が細かい大量のデータのことを「高頻度データ」(High frequency data)といい、近年積極的に研究がなされている。高頻度データを扱うときに、例えば約定時間間隔などは一定ではなく、従来の時系列分析の枠組みで直接分析することができなくなる。そこで約定時間間隔そのものが確率的に支配されているとし、分析を試みた重要な先行研究として Engle & Russell[2],[3]がある。これらの論文では GARCH モデルを転用した、ACD(Autoregressive Conditional Duration) モデルが提案されている。しかし、ACD モデルなどでは複数の系列がお互いに影響しあって、確率的に発生していくような高次元のモデルを扱うのは難しかった。そこで近年注目を浴びているモデルが Hawkes 過程という自己励起性を持った点過程モデルである。もともとは地震の発生などを表現するモデルであり有名な論文として、Ogata[4]がある。地震以外にも、遺伝学 (Reynaud-Bouret & Schbath[5])、神経科学 (Johnson[6])、天文学 (Baddeley[7]) など様々な分野に応用されており、ユニークなものでは犯罪学への応用 (Mohler et al.[8]) もある。自己励起性を持ち、確率的に発生する事象をモデル化しようとしたときに、Hawkes 過程は分野を問わずに非常に大きな威力を発揮する。しかし、近年最も Hawkes 過程が応用されている領域こそが金融分野である。Hawkes 過程の金融分野への応用例としては、金融高頻度データ分析のテーマの一つであるマイクロストラクチャー・ノイズを Hawkes 過程を利用して分析を行った Bacry et al.[9]、マーケット・メイキングをテーマとして取り上げた Toke[10]、注文板のモデル化を試みた Large[11] などがある。その他の応用例は Toke[12] や Bacry et al.[13] で多く紹介されている。

本論文のメインテーマは Hawkes 過程を用いて、ビットコイン為替市場の約定データを分析する

ということである。先行研究として、FX の約定データを BUY と SELL の 2 つの系列に分けて 2 変量 Hawkes 過程に当てはめて分析を行った Hewlett[14] や、同様の分析を株式に対して行い、最適執行戦略を考えた Calsson et al.[15] などがある。本論文の貢献は、ビットコイン為替市場に対して Hawkes 過程を当てはめて分析を行ったことにある。ビットコイン為替市場を分析しようと考えた動機は、取引所の API を通じて容易に約定データが入手できるということと、以下に挙げるビットコイン特有のいくつかの特徴が、分析結果にどのような影響を及ぼすか検証したかったということが挙げられる。

表 1: ビットコインの主な特徴

-
- ・ボラティリティが非常に大きい
 - ・取引情報がオープンで、すべてのマーケット参加者に対して公平である
 - ・国の中央銀行などの発行主体が存在しない分散型通貨である
 - ・発行量に上限があるため、原理上インフレは起きない
-

本論文の構成は、まず 2 章で Hawkes 過程の理論的背景を概観する。3 章では Hawkes 過程のパラメータ推定法について説明する。本論文では最尤法を用いて推定を行っているため、Hawkes 過程がどのような尤度関数を有するのかを明示する。4 章では実証分析を行なう。使用したデータについて詳しく説明し、その後パラメータの推定結果を考察し、モデルの当てはまりの良さを確かめるために、QQ プロットとのちに詳しく説明するバーコードプロットという 2 種類のグラフを用いる。最後に推定したパラメータとシミュレーション・アルゴリズムを利用して、シミュレーション結果をグラフで描写し、5 章でまとめと今後の研究の展望について述べる。

2 Hawkes 過程の理論的背景

Hawkes 過程の定義の前に、大枠である点過程について説明しなければならない。点過程とは確率過程の一種であり、時間や地理的な空間上の独立した点の集合をモデル化したものである。点過程は確率空間上のある事象 ω に対して決まる点列として次のように定義される。

定義 2.1 (点過程) $(\Omega, \mathcal{F}, \mathbf{P})$ を確率空間とする。任意の $i \in \mathbb{N}$ について、 $t_i < t_{i+1}$ であるような非負の確率変数列 $(t_i)_{i \in \mathbb{N}}$ を \mathbb{R}_+ 上の点過程という。

次に、計数過程を以下のように定義する。

定義 2.2 (計数過程) $(t_i)_{i \in \mathbb{N}}$ を点過程としたとき

$$N(t) = \sum_{i \in \mathbb{N}} \mathbf{1}_{t_i \leq t} \quad (1)$$

を $(t_i)_{i \in \mathbb{N}}$ に関連した計数過程という。

$\mathbf{1}_{t_i \leq t}$ は定義関数であり、 t より過去の事象の数が $N(t)$ となることを意味している。 $(t_i)_{i \in \mathbb{N}}$ の代わりに $N(t)$ が与えられたとしても、 $t_{i+1} = \inf\{t > t_i \mid N(t) - N(t_i) = 1\}$ より $(t_i)_{i \in \mathbb{N}}$ が取り出せるので、 $N(t)$ 自体を点過程としても差し支えない。また計数過程と同様に、点過程に関連付けられてデュレーション過程を次のように定義できる。

定義 2.3 (デュレーション過程) $(t_i)_{i \in \mathbb{N}}$ を点過程としたとき

$$\forall i \in \mathbb{N}, \delta t_i = t_i - t_{i-1} \quad (2)$$

を $(t_i)_{i \in \mathbb{N}}$ に関連したデュレーション過程という。

最後に強度関数という概念を導入する。

定義 2.4 (強度関数) $N(t)$ があるフィルトレーション \mathcal{F}_t に適合している*³点過程であるとする、左連続な強度関数が以下のように定義される。

$$\lambda(t|\mathcal{F}_t) = \lim_{h \downarrow 0} \mathbf{E} \left[\frac{N(t+h) - N(t)}{h} \middle| \mathcal{F}_t \right] \quad (3)$$

また、同様に

$$\lambda(t|\mathcal{F}_t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbf{P}[N(t+h) - N(t) > 0 | \mathcal{F}_t] \quad (4)$$

*³ ここでいう $N(t)$ がフィルトレーション \mathcal{F}_t に適合しているとは、各時点 t において $N(t)$ が \mathcal{F}_t -可測であるということの意味する。

(4) から分かるように、強度関数は時点 t までの情報が与えられたもとの、その次の事象が観測される確率密度を表している。以後、 $N(t)$ に対してフィルトレーション \mathcal{F}_t が常に存在すると仮定して、煩わしさを避けるために $\lambda(t|\mathcal{F}_t)$ を $\lambda(t)$ と省略して表記する。

最も単純な点過程モデルとして、ポアソン過程がある。

定義 2.5 (ポアソン過程) λ を 0 以上の実数としたとき、以下で定義される点過程をポアソン過程という。

$$P[N(t+h) - N(t) = 1 | \mathcal{F}_t] = \lambda h + o(h) \quad (5)$$

$$P[N(t+h) - N(t) > 1 | \mathcal{F}_t] = o(h) \quad (6)$$

ポアソン過程は過去の系列に依存しないということと、デュレーションがパラメーター λ の指数分布に従うという特徴を持っている。

Hawkes[16] において、自己励起性を持つ点過程である Hawkes 過程が提案した。Hawkes 過程はポアソン過程をより一般化したモデルであり、ポアソン過程では一定値だった強度関数 λ が過去の事象によって変化するようになっている。ポアソン過程との違いは強度関数だけであるので、以下では Hawkes 過程の強度関数がどのように表現されるかを、一変量のケースと多変量のケースに分けて示す。

定義 2.6 (一変量 Hawkes 過程の強度関数)

$$\begin{aligned} \lambda(t) &= \lambda_0(t) + \int_0^t \sum_{j=1}^P \alpha_j e^{-\beta_j(t-s)} dN_s \\ &= \lambda_0(t) + \sum_{t_i < t} \sum_{j=1}^P \alpha_j e^{-\beta_j(t-t_i)} \end{aligned} \quad (7)$$

定義 2.7 (多変量 Hawkes 過程の強度関数)

$$\begin{aligned} \lambda^m(t) &= \lambda_0^m(t) + \sum_{n=1}^M \int_0^t \sum_{j=1}^P \alpha_j^{mn} e^{-\beta_j^{mn}(t-s)} dN_s^n \\ &= \lambda_0^m(t) + \sum_{t_i^n < t} \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{mn} e^{-\beta_j^{mn}(t-s)} \end{aligned} \quad (8)$$

定義式から分かるように、Hawkes 過程の強度関数は過去の事象との時間差の関数 (指数関数) の線形和で現在の強度関数の値が決まるという形を成している。例えば簡単のため $P = 1$ としたとき、一変量 Hawkes 過程において、 α は事象が起こった時の強度関数のジャンプ幅を表し、 β がその後指数関数的に減衰していく度合を表している。また、 $\lambda_0(t)$ を時間に依存せず一定とすれば、常に最低限保たれるベースの強度を表している。以上のことは、事象の発生系列を確定的に与えたとき、 $\alpha = 0.2$ 、 $\beta = 0.5$ 、 $\lambda_0 = 0.1$ とした Hawkes 過程の強度は次の図 2 のようになることから確認

できる。

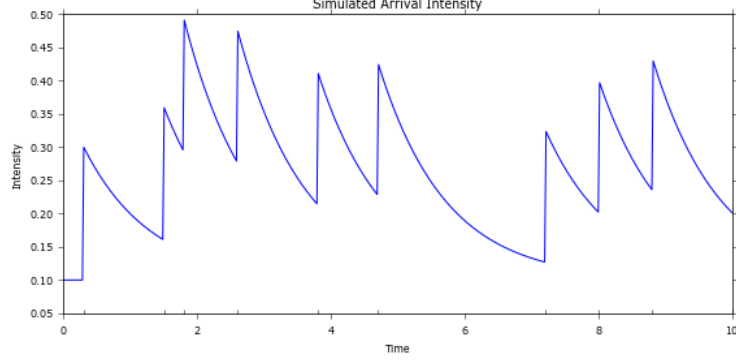


図 2: パラメーターを $\alpha = 0.2$, $\beta = 0.5$, $\lambda_0 = 0.1$ とし、発生時間系列を $t = \{0.3, 1.5, 1.8, 2.6, 3.8, 4.7, 7.2, 8.0, 8.8\}$ と与えたときの Hawkes 過程の強度関数の様子

点過程モデルには、唯一、定常性のみが仮定されている。多変量 Hawkes 過程における定常性の条件は次のように示される。

定理 2.1 (多変量 Hawkes 過程の定常条件) 多変量 Hawkes 過程に対して以下の行列

$$\mathbf{G}(t) = \left(\alpha^{mn} e^{-\beta^{mn}(t-s)} \right)_{m,n=1,\dots,M} \quad (9)$$

のスペクトル半径

$$\mathbf{\Gamma} = \int_0^\infty \mathbf{G}(u) du = \left(\frac{\alpha^{mn}}{\beta^{mn}} \right)_{m,n=1,\dots,M} \quad (10)$$

の値が 1 未満ならば定常である。

つまり一変量 Hawkes 過程ならば、 $\frac{\alpha}{\beta} < 1$ を満たせば定常である。

点過程モデルにおいて重要な定理に次の時間変更定理というものがある。詳しい証明は Daley & Vere-Jones[17] を参照。

定理 2.2 (時間変更定理) 所与の時間列 $\{t_i\}$ に対して、新たな時間列 $\{\tau_i\}$ を以下のように定義する。

$$\tau_0 = \int_0^{t_0} \lambda_0 ds = \Lambda(0, t_0) \quad (11)$$

$$\tau_i = \tau_{i-1} + \Lambda(t_{i-1}, t_i) \quad (12)$$

ただし、

$$\Lambda(t_{i-1}, t_i) = \int_{t_{i-1}}^{t_i} \lambda_0(s) ds + \sum_{j=1}^P \frac{\alpha_j}{\beta_j} \left(1 - e^{-\beta_j(t_i - t_{i-1})} \right) A_j(i-1) \quad (13)$$

であり、 $A_j(i)$ は以下のように再帰的に計算される。 $A_j(0)$ は 0。

$$A_j(i-1) = 1 + e^{-\beta_j(t_{i-1}-t_{i-2})} A_j(i-2) \quad (14)$$

以上の時間変更で得られた $\{\tau_i\}$ はポアソン過程に従う。

この事実が成り立つことより、Hawkes 過程で推定した結果を用いて、時間変更を行った系列がポアソン過程に従っているかどうかを確かめることで、モデルそのものの当てはまりの良さを検証できる。Hawkes 過程を用いた分析で当てはまりの良さを確かめる際に最も頻繁に行われている手法であり、本論文でも時間変更を利用してモデルの良さを検証する。

Ogata & Akaike[18] において、Hawkes 過程のシミュレーション・アルゴリズムが提案された。そのアルゴリズムを用いることで、確定的に事象の発生時間系列を与えた図 2 とは別に、パラメータを指定するだけで Hawkes 過程を次の図のようにシミュレーションできる。

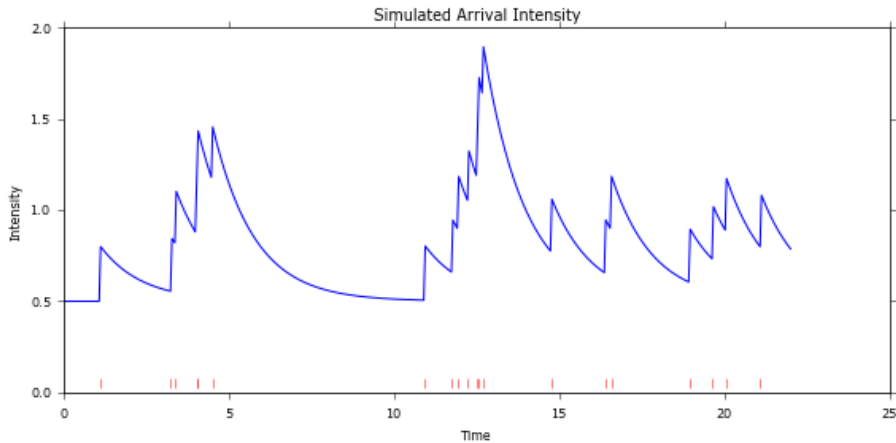


図 3: $\alpha = 0.3$, $\beta = 0.8$, $\lambda_0 = 0.5$ としたときの一変量 Hawkes 過程のシミュレーション

3 パラメーター推定

Hawkes 過程のパラメーター推定法は、Ogata[19] において提案された最尤法が最もよく知られている。他に、EM アルゴリズムを用いた推定法は Veen and Shoenberg[20]、ノンパラメトリック推定法は Bacry et al.[21] で行われている。最尤法は単純で最も実装しやすいが、Hawkes 過程の次元が上がると忽ち上手く推定できなくなるという欠点がある。これは、Hawkes 過程における推定すべきパラメーターが次数の数に対して 2 乗のオーダーで増加していくため、高次元であればあるほど尤度関数が複雑になり、最適化が困難になるからである。高次元の Hawkes 過程を取り扱った論文として、Linderman and Adams[22] などがあるが、ギブス・サンプラーなどのベイズ分析的手法を用いて計算上の問題をクリアしている。本論文では、一変量 Hawkes 過程を取り扱う

ので、最尤法でのパラメーター推定を考える。

Ogata[19]において、Hawkes 過程の対数尤度関数は一変量と多変量の場合でそれぞれ次のように表現される。

定理 3.1 (一変量 Hawkes 過程の対数尤度関数)

$$\begin{aligned} \ln \mathcal{L}(\{t_i\}_{i=1, \dots, n}) &= t_n - \int_0^{t_n} \lambda_0(s) ds - \sum_{i=1}^n \sum_{j=1}^P \frac{\alpha_j}{\beta_j} (1 - \exp -\beta_j(t_n - t_i)) \\ &\quad + \sum_{i=1}^n \ln \left[\lambda_0(t_i) + \sum_{j=1}^P \alpha_j R_j(i) \right] \end{aligned} \quad (15)$$

ただし、 $R_j(i)$ は以下のように再帰的に計算される。 $R_j(0)$ は 0 とする。

$$R_j(i) = \exp -\beta_j(t_i - t_{i-1})(1 + R_j(i-1)) \quad (16)$$

定理 3.2 (多変量 Hawkes 過程の対数尤度関数) 多変量 Hawkes 過程の対数尤度は、各次元の対数尤度の和で表現できるので、

$$\ln \mathcal{L}(\{N(t)\}_{t \leq T}) = \sum_{m=1}^M \ln \mathcal{L}^m(\{N^m(t)\}_{t \leq T}) \quad (17)$$

であり、また各次元の対数尤度は

$$\begin{aligned} \ln \mathcal{L}^m(\{N^m(t)\}_{t \leq T}) &= \int_0^T (1 - \lambda^m(s)) ds + \int_0^T \ln \lambda^m(s) dN^m(s) \\ &= T - \int_0^T \lambda^m(s) ds \\ &\quad + \sum_{i: T_i \leq T} \mathbf{1}_{\{Z_i=m\}} \ln \left[\lambda_0^m(T_i) + \sum_{n=1}^M \sum_{j=1}^P \sum_{T_k^n < T_i} \alpha_j^{mn} e^{-\beta_j^{mn}(T_i - T_k^n)} \right] \end{aligned} \quad (18)$$

である。これを変形すると

$$\ln \mathcal{L}^m(\{N^m(t)\}_{t \leq T}) = T - \int_0^T \lambda^m(s) ds \quad (19)$$

$$\begin{aligned} &- \sum_{i: T_i \leq T} \sum_{n=1}^M \sum_{j=1}^P \frac{\alpha_j^{mn}}{\beta_j^{mn}} (1 - e^{-\beta_j^{mn}(T - T_i)}) \\ &+ \sum_{l: T_l^m \leq T} \ln \left[\lambda_0^m(T_l^m) + \sum_{n=1}^M \sum_{j=1}^P \alpha_j^{mn} R_j^{mn}(l) \right] \end{aligned} \quad (20)$$

と表せる。ただし $R_j^{mn}(l)$ は次のように再帰的に計算される。 $R_j^{mn}(0) = 0$ である。

$$\begin{aligned}
 R_j^{mn}(l) &= \sum_{T_k^n < T_l^m} e^{-\beta_j^{mn}(T_l^m - T_k^n)} \\
 &= \begin{cases} e^{-\beta_j^{mn}(T_l^m - T_{l-1}^m)} R_j^{mn}(l-1) + \sum_{T_{l-1}^m \leq T_k^n < T_l^m} e^{-\beta_j^{mn}(T_l^m - T_k^n)} & \text{if } m \neq n \\ e^{-\beta_j^{mn}(T_l^m - T_{l-1}^m)} (1 + R_j^{mn}(l-1)) & \text{if } m = n \end{cases}
 \end{aligned} \tag{21}$$

これら尤度関数をプログラム上に定義し、最適化アルゴリズムを適用することによって Hawkes 過程のパラメーター推定を行うことができる。

4 実証分析

データは 2016/6/23/00:00 から 2016/6/30/00:00 までの一週間分の BitFlyer の約定履歴を使用する。これらのデータを 1 日単位で分割し、それぞれに対して一変量 Hawkes 過程へフィッティングを行う。データ区間として 2016 年の 6 月下旬を選んだ理由は、図 4 を見ればわかる通り、激しく為替レートが動いているため分析対象として興味深いと考えたからである。



図 4: 2016 年 6 月下旬の BTC/JPY 為替レート

BitFlyer の約定履歴は、取引所が公開している API を通じて簡単に取得することができる。得られる約定データのタイムスタンプはミリ秒単位であり、非常に細かいデータを得ることができる。しかし、データを扱う前に注意しなければならないことがある。それは、取得した大量のデータの中に割と頻繁に同じ時間で記録されている約定データが混ざっているということである。点過程モデルの前提として、同系列に同じ時間が存在してはならないので、修正を施さなければならない。そもそもタイムスタンプがミリ秒単位であるのに同じ時間の記録が頻繁に現れるのはおかしいのだが、実は API で得られる約定データにはいつの売り注文と買い注文がマッチングしたまで記録されているため、このような現象が起こってしまっている。例えば、1 ビットコインの買い注文を出したとき、瞬時に約定されたとして、その内訳が 0.2 ビットコインはある売り注文、残りの

0.8 ビットコインは先ほどの売り注文とは別の売り注文とマッチングしたとなっていれば、この1 ビットコインの買い注文の約定は同時刻で2つ記録されてしまう。このようなことが原因で、約定データに同じ時間が存在するのだが、約定という事象の発生は一回しか起こっていないと捉えれば、同時刻の複数のタイムスタンプは一つの約定データとして扱うのが妥当である。従って、点過程の前提を崩さないためにも、今回分析するにあたって同じ時間のタイムスタンプの塊があった場合、1つだけ残して残りはデータから取り除くといった処理を施した。

日付ごとのデータに対して、それぞれ最尤法を行いパラメータを推定した結果、次の表のようになった。

表 2: パラメーターの推定結果と各指標の値

日付	約定回数	λ_0	α	β	$E[\lambda]$	AIC	branchingRatio	$N_{rsponce}$	halfLife
2016/6/23	19736	0.0719	0.190	0.277	0.229	-59.5	0.685	2.18	2.50
2016/6/24	18448	0.0632	0.127	0.181	0.214	-58.7	0.704	2.38	3.83
2016/6/25	10765	0.0528	0.102	0.177	0.125	-42.6	0.576	1.36	3.92
2016/6/26	9063	0.0322	0.102	0.147	0.105	-35.9	0.693	2.26	4.70
2016/6/27	8663	0.0347	0.150	0.229	0.100	-34.2	0.655	1.90	3.02
2016/6/28	8722	0.0311	0.147	0.212	0.101	-33.9	0.692	2.25	3.27
2016/6/29	13572	0.0750	0.324	0.630	0.154	-46.8	0.515	1.06	1.10

数値結果を解釈する前に、いくつかの指標について説明しなければならない。それらの指標については Huffine[23] を参考にした。branchingRatio は以下のように定義される量である。

$$\text{branchingRatio} = \frac{\alpha}{\beta} \quad (22)$$

この指標が示す意味は、「過去の事象に起因して発生した事象の割合」である。Hawkes 過程に従って発生する系列は、まったく外部の影響から発生した immigrants と、すでに起こっている事象から生み出される offsprings のどちらかに分類することができるという考え方*4がある。branchingRatio は全体の事象のうち、offsprings はどれだけいるのかを表す割合なのである。

$N_{rsponce}$ は

$$N_{rsponce} = \frac{\alpha}{\beta - \alpha} \quad (23)$$

のように定義される量で、一回のジャンプによって起因される事象の数の期待値を表している。また、halfLife は

$$\text{halfLife} = \frac{\ln 2}{\beta} \quad (24)$$

*4 確率過程における分岐過程と同様な考え方で、Galton-Watson 過程などが有名。

のように定義され、ジャンプによって上がった強度が半分に減衰するまでの時間を表している。

以上を踏まえて表の数値に目を通すと、約定回数そのものが多く場が活発だった 6/23、6/24 はベースとなる強度 λ_0 の値が他に比べて大きく推定されている。また、6/23、6/24 を比べると 6/23 の方がジャンプ幅 α が大きい分 β も大きいので減衰は早い。従って、halfLife の値も 6/23 の方が小さい。6/25 から 6/28 はどの日も約定回数と $E[\lambda]$ の値は同じくらいであるが、6/25、6/26 と 6/27、6/28 とではジャンプ幅と減衰の度合いが微妙に異なっている。6/29 はベースとなる強度 λ_0 が一番高く、ジャンプ幅も最も大きい、 β も他と比べ著しく大きいため、約定回数や $E[\lambda]$ の値はあまり大きくならなかった。 $N_{response}$ の値も小さいため、一回のジャンプに対して他の約定が発生しづらかったと解釈できる。branchingRatio の値を比較しても分かるが、おおむね 6 から 7 割ほどの値をとるのに対し、6/29 は 5 割ほどにとどまった。以上により、日付ごとに約定系列の特徴があり、同じ約定回数でもその質は異なるということが確認できる。

また、ビットコインの特性という観点から表の数値を見てみると、branchingRatio が軒並み 5~7 割の値をとったというのはビットコイン為替市場ならではの特徴であると考えることができる。branchingRatio が 5~7 割ということはつまり、内因的に自己励起の効果によって発生した約定が全体の 5~7 割を占めるということであるが、これはかなり大きい割合である。ビットコインは発行主体を持たないので、その価値を保証する国などは存在しない。その点から、一般の通貨に比べて外因的影響を受けにくいと考えることができるので、branchingRatio がこのような大きな値を取り得ると考えられる。

次に、日付ごとに一変量 Hawkes 過程を当てはめて推定した結果を利用して、時間変更を行った後に、デュレーションが指数分布に従うかどうかを QQ プロットを用いて検証した。日付ごとの QQ プロットは次の図 5 のようになり、結果として、どの日付においても分布の裾の部分での当てはまりが良くないということが分かった。

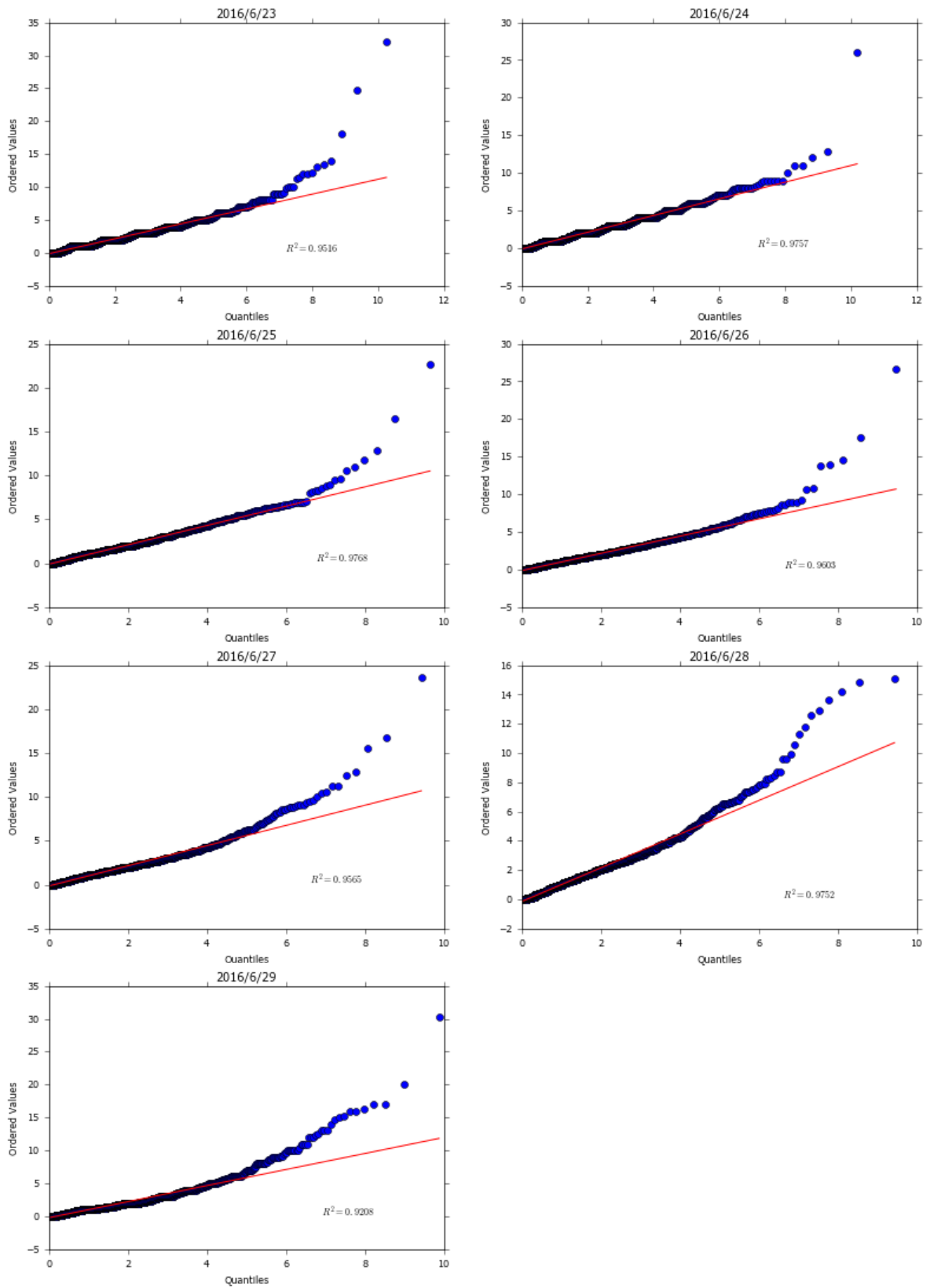
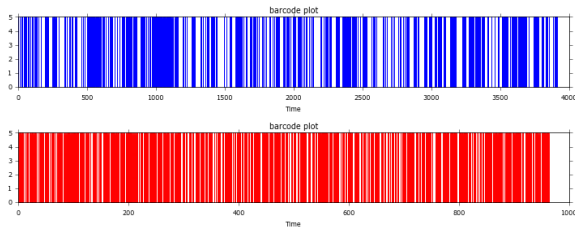


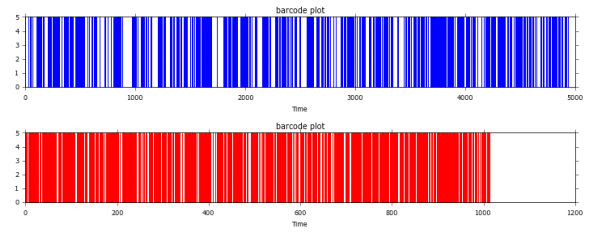
図 5: 2016/6/23~6/29 間のそれぞれの日付における QQ プロット

Embrechts et al.[24] で、Hawkes 過程の当てはまりの良さの検証法としてバーコードプロットを用いる方法が行われた。時間変更前と後のバーコードプロットを比較して、クラスターが消失したかどうかを見ることによって確かめる方法である。Hawkes 過程に従う系列に時間変更を施すとポアソン過程に従うはずなので、モデルが良く当てはまっていれば、時間変更前に存在したクラスターがなくなるはずであるという考え方である。

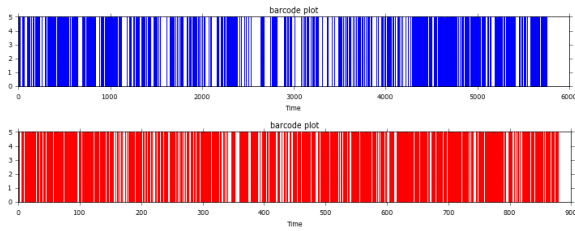
実際に、日付ごとに時間が早い方から 1000 個の約定系列を取り出しそれぞれバーコードプロットを描写した結果、次の図 6 のようになった。青が時間変更前で、赤が時間変更後である。



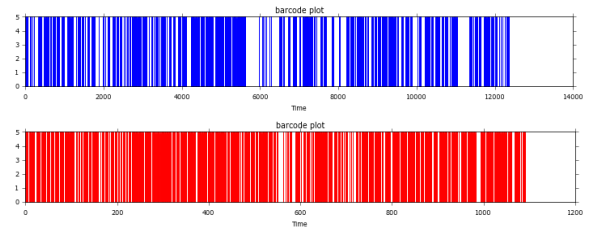
(a) 6/23



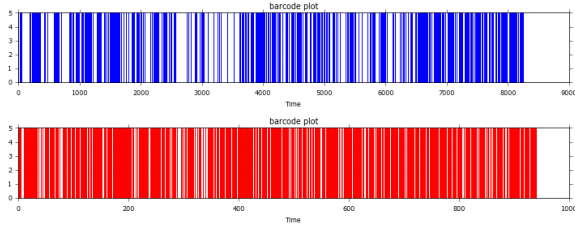
(b) 6/24



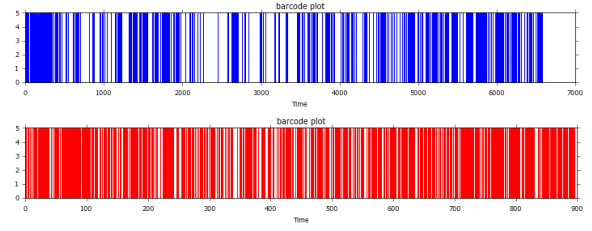
(c) 6/25



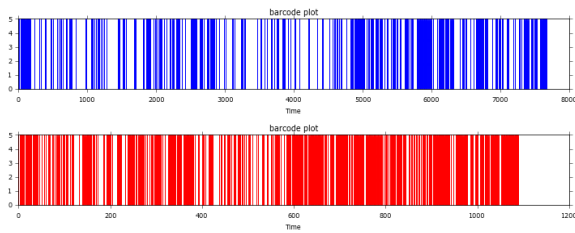
(d) 6/26



(e) 6/27



(f) 6/28



(g) 6/29

図 6: 2016/6/23~29 間のそれぞれの日付におけるバーコードプロットの比較

結果として、時間変更を行った後ではクラスターの数減っているが、6/28 や 6/29 などともと約定発生時間の偏りが激しい日付は、時間変更を行った後もポアソン過程に従って言えるとはいいがたいグラフとなっている。逆に、6/23 や 6/24 などともとの系列にクラスターがあまり存在していないような日付では時間変更がうまくいっているように見える。QQ プロットの結果も 6/23 や 6/24 が良いので、その結果とも適合している。

2016 年 6 月 23 日における推定パラメーターを利用し、2 章で紹介した Ogata のシミュレーション・アルゴリズムを実行して、1000 回分の約定系列をシミュレートすると次の図 7 のようになった。1000 回約定が発生するまで 3800 秒、つまり一時間ほど時間が掛かると見積ることができる。

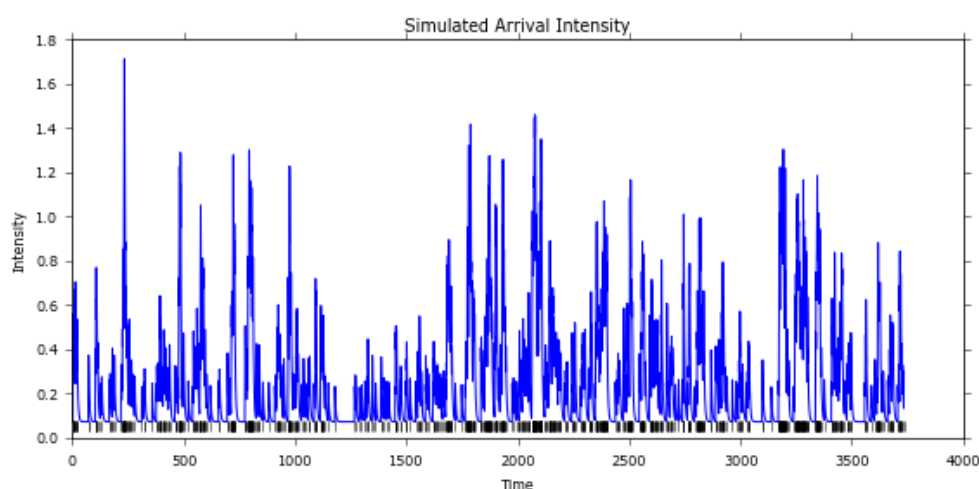


図 7: 2016 年 6 月 23 日の推定結果からのシミュレーション

5 結論

今回の研究では、一変量の Hawkes 過程を用いて BTC/JPY 為替市場の約定データを分析したが、最も単純な一変量 Hawkes 過程では約定系列を完全にモデル化できているとは言い難い結果となった。やはり、変数を増やして系列間の交互作用まで表現できるようにしたり非線形の Hawkes 過程を考えたりなど、モデルを複雑化していかなければより良いモデリングは難しい。しかし、一変量 Hawkes 過程というシンプルなモデルを扱ったがゆえに、結果の解釈が容易であったおかげで、branchingRatio が 5~7 割ほどだったという事実を実証でき、外因的影響を受けることが FX などに比べて少ないビットコイン為替市場ならではの結果であるという重要な示唆を得られた。今後の研究の展望としては、多変量の Hawkes 過程を取り扱うときに生じる計算上の問題について、Rasmussen[25] で研究されているようなベイズ分析的アプローチ対処することを考えたり、また今回はモデルの当てはめまでにとどまったが、推定結果から最適執行戦略を考えていくような研究も非常に有意義である。また今回利用したデータは約定データであったが、売り注文や買い注文など

の気配値を対象にして Hawkes 過程を当てはめて、注文板をモデリングするといった研究も今後頻繁に行われていくだろうと考えられる。

参考文献

- [1] Satoshi Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. (<https://bitcoin.org/bitcoin.pdf>).
- [2] Robert F Engle and Jeffrey R Russell. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *Journal of empirical finance*, Vol. 4, No. 2, pp. 187–212, 1997.
- [3] Robert F Engle and Jeffrey R Russell. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Source: Econometrica*, Vol. 66, No. 5, pp. 1127–1162, 1998.
- [4] Yoshihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, Vol. 83, No. 401, pp. 9–27, 1988.
- [5] Patricia Reynaud-Bouret and Sophie Schbath. Adaptive estimation for hawkes processes; Application to genome analysis. *Annals of Statistics*, 2010.
- [6] Don H. Johnson. Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, 1996.
- [7] Adrian Baddeley. Spatial Point Processes and their Applications.
- [8] G O Mohler, M B Short, P J Brantingham, F P Schoenberg, and G E Tita. Self-Exciting Point Process Modeling of Crime. 2011.
- [9] E Bacry, S Delattre, M Hoffmann, and J F Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, Vol. 13, No. 1, pp. 65–77, 2013.
- [10] Ioane Muni Toke. “Market Making{”} in an Order Book Model and Its Impact on the Spread. In *ECONOPHYSICS OF ORDER-DRIVEN MARKETS*, pp. 49–64, 2011.
- [11] Jeremy Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, Vol. 10, No. 1, pp. 1–25, 2007.
- [12] Ioane Muni Toke. An Introduction to Hawkes Processes with Applications to Finance Ioane Muni Toke (ECP -BNPP Chair) An Introduction to Hawkes Processes. 2011.
- [13] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy. Hawkes processes in finance. p. 48, 2015.
- [14] Patrick Hewlett. Clustering of order arrivals, price impact and trade path optimisation. *Workshop on Financial Modeling with Jump Processes*, 2006.
- [15] John Carlsson, Mao-Ching Foo, Hui-Huang Lee, and Howard Shek. High frequency trade prediction with bivariate Hawkes Process. Technical Report June, 2007.
- [16] Alan G Hawkes. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, Vol. 58, No. 1, pp. 83–90, 1971.

- [17] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.
- [18] Yoshihiko Ogata and Hirotugu Akaike. On Linear Intensity Models for Mixed Doubly Stochastic Poisson and Self- Exciting Point Processes. *Journal of the Royal Statistical Society. Series B {(Methodological)}*, Vol. 44, No. 1, pp. 102–107, 1982.
- [19] Yoshiko Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, Vol. 30, No. 1, pp. 243–261, 1978.
- [20] Alejandro Veen and Frederic P Schoenberg. Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, Vol. 103, No. 482, pp. 614–624, 2008.
- [21] E. Bacry, K. Dayri, and J. F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *European Physical Journal B*, 2012.
- [22] Scott W Linderman and Ryan P Adams. Discovering latent network structure in point process data. In *ICML*, pp. 1413–1421, 2014.
- [23] Carroll Huffine. High Frequency Final Project: The Hawkes Process. 2015. (<https://github.com/treyhuffine/Hawkes-Process/blob/master/Hawkes%20Process%20Report.pdf>).
- [24] Paul Embrechts, Thomas Liniger, Eth Zürich, and Lu Lin. NEW FRONTIERS IN APPLIED PROBABILITY MULTIVARIATE HAWKES PROCESSES: AN APPLICATION TO FINANCIAL DATA MULTIVARIATE HAWKES PROCESSES: AN APPLICATION TO FINANCIAL DATA. *J. Appl. Prob. Spec*, Vol. 48, pp. 367–378, 2011.
- [25] Jakob Gulddahl Rasmussen. Bayesian Inference for Hawkes Processes. *Methodology and Computing in Applied Probability*, Vol. 15, No. 3, pp. 623–642, 2013.