

# 高頻度金融時系列を用いた外国為替市場の大規模分析

京都大学大学院情報学研究科数理工学専攻

佐藤 彰洋\*

## 概要

外国為替市場の 24 種類の通貨からなる 46 種類の通貨ペアの通貨交換レートと注文行動頻度時系列を用い、市場参加者の行動特性をとらえる手法を提案する。提案手法として多変量時系列の類似性をスペクトル形状の類似性から計量するスペクトル距離に基づく手法と、メジアンフィルタを用いた時系列中のパルス的な変動箇所をとらえる手法について述べる。提案手法を用い、市場参加者の行動を類似性と同期性の観点から実証分析をおこなった。

The methods to quantify behavior of market participants in the foreign exchange market by using high-resolution data which include 46 kinds of currency pairs consisting of 24 kinds of currencies are proposed. As proposal methods spectral distances based on shapes of normalized power spectra and detection of impulsive variations in multiple time series by means of median filtering are considered. Empirical investigation from perspectives of both similarity of collective motions and synchronous singular behavior of market participants is carried out.

## 1 はじめに

Bateson の意味論に基づく情報の定義にしたがうと、情報とは変化を作る違い (a difference which makes a difference) として定義される [1, 2]. これはメッセージとそのメッセージを受け取る受け手の状態変化・応答の組みを考え、受け手の行動や状態に変化を作り得るメッセージ群に情報=意味があると考えられるものである。そして、情報の生成は情報の受け手の状態の変化、あるいは応答の変化を通じて観測される。

更に、マルチエージェント系において、過去のエージェント集団の状態変化や応答が現在の各エージェントへのメッセージとなることによりメッセージの生起と状態変化の連鎖が生じうる。この現象は古来より、思い・言葉・行いの三業として、因果論の文脈で考察が深められている。そのため、我々の社会のようなマルチエージェント系において、エージェント集団の状態を高精度に計測することは、社会理解の精度向上に繋るため、強い関心が持たれている。

人間社会を物理的観点から考察するために、社会の構成単位である人の応答速度と記憶容量について考えてみる。人の情報処理速度は、脳内のニューロンパルスの発火間隔が約 1ms であることから、1ms 以下の応答は原理的に困難である。それゆえ人が識別できる時間の最小単位 (一瞬) が存在し、その時間は 50ms 程度と見積もられている [3]. このことは、連続的に静止画像を表示しているにもかかわらず、静止画の切りかわりが高速であると、あたかもそれらの静止画像が連続な動画であるように感じることや、交流電流により白熱灯が 50Hz-60Hz で点滅しているにもかかわらず、点灯していると感じることからも感覚的に理解される。更に、詳細に、視覚刺激から応答を行うまでの

\*京都市左京区吉田本町, TEL:075-753-5515, FAX:075-753-4919, E-mail:aki@i.kyoto-u.ac.jp, URL:http://amech.amp.i.kyoto-u.ac.jp/~aki/

最小時間を解剖学的に見積った研究によると、人間の最小応答時間は 150ms 程度であると考えられている [3].

人間の記憶容量が有限であることは脳細胞数が極めて多いながらも有限であることから予想できる。記憶容量の上限については、研究者により様々な見積りがおこなわれている。たとえば、von Neumann は 28PBytes, McChloch は 1-100GBytes を提案している [3].

一方、現在日常的に使用されているパソコン (Core2Duo, CPU2.6GHz) の処理速度が 15GFLOPS 程度 (浮動小数点の足し算や引き算を 1 秒間に 150 億回可能) であり、巨大磁気抵抗 (Giant Magneto Resistance; GMR) 技術の発展により (2008 年 7 月現在 600Gbits 平方インチを越えている), 1TBytes を越える記憶容量を持つ HDD が普及段階となっている。そのため、現在のパソコン程度のコンピュータを用いても、人間の一生 ( $10^9$  s) で実行可能な操作 ( $6 \times 10^9$  operations) を越える演算や、人間の一生内に消費不可能な文章や音楽などのデジタルデータを蓄積することが可能となっている。

その為、これまで人間の処理能力では困難と思われてきた大量のデータ操作を日常的におこなうことができ、この仕組みが現代社会の高度な利便性と関係している。更に ICT 技術を用いることにより、大量の人間活動に関するデータの収集、蓄積ができるようになってきている。そのため、これらの計算環境とデータとを用い、これまで不可能と思われていた規模と解像度で人間集団の状態理解を行うことが原理的に可能である。

人間の社会的・経済的活動から生成される大量のデータを用いる学術分野として、情報学、経済学、物理学、社会学の間での学際的研究領域が形成されつつある。特に、情報学の分野では、年々社会に蓄積されるデジタルデータ量の爆発的な増加 (情報爆発) が認識されている。近年の予想では、世界全体のデジタルデータ生成量が 2010 年頃に 1ZBytes/年 ( $10^{21}$  Bytes) を越えると予想され、この情報爆発現象の概念に基づき研究がすすめられている [4]。また、物理学の分野では、Physics in Society [5] と呼ばれる経済・社会現象におけるデータ生成の物理的過程を考慮した実証的な社会研究分野が形成されつつある。

本稿では、金融市場をエージェント行動の結果生じるメッセージに基づき、他のエージェントが逐次的に行動を生起する、メッセージと状態変化の生成連鎖が生じている体系であると仮定する。そして、エージェント状態 (行動) の時間発展と行動のタイミングを定量化することにより、メッセージ (= 意味) の生起の様子をとらえることを目的とする。特に、金融市場として、外国為替市場を取り上げ、市場参加者行動の時間発展をエージェントの行動タイミングの類似性と特異的な行動の同期性および、レート変化時系列の類似性を通じて調べる。

本原稿は以下の構成となっている。2 節において、分析に用いる外国為替市場の高解像度データについて述べる。3 節において、エージェント行動の時間発展を観測、定量化するための数的手法を述べる。4 節において、メジアンフィルタを用いた注文行動頻度の急激な変化箇所をとらえる方法と、注文行動頻度の特異的は変化の同期性指標について述べる。5 節はまとめと今後の課題である。

## 2 外国為替市場

外国為替市場は、世界でもっとも大きな金融市場であり、各国の通貨を交換するために、ブローカ、トレーダ、投資家から形成されるネットワーク構造を形成している。近年の情報通信技術の進歩によって、高解像度かつ網羅的な取引に関するデータを利用することができるようになってきている。

本分析では、CQG 社から購入した、Time & Sales (T&S) データと呼ばれる外国為替市場で取引される通貨に関する気配値提示データを用いた。分析に使用したデータには 24 種類<sup>1</sup>の通貨からなる

<sup>1</sup>NZD: ニュージーランド・ドル, AUD: オーストラリア・ドル, JPY: 日本・円, KRW: 韓国・ウォン, SGD: シンガポール・ドル, HKD: 香港・ドル, THB: タイ・バーツ, PRK: パキスタン・ルピー, INR: インド・ルピー, TRY: トルコ・リラ,

46種類の通貨ペア<sup>2</sup>に関する注文価格、売り/買いの別を含む1分間解像度のタイムスタンプを1レコードとするデータが含まれている。2007年6月1日から2007年7月31までのデータ期間内(土日を除く43日間)に含まれるコンマ区切り形式のテキストファイルのレコード総数は283,098,833 records, 総ファイルサイズは12,036,977,108Bytes(約12GB)であった。テキストファイルとして記録されている約2億8千万行の文字列に対する操作は人の一生である $6 \times 10^9$  operationsを越えているため、人力でこれを操作することは不可能である。そのため、何らかの数理的記述を用い機械的に定量化をおこなう必要がある。ここでは、時間区間毎での注文行動の生起総数と通貨交換レートに着目して分析を行う。

注文行動頻度と交換レートは以下のように各通貨ペアに対して抽出した。 $A_j(k)$  ( $j = 1, \dots, M; k = 0, 1, \dots, Q - 1$ )を期間 $[k\Delta t, (k+1)\Delta t]$ に計量される通貨ペア $j$ の注文数とする。ここで、外国為替市場においては、two-way quotationと呼ばれる売値と買値を同時に提示する価格注文方式が利用されることが多いので、売り注文と買い注文の数はほぼ同数である。そのため、ここでは売り注文数を注文数とみなした。また、通貨ペア $a/b$ の通貨交換レートは第一通貨 $a$ 1単位を第二通貨 $b$ の尺度で買う(Bid)/売る(Ask)値段として定義される。ここでは、期間 $[k\Delta t, (k+1)\Delta t]$ 内で見い出された最小の売値(best ask)を交換レート

$$b_j(k) = \min_{k\Delta t \leq t < (k+1)\Delta t} [ask_j(t)] \quad (j = 1, \dots, M; k = 0, 1, 2, \dots, Q - 1), \quad (1)$$

$b_j(k)$ と定義した。ここで $ask_j(t)$ は時刻 $t$ において注文された通貨ペア $j$ の売り値を表わす。もし、 $A_j(k) = 0$ ならば $b_j(k) = b_j(k-1)$ とした。この補間方法はいくつかの金融市場において実際に市場価格を知らせるために利用されている方法である。図. 1に $\Delta t = 1[\text{min}]$ としたときの、2007年6月18日から22日までのUSJ/JPYの注文行動頻度とベストアスクレートの時系列を示す。1週間で注文回数に5つの山を認めることができる。この山は地球の自転1回分(24時間)と対応している。また、ベストアスクレートの変化はランダムに見える動きをしている。

### 3 スペクトル距離

#### 3.1 スペクトル距離とは

ここでは、多変量時系列の類似度を計量する方法として、規格化パワースペクトルの形状類似性に着目したスペクトル距離による方法を提案する[6]。

時系列を生成する確率力学の詳細なモデル化を行うことなく時系列を生成する過程がどれほど似ているかを評価したい場合がある。例えば、簡単な例として平均0のノイズ $\xi_1(t)$ と $\xi_2(t)$ に駆動される2つのAR(1)モデル

$$x_1(t) = a_1 x_1(t-1) + \xi_1(t), \quad (2)$$

$$x_2(t) = a_2 x_2(t-1) + \xi_2(t), \quad (3)$$

ZAR:南アフリカ・ランド, CZK:チェコ・コルナ, PLN:ポーランド・ズォティ, HUF:ハンガリー・フォリント, SEK:ウェーデン・クローナ, NOK:ノルウェー・クローネ, DKK:デンマーク・クローネ, EUR:欧州・ユーロ, CHF:スイス・フラン, GBP:イギリス・ポンド, CAD:カナダ・ドル, MXN:メキシコ・ペソ, BRL:ブラジル・レアル, USD:アメリカ合衆国・ドル(表記はISO4217に基づく)を含む。

<sup>2</sup>以下の通貨ペアの分析をおこなった: AUD/CHF, AUD/JPY, AUD/NZD, AUD/SGD, AUD/USD, CAD/CHF, CAD/JPY, CHF/JPY, EUR/AUD, EUR/CHF, EUR/CZK, EUR/GBP, EUR/HUF, EUR/JPY, EUR/NOK, EUR/PLN, EUR/SEK, EUR/TRL, EUR/USD, GBP/AUD, GBP/CAD, GBP/CHF, GBP/INR, GBP/JPY, GBP/NZD, GBP/USD, NOK/SEK, NZD/USD, USD/BRL, USD/CAD, USD/CHF, USD/DKK, USD/HKD, USD/HUF, USD/INR, USD/JPY, USD/KRW, USD/MXN, USD/NOK, USD/PKR, USD/PLN, USD/SEK, USD/SGD, USD/THB, USD/TRL, USD/ZAR.

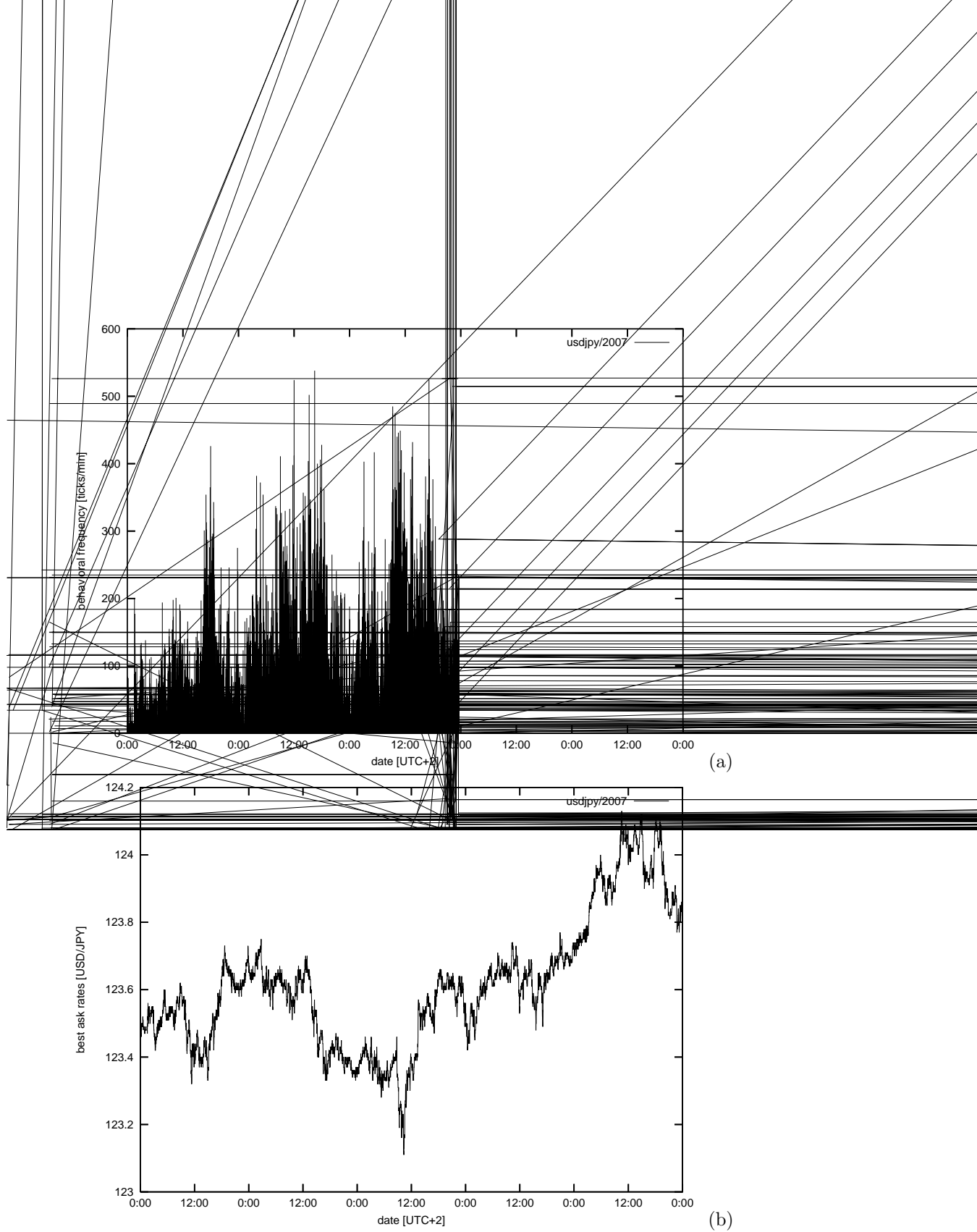


図 1: 2007 年 6 月 18 日から 2007 年 6 月 22 日までの USD/JPY の注文行動頻度 (a) とベストアスクレート (b).  $\Delta t = 1[\text{min}]$ .

から生成される時系列のみを用いて、2つの AR(1) モデルの類似性、すなわちパラメータ  $a_1$  と  $a_2$  の類似性を知るという問題を考える。式 (2), 式 (3) の相互相関関数を求めると

$$C_{12}(0) = \frac{\langle x_1 x_2 \rangle}{\sqrt{\langle x_1^2 \rangle} \sqrt{\langle x_2^2 \rangle}} = \frac{\langle \xi_1 \xi_2 \rangle}{\sqrt{\langle \xi_1^2 \rangle} \sqrt{\langle \xi_2^2 \rangle}} \frac{\sqrt{1-a_1^2} \sqrt{1-a_2^2}}{1-a_1 a_2}, \quad (4)$$

となり、ノイズの相互相関関数の値に依存していることがわかる。 $a_1 = a_2$  ならば  $C_{12}(0) = \frac{\langle \xi_1 \xi_2 \rangle}{\sqrt{\langle \xi_1^2 \rangle}}$

$H(p_j, t)$  が大きな値を取るとき、時系列は白色ノイズに近く多くの周波数成分を均等に含む時系列であることを意味する。反対に、 $H(p_j, t)$  が小さな値を取るとき、時系列は少数の周波数の正弦波によりよく近似できる。そのため、規格化パワースペクトルに対して、確率密度関数を用いた情報理論のアナロジーが成立する。

ここでは、パワースペクトルの形状を比較する方法として、情報理論で研究が深められている、Jensen-Shannon divergence(JS) と Kullback-Leibler divergence(KL) を用いる。Jensen-Shannon divergence は Lin により提案された divergence であり、Shannon entropy を用いて定義される複数の分布関数の総合的類似度を計量することができる divergence である [8]。ここでは、多変量時系列のパワースペクトル形状の総合的な類似性を比較するために利用する。時系列の注目度を定める重み  $\sum_{j=1}^M \pi_j = 1$  かつ  $\pi_j > 0$  なる  $\pi_j$  を考える。JS は式 (10) を用い、

$$JS(t) = H\left(\sum_{j=1}^M \pi_j p_j, t\right) - \sum_{j=1}^M \pi_j H(p_j, t), \quad (11)$$

と定義される。特に、全ての時系列に対して同じ重みで計量する場合は、 $\pi_j = 1/M$  と取る。この場合、 $JS(t)$  は全ての  $p_j$  が一致したときのみ 0 となる。そのため、Jensen-Shannon divergence を用いたスペクトル距離はダイナミクスの類似性を総合的に評価する量となることが期待される。別のスペクトル距離として、Kullback-Leibler divergence を用いた方法が提案されている [9, 10]。

$$KL_{lm}(t) = \sum_{n=1}^{N-1} p_l(f_n, t) \log \frac{p_l(f_n, t)}{p_m(f_n, t)}. \quad (12)$$

$KL$  の定義から、 $p_l(f_n, t) = p_m(f_n, t)$  が全ての  $f_n$  に対して成立する場合にのみ  $KL_{lm}(t) = 0$  となる。

### 3.3 実証分析

スペクトル距離を用い、注文行動回数および、価格変動のスペクトル分析を 46 通貨ペアに対して行った ( $M = 46$ ,  $\Delta t = 1[\text{min}]$ )。窓関数の幅は、人間の活動時間の標準である  $N = 480[\text{min}] = 8[\text{h}]$  とした。図. 2 と図. 3 に 2007 年 6 月 18 日から 22 日までの JS の時間的変化と、KL のスナップショットを示す。地球の自転に対応して JS の時間的に変動する様子を確認することができる。更に、KL は通貨ペアごとに類似性の高い通貨ペアと類似性の低い通貨ペアの組が存在することを確認することができる。注文行動頻度およびベストアスクレートの類似・非類似の組は時間的に変化していることが確認される。

## 4 時系列のパルスの変化の同期性

図. 1(a) の注文行動頻度時系列を見ると、注文行動頻度がしばしば急激に変化する場合があることに気がつく。注文行動頻度が急激に上昇することは、多くの市場参加者がほぼ同じタイミングで売りや買いの注文を提示したことを意味するので、このような注文行動頻度の急激な上昇が外国為替市場全体にわたってどのようになっているかを知ることが、市場参加者をとりまく環境状態の変化のなかで特に多くの市場参加者に同時にメッセージとなった変化の発生を知ることにつながる。環境変化をとらえる上で有益な知見をもたらすと期待される。

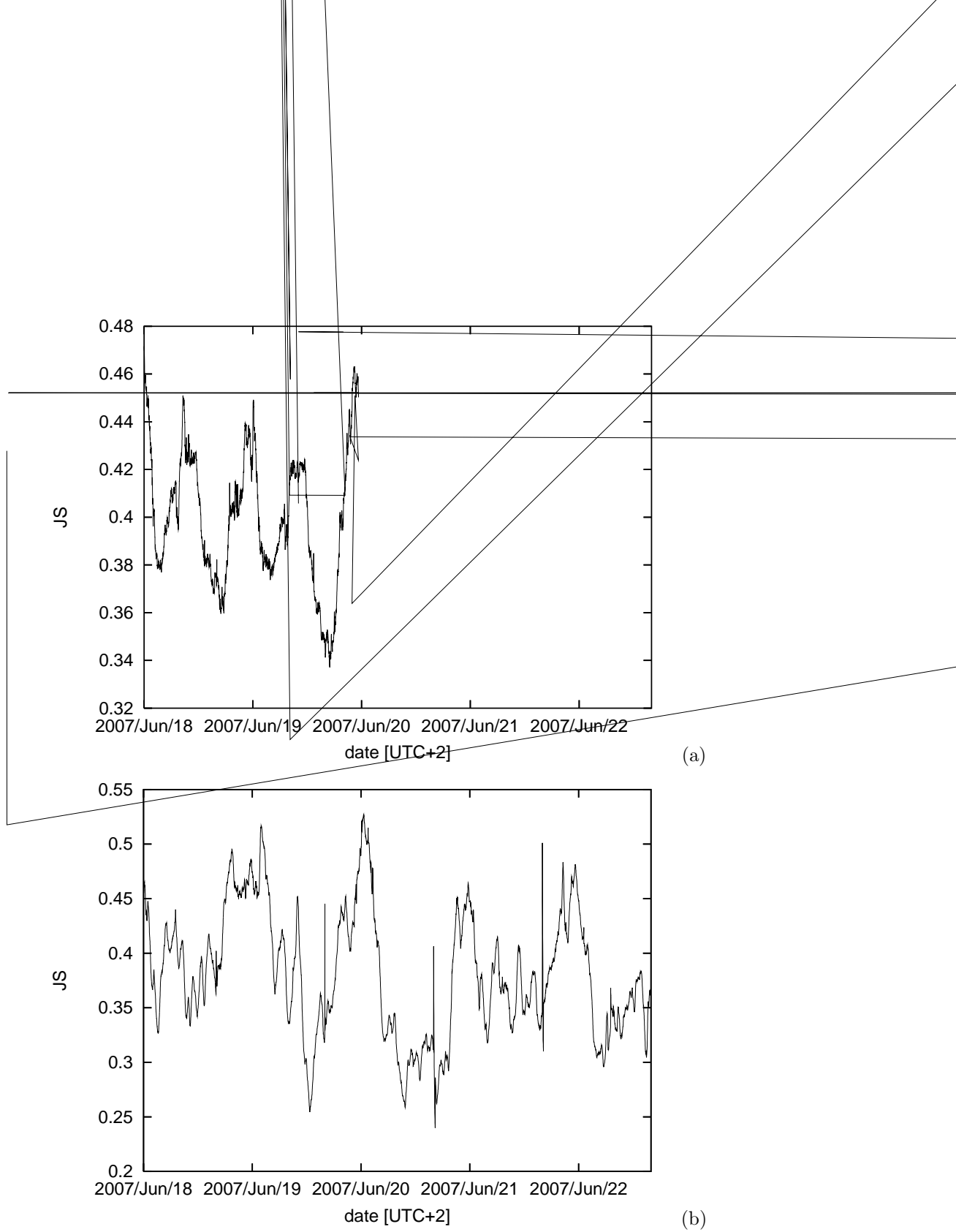
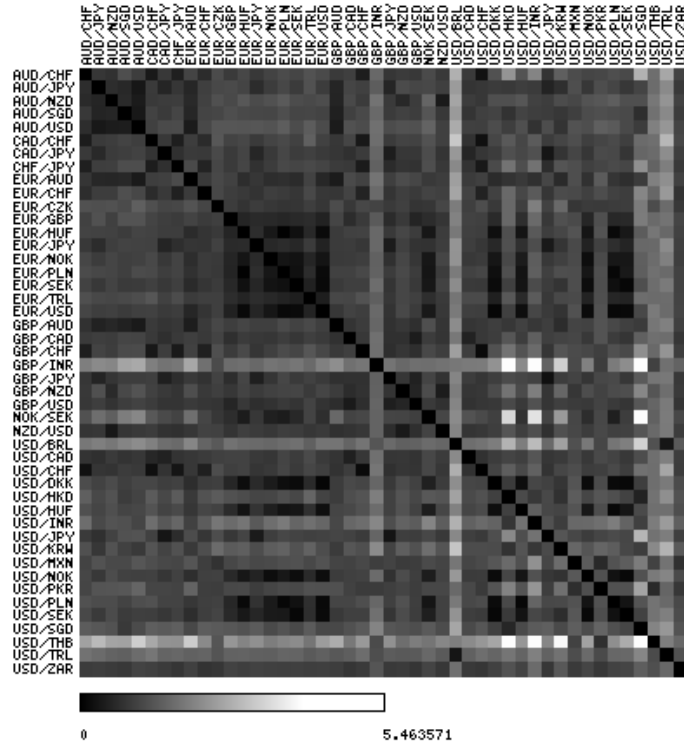


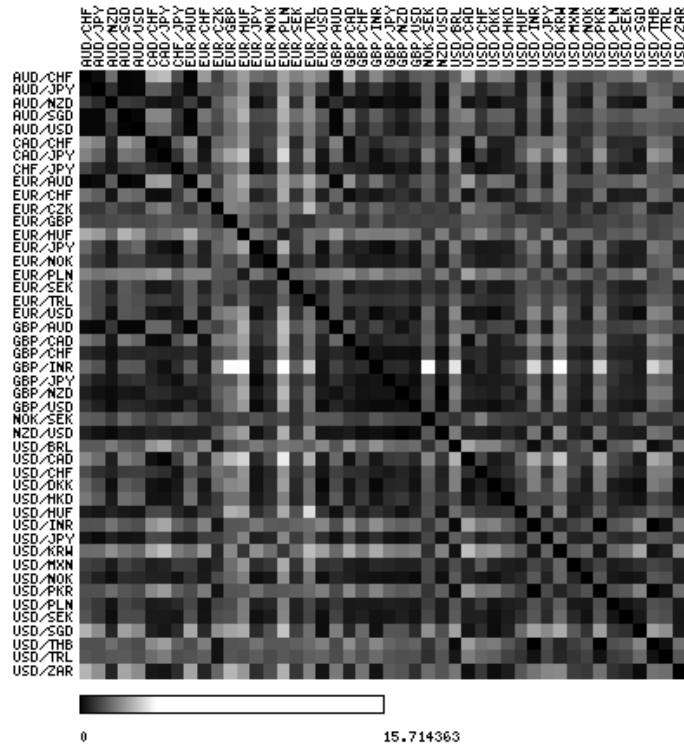
図 2: 2007 年 6 月 18 日から 22 日までの 46 通貨ペアの注文回数 (a) とベストアスクレート (b) を用いて計算した Jensen-Shannon spectral divergence の変化.  $\Delta t = 1[\text{min}]$ , 窓関数の幅は  $N = 480[\text{min}]$  として計算を行った.

2007/06/18 00:00(UTC+2) : Asia (A) : SKL



(a)

2007/06/18 00:00(UTC+2) : Asia (A) : SKL



(b)

図 3: 2007 年 6 月 18 日 0:00-8:00(UTC+2) における 46 通貨ペアの注文回数 (a) とベストアスクレート (b) を用いて計算した Kulback-Leiber spectral divergence のスナップショット.  $\Delta t = 1[\text{min}]$ , 窓関数の幅  $N = 480[\text{min}]$  として計算を行った.



ここでは、注文行動頻度の急激な上昇に着目し、これを時系列から検出する数理的方法を提案する [11]. 基本的アイデアは、時系列からメジアンフィルタを用いインパルス的な変動を除去した補完的時系列を作成して、その補完的時系列と元時系列との差が閾値を越えた時刻にインパルス状の変化が存在しているとするものである.

#### 4.1 メジアンフィルター

インパルス性雑音の除去方法はこれまで画像処理や音声処理の分野において長く研究されてきた [12, 13, 14]. 特にメジアンフィルタを用いる方法は、計算量が少なく、性能がよいことが知られている [13, 14]. メジアンフィルタとは対象とする時系列の中間値を出力する順序統計フィルター的一种である. 時系列  $x(k)$  に対してメジアンフィルタは以下で定義される.

$$\tilde{x}(k) = \text{Median}_{k' \in [k-K, k+K]}(x(k')) \quad (K > 0), \quad (13)$$

ここで、 $\text{Median}_{k' \in [k-K, k+K]}(x(k'))$  は区間  $[k-K, k+K]$  の間で  $x(k')$  の値の大きい順序に並べたときの中間の値を意味する.

#### 4.2 パルス位置の検出

メジアンフィルタを用いるとインパルス的に急激に値が変化する箇所は除去される. 一方、インパルス的ではない  $k$  における  $\tilde{x}(k)$  と  $x(k)$  はほとんど変わらないので、もとの時系列とインパルス性雑音を除去した時系列との差  $x(k) - \tilde{x}(k)$  を調べることで、インパルス性の急激な変動を捉えることができる. すなわち、

$$y(k) = |x(k) - \tilde{x}(k)|, \quad (14)$$

と定義すると、適当な  $\theta > 0$  に対して、 $y(k) \geq \theta$  が成立する  $k$  をインパルスの存在する箇所と判定できる. インパルスと判定する閾値には任意性があるが、ここで閾値  $\theta$  は  $y(k)$  の平均値  $m$  からのずれが標準偏差  $\sigma$  の  $D$  倍以上である場合に特異的な値であるとみなすこととし、閾値を

$$\theta = m + D\sigma \quad (D > 0), \quad (15)$$

とした.  $D$  を変化させてインパルス位置を検出した結果、 $D$  は着目している特異性の度合いに応じた観測者により指定される量である. ここでは、 $D = 5.0$  とし、特に大きなインパルス状変動に着目した.  $D = 5.0$  の場合に 46 通貨ペアに対する時系列を用いて推定した注文行動頻度のパルス的な変化が見出される時刻を図 4(a) に示す. 多くの通貨ペアに対して、注文行動頻度が同期してパルス的に急上昇する事象が起りえることが確認される.

#### 4.3 同期の定量化

$M$  種類の通貨ペアのうち  $j$  番目の通貨ペアに対して、インパルス状の注文行動頻度の急上昇が検出されたとき  $z_j(k) = 1$ , 検出されなかったとき  $z_j(k) = 0$  とする. このとき、

$$q(k) = \frac{1}{M} \sum_{j=1}^M z_j(k), \quad (16)$$

は金融市場全体での同期の指標 ( $0 \leq q(k) \leq 1$ ) となりえる。  $q(k)$  が大きい場合、多くの通貨ペアに対して同期して注文行動頻度に対してインパルス状の急変動が生じていると解釈することができる。 4(b) に 2007 年 6 月 18 日から 22 日までの 1 週間に対して同期の指標の時系列を示す。  $q(k)$  が大きな値を取る前後では、  $q(k)$  のゆらぎが大きくなる傾向が持続することが確認される。 これは、少数の通貨ペアに対して発生した市場参加者のパルスの注文行動変化の同期が、市場全体に波及していく様子および収束していく様子に対応していると考えられる。 観測期間内で、最大 80%以上の通貨ペアに対して同期して、注文行動のインパルス的な変化が生じたことが確認された。

## 5 まとめと今後の課題

外国為替市場の注文行動とベストアスクレートの通貨全体にわたる生起の様子を時系列間のスペクトル形状の類似性から定量化、可視化する方法の提案をおこなった。 注文行動頻度間の類似性とレート間の類似性は時間的に変動していることがわかった。 スペクトル距離の定義から、注文行動頻度とレートの類似性の変化は、これらを生成するダイナミクスを決める市場参加者行動の時間変化と関係していると推察される。

市場参加者の行動変化が生じる原因は、Bateson 的な情報の観点から推察すると市場参加者をとりまく環境状態の変化と深く関係していると考えられる。 そのため、市場参加者の状態変化を注意深く計量することは、社会環境に対するある種の高精度なセンサーとして利用できる。

更に、メジアンフィルタを用いた時系列のパルス状の変化箇所をとらえる方法を提案した。 提案手法を用いることにより、注文行動頻度のパルス的な急上昇の同期が市場全体にわたり大規模に生じる場合があることがわかった。 注文行動頻度の急激な上昇が同期して生じる場合、多くの市場参加者の行動がその瞬間に同調したことを意味するので、このような同調がどのような環境状態の変化に起因するものかを調べることは、金融市場の特性を理解する手掛かりになると期待される。 今後の課題として、注文行動頻度のパルス的な変化の同期と通貨交換レートの同調との関係を調べることで、金融市場参加者へのアンケート調査を通じた、現実の市場理解との対応付けが挙げられる。

本研究を深めることにより、網羅的かつ高解像度の外国為替市場参加者の活動の様子を通じ、世界経済の状態理解を深める技術につながることを期待される。 そして、この技術が、我々をとりまく世界状態の高精度理解を通じ、意思決定者集団に世界の兆しを伝える技術に発展していくことを切望する。

## 謝辞

本研究は京都大学 GCOE プログラム「知識循環社会のための情報学教育研究拠点」の部分的援助を受けて行われた。

## 参考文献

- [1] G. Bateson, *Steps to Ecology of Mind*, University of Chicago Press, Chicago, 2000 [Originally published in 1972].
- [2] Y. Neuman, *Information Sciences*, **176** (2006) 1435.
- [3] 大島正光編集, *アーゴノミクス*, 朝倉書店 (1965).

- [4] ” 情報爆発時代におけるわくわくする IT の創出を目指して”, 情報処理, **49** (2008) 879.
- [5] A. Carbone, G. Kaniadakis, and A.M. Scarfone, “Tails and Ties, Editorial: Topical Issue on Physics in Society”, The European Physical Journal B, **57**, (2007) 121–125.
- [6] A.-H. Sato, Physica A, **387** (2008) 3960.
- [7] B. Bein, *Best Practice & Research Clinical Anaesthesiology*, **20** (2006) 101.
- [8] J. Lin, IEEE Transactions on Information Theory, **37** (1991) 145.
- [9] A.-H. Sato, Physica A, **382** (2007) 258.
- [10] R. Veldhuis, E. Klabbers, IEEE Transactions on Speech and Audio Processing, **37** (1991) 145.
- [11] 佐藤 彰洋, “外国為替市場参加者行動の特異性検出方法: 特異性の伝播と同期”, 情報処理学会論文誌数理モデル化と応用, 採録.
- [12] 雛元 孝夫, 棟安 実治, 田口 亮, 非線形デジタル信号処理 朝倉書店 (1999).
- [13] J.W. Tukey, *Cong. Rec. EASCON'74* (1974) 674.
- [14] A. Asano, K. Itoh, and Y. Ichioka, Pattern Recognition, **23** (1990) 1059.

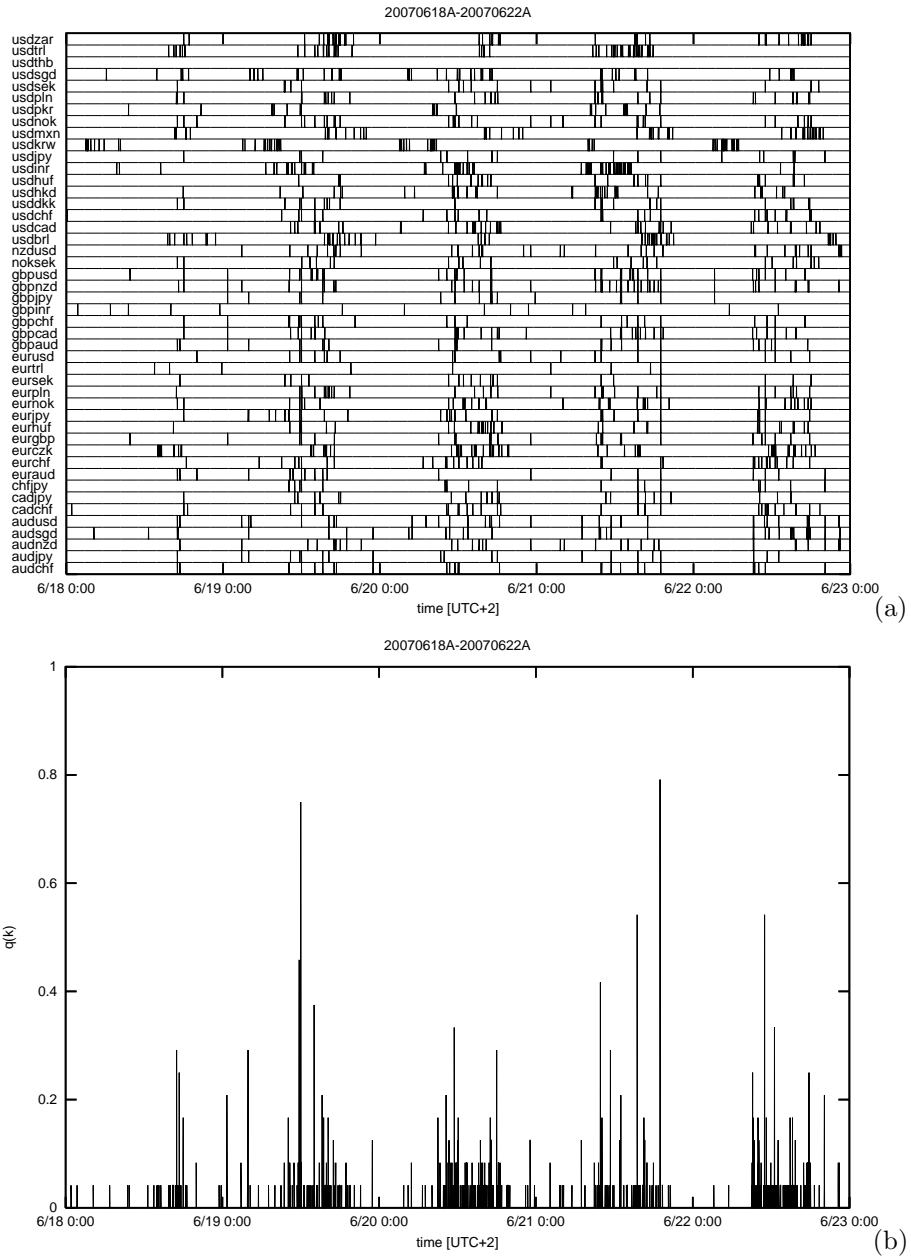


図 4: 2007 年 6 月 18 日から 22 日までの, 46 通貨ペアの 1 分間当りの注文回数に対するパルス的な変動とみなされる時刻 (a) と同期指標 (b).  $D = 5.0$  として計算を行った.