



日本取引所グループ
JAPAN EXCHANGE GROUP

JPX WORKING PAPER

JPXワーキング・ペーパー

国内高速3株式市場間の注文板形成の先行遅行関係分析

林 高樹

March 31, 2016

Vol. 13

備考

JPX ワーキング・ペーパーは、株式会社日本取引所グループ及びその子会社・関連会社（以下「日本取引所グループ等」という。）の役職員及び外部研究者による調査・研究の成果を取りまとめたものであり、学会、研究機関、市場関係者他、関連する方々から幅広くコメントを頂戴することを意図しております。なお、掲載されているペーパーの内容や意見は執筆者個人に属し、日本取引所グループ等の公式見解を示すものではありません。

国内高速 3 株式市場間の注文板形成の先行遅行関係分析

慶應義塾大学, JST-CREST 林 高樹*

ジャフイー・ジャーナル 2016 年 3 月号掲載予定論文

Abstract

本研究は、東京証券取引所(主市場)と2つの私設証券取引所-チャイエックス, ジャパンネクスト PTS-の国内 3 市場にて同時に取引されている現物株式について、注文板形成の先行遅行関係の存在やその大きさを実証的に調査することを目的とする。ステップ 1 では、高頻度注文板データを用いて先行遅行時間を計測する。まず、新規注文や取り消し、約定等のイベント発生順に不等間隔に並べられた注文板の情報を、各時点ごとに最良気配値の加重平均値である“マイクロ・プライス”に集約する、次に、3 市場におけるこれらのマイクロ・プライス系列の先行遅行時間を、Hoffman, et al.(10,13)による先行遅行関係推定法により“計測”する。これら個別銘柄の日々の計測値を要約することで市場間の先行遅行関係の全体的傾向を把握する。ステップ 2 では、個々の先行遅行時間計測値を多変量時系列データ(銘柄×データ期間)に配置し、パネル回帰分析を行い、銘柄に共通な特徴や相違点を抽出し先行遅行要因を探る。分析対象銘柄は、東京証券取引所において時価総額および流動性の特に高い 30 銘柄から成る TOPIX Core30 構成銘柄、分析期間は 2013 年 1 月 4 日から 2014 年 12 月 30 日(489 営業日)であり、東証における 2 回のティックサイズ変更のタイミングも含んでいる。

主な実証的発見は次の通りである。まず、3 市場間の先行遅行関係に関する平均的な傾向として、銘柄や時間帯による相違はあるものの、全般的に、東証が他二市場に先行する様子が見られた。これは、林(2015a)における代替的指標(“LLR”)によるものと異なった。2014 年の東証におけるティックサイズ変更は、全般に ChiX が他二市場に対して先行度合をやや高めた様子が観察された。次に、先行遅行関係を説明する変数として、ボラティリティやスプレッド率、約定枚数が相対的に高まる市場は、より先行度合が高まる傾向が見られた。その一方、気配数量(デプス)が相対的に高まる市場は逆に先行度合が低下(遅行度合が高まる)傾向が見られた。先行遅行時間が一日の中で変化する様子も観察された。

Keywords: 高頻度データ, 高頻度トレード, アルゴリズム取引, lead-lag 分析, Hayashi-Yoshida 推定量, マーケット・マイクロストラクチャ。

1 研究の背景

世界の株式市場において近年電子市場が相次いで設立され市場の分裂化(fragmentation)が進み市場間競争が加速する中、我が国では、東京証券取引所(東証)において arrowhead が 2010 年 1 月に導入され、一方、代替的市場として高速約定と刻み値の小ささを売りとする 2 大私設取引所(PTS)- ジャパンネクスト PTS(運営会社 SBI ジャパンネクスト証券)は 2008 年 10 月、チャイエックス(運営会社チャイエックス・ジャパン)は 2010 年 7 月- が相次いで運営を開始している。これら複数市場に対して最も価格の良い市場を見つけて顧客のために売買する SOR(Smart Order Routing)技術が普及拡大し、一方、日本証券クリアリング機構は PTS 取引を決済対象に加え(2010 年 7 月)、さらに PTS 取引が「5% ルール」適用除外とされる(2012 年 10 月)など、インフラ面、制度面でも代替的市場の利便性が高まっている。

* 慶應義塾大学大学院 経営管理研究科, 独立行政法人科学技術振興機構戦略的創造事業(JST-CREST). takaki@kbs.keio.ac.jp. 本論文は、証券アナリスト誌第 53 巻第 4 号(2015 年 4 月)に掲載された拙著論文“高頻度注文板データによる 2014 年東証ティックサイズ変更の国内株式市場への影響分析”(林(2015a))の姉妹論文であり、分析結果は相互補完的である。

このような中、これら国内3市場における注文板情報をリアルタイムで比較しながら収益機会を狙い、高速かつ高頻度で売買を行うHFTなどの市場参加者は相当数存在していると思われる。本研究は、国内株式市場間の短時間での連動性について焦点を当て、高頻度の注文板データから観察される現象について報告する。分析用データには市場参加者に関する情報が含まれていないため、例えば近年国内市場においても存在感を増しているHFTの行動を直接分析することはできないが、分析を通じて彼らの行動を間接的に理解する一助となることも期待される。

分析に用いるデータは、東証、チャイエックス、ジャパンネクストPTSの3市場より提供を受けた気配更新や約定を全て記録した高頻度注文板データで、タイムスタンプの時間解像度はミリ秒、データ期間は2013年1月4日～12月28日(245営業日)、および2014年1月6日～同12月30日(244営業日)、計489営業日である。当該期間は東証における2回のティックサイズ変更のタイミング(1月14日、7月22日)を含んでいる。分析対象銘柄は、東証において時価総額および流動性の特に高い30銘柄から成るTOPIX Core30構成銘柄である。

本研究では、特に3市場間の先行運行関係に興味がある。ベースとなる統計的方法論は、Hoffmann et al. (2013)による先行運行関係推定法である。日々の先行運行計測値よりデータ期間内中央値を計算し、市場間の先行運行関係の全体的傾向を把握する。さらに、個々に計測された先行運行時間を多変量時系列データ(銘柄×データ期間)に配置し、パネル回帰分析を行い、銘柄に共通な特徴や相違点を抽出し先行運行要因を探る。

Hoffmannらの方法論は提唱されて日が浅く、推定精度など今後理論的整備がなされていくべきものである。また、同じ枠組内のもう一つの先行運行指標による分析(林(2015a))とは異なる結果が得られている。本稿の分析結果は、あくまで暫定的なものであり、同論文の結果と共に手法の妥当性も含めて今後検証が必要である。^{*1}

2 国内高速3市場の概要およびデータセット

2.1 3市場の概要

本研究で分析する、東証、チャイエックス(以下、“ChiX”)とジャパンネクストPTS(以下、“JNX”)の3市場の概要を表1にまとめる(2013年末時点)。同表にあるように、マッチング・エンジンのスピードの差や、ティックサイズの違い等に加え、市場参加者が約定時に各市場に対して支払う取引手数料にも相違があることから、市場の参加者タイプの構成割合は市場間で異なっている可能性が高い。^{*2}

東証は、2014年1月14日と7月22日の2回にわたり、投資家の“利便性の向上”を目的として、“流動性が高い銘柄について細かい呼値の単位を導入することで、約定価格の改善や、指値注文における値段の選択肢の広がりによる板での順番待ちの緩和”を図った。^{*3} 第1回目の変更(フェーズ1)では対象銘柄群(TOPIX100構成銘柄)の内の呼値が3000円超に対して導入され、第2回目(フェーズ2)では対象銘柄群の残りの全て(呼値3000円以下)の銘柄にまで拡大された。その結果該当銘柄のティックサイズはPTS二市場と競合する水準にまで引き下げられた。東証ティックサイズ変更の影響については、林(2015a))による3市場間比較の他、近藤(2015)が東証にフォーカスした詳細な分析を行っている。

HFTの戦略には、注文板への指値注文を通じた流動性供給を行う“メイカー”戦略と、注文板にある指値注文に対して取引を図る需要する“テイカー”戦略とに大別される。前者はビッドとアスクの差(ビッド・アスク・スプレッド)が主な収益源であり、後者は価格変化が収益源である(大墳(2014))。前者の方がより高速・高頻度に注文行動を行うとされている。

^{*1} 現時点において本研究における計測・分析対象である(“マイクロ・プライス”をベースにした)注文板形成の先行度合と市場“クオリティ”の関連性は明確ではない。計測値の精度の検証も残されている中、今回の先行運行度合に関する分析結果が市場間の優劣を直接示唆するものではないことに注意されたい。

^{*2} PTS二市場の特徴については宇野(2012)に詳しい。

^{*3} 東証作成リーフレット(2013年12月1日付)より。

	東証 (内国株式)	ChiX	Japannext PTS
取引時間	前場: 9-11:30 後場: 12:30-15	8-16	デイトタイム: 8:20-16 ナイトタイム: 19-23:59
取扱銘柄	一部 1774 二部 559	約 3500	約 3500
ティックサイズ	≤ 3,000, 1 円 ≤ 5,000, 5 円 ≤ 30,000, 10 円, ...	≤ 5,000, 0.1 円 ≤ 100,000, 1 円 > 100,000, 10 円	東証の 1/10
取引手数料	基本料金 0.29bp*	テイカー 0.2bp メイカー 0bp	0.2 bp
スピード	1 ミリ秒	0.2 ミリ秒	0.4 ミリ秒
特殊注文条件	寄指/成, 引指/成 不成, IOC	FOK, IOC, Post-Only Mid Peg 等, アイスバーグ	FOK, IOC, Post-Only アイスバーグ

Table1 三市場の比較 (2013 年末時点). * 月間の総売買代金及び証券会社毎の売買代金に応じて割引かれる.

前者のタイプによる一連の取引行動に要する時間はミリ秒未満とも言われる。そのような高速の行動を一般の研究者が入手可能なデータの仕様, すなわち, 取引主体の ID が付与されてなく, かつ時間解像度がミリ秒以上の “低速” の高頻度データで把握することには無理がある。しかしながら, それら “超” 高速部分を除いた領域においては, 注文板データに反映される限りにおいて, より広い範囲での HFT 戦略やそれ以外の参加者による短時間の行動—例えば, SOR を介した注文の自動回送や東証気配に連動したペグ注文なども含めて—について, 適切なデータ分析を通じて間接的な理解が進むことが期待される。

高速化が進む我が国の株式市場に関する実証分析として, 宇野 (2012), 太田 (2013), また, 国内における HFT にフォーカスした実証分析として保坂 (2014) などがあるが, 市場内外の急激な変化に実証研究が追いついていないのが現状である。高頻度注文板データの分析を行う意義は大きい。

2.2 分析データセット

今回の分析にあたり, 3 市場から高頻度注文板データを入手した。分析データセットの時間解像度はいずれの市場ともミリ秒単位である。各市場とも, データに記録されているタイムスタンプは, 個別注文が各々のマッチング・エンジンに到着した時刻とされる。

ここでは, 東証一部銘柄の中で時価総額や流動性の特に高いものから構成される TOPIX Core30 構成銘柄 (2013 年 10 月末時点) を分析対象とする (以下, “Core30”)

- Core30 (分析対象群: 日本たばこ (2914)*, セブン&アイ (3382)*, 信越化学 (4063)*, 武田 (4502)*, アステラス製薬 (4503), 新日鉄住金 (5401), コマツ (6301), 日立 (6501), パナソニック (6752), ソニー (6758), デンソー (6902)*, ファナック (6954)*, 日産自動車 (7201), トヨタ自動車 (7203)*, ホンダ (7267)*, キヤノン (7751)*, 三井物産 (8031), 三菱商事 (8058), 三菱 UFJ FG (8306), 三井住友 FG (8316)*, みずほ FG (8411), 野村 (8604), 東京海上 (8766)*, 三井不動産 (8801)*, 三菱地所 (8802), JR 東日本 (9020)*, NTT (9433)*, KDDI (9433)*, NTT ドコモ (9437), ソフトバンク (9984)*. (* 印は, 2014 年 1 月 14 日 (フェーズ 1 適用開始日) における東証始値が 3000 円超となった銘柄.)

一方, 分析のための比較対照銘柄群として, 2013 年 1 月時点において日経平均採用銘柄ではあるが TOPIX100(2013 年 10 月末時点) に属しない銘柄群のうち, 分析期間中上場廃止等のイベントの生じなかった 105 銘柄の中から非

復元無作為抽出により 30 銘柄を選定した (便宜上, “JAFEE30” と命名する).^{*4}

- JAFEE30 (比較対照群): 大林組 (1802), 双日 (2768), キッコーマン (2801), 三越伊勢丹 (3099), 日清紡 (3105), 王子 HD(3861), 日産化学 (4021), 日本曹達 (4041), 宇部興産 (4208), 日本化薬 (4272), 中外製薬 (4519), 昭和シェル石油 (5002), T O T O (5332), 神戸製鋼所 (5406), 大平洋金属 (5541), 東洋製罐 (5901), 荏原製作所 (6361), N T N (6472), 富士電機 (6504), 安川電機 (6506), 明電舎 (6508), アルプス電気 (6770), カンオ計算機 (6952), 太陽誘電 (6976), 凸版印刷 (7911), 丸井 (8252), ユニーグループ (8270), 商船三井 (9104), NTT データ (9613), 東京ドーム (9681).

今回の分析対象である Core30 はいずれも, 2014 年の東証ティックサイズ変更の対象銘柄群 (2013 年 10 月末時点での TOPIX100 構成銘柄) に含まれているが, 他方 JAFEE30 は含まれていない. データ期間は, 2013 年 (245 営業日), および 2014 年 (244 営業日) の 2 年間計 489 営業日である.

PTS 二市場は昼休みがないが, 東証との先行遅行関係をみるため, 東証の立会時間中のデータのみを分析に使用する.

3 分析の方法論

3.1 マイクロ・プライスの使用

注文板上の売買両サイドの最良気配値および注文枚数より, 各気配レコードにおける “マイクロ・プライス” を計算する (例, Gatheral and Oomen (2010)).

$$M = \frac{B \cdot Q^A + A \cdot Q^B}{Q^A + Q^B},$$

但し, B (A) は最良買 (売) 気配値, Q^B (Q^A) は最良買 (売) 注文に対する注文枚数合計である.^{*5}

本研究は市場参加者の注文・キャンセルおよび売買行動の結果としての高速での注文板の変化に興味がある. マイクロ・プライスは, キャンセルを含めた市場参加者の動きを直接に反映することから, 本分析の目的に照らして適当と考える. 分析に用いる代替的な価格系列としては, 通常ファイナンス/マイクロストラクチャ研究においては, マイクロ・プライスではなく, 最良気配間の仲値 (単純平均) か約定価格を用いられるが, 仲値では売り買い 2 つの最良気配水準が変わらない中での注文量の増加・現象は反映されない. 結果的に価格変動のない気配更新 (“ゼロ・リターン”) データが増えてしまい, 注文板データに含まれている市場参加者の動きに関する情報を有効に活用できなくなる可能性が高い. また, 約定データにおいては, 気配更新データに比べデータ量が大幅に減少すること, 約定データにおいてはゼロ・リターン現象に加えて “ビッド・アスク・バウンス” 現象への考慮が必要となってくる. 一方, Gatheral and Oomen (2010) が指摘している通り, マイクロ・プライスを用いた方が価格系列が拡散過程に “見える” ことから, 本研究で採用する Hoffmann et al. (2013) の方法論の前提条件との整合性においても, マイクロ・プライスの採用が妥当と考える. なお, マイクロ・プライスは最良気配数量に対するマルコフモデルを仮定することで次の約定価格の “期待値” と解釈できることが Avellaneda et al. (2011) によって示された.

3.2 時系列データ間の先行遅行時間

2 変量の時系列データの先行・遅行関係を分析するのに標準的な道具は, 両者の相互相関係数 (cross-correlation) を計測することであろう. しかし, 不等間隔に並ぶティックデータにおいては, 2 変量時系列は非同期に観測されるため (nonsynchronicity), 同期観測データを前提とする標準的アプローチを用いようとするれば, データを補完した

^{*4} 抽出に際しては R 関数 `sample()` を使用. なお, 計算時間の節約のため, 銘柄数は Core30 と同じく 30 銘柄としたが, 分析目的上は同数に揃える必要はない.

^{*5} 林 (2015b,a)) では, Q^B (Q^A) として, 最良気配値を入れて内側から 3 つの価格板の注文数量合計を用いて, マイクロ・プライスを計算していた.

り一部を捨てるなどすることによって、2変量を一旦同期化 (synchronization) させねばならない。Hayashi and Yoshida (2005) は、同時相関 (先行遅行ラグがゼロ) の計測において、この非同期データの同期化がもたらすバイアスの存在に注意を払い、データの同期化を行わずに共分散/相関係数を推定するアプローチを提案した。彼らは、時系列データが拡散型確率過程の離散時点観測であるというケースにおいて、観測区間固定の下でサンプル頻度が増大する (高頻度観測) になるという状況において、彼らの共分散推定量 (以下、“HY 推定量”) が真の共分散に対する一致推定量となることを示した (後に、Hayashi and Yoshida (2008, 2011) にて拡張)。

Hoffmann et al. (2013)(以下、“HRY”) は、HY 推定量の持つ、2時系列のサンプル時点が非同期でも構わないという特性を活かして、HY 推定量をベースにした、2変量の先行・遅行関係を推定する方法を提案した。

以下、彼らの表記法に従う。観測区間 $[0, T]$ において、(非同期に) 離散のかつランダム時点に観測される2つの確率過程 X, Y とする。それぞれに対する第 i 番目の観測時点を S^i, T^i , 対応する小区間を $I^i = (S^{i-1}, S^i]$, $J^i = (T^{i-1}, T^i]$ と表す。さらに、小区間 J^i を“左側”(過去時点) に θ だけシフトした区間を $J_{-\theta}^i = (T^{i-1} - \theta, T^i - \theta]$ 等と表すこととする。さらに、 $Y_t = \tilde{Y}_{t-\theta^*}$ であるような潜在的な確率過程 \tilde{Y} が存在すると仮定する; すなわち、 \tilde{Y} はリアルタイムでは観測されず、真 (未知) のラグ θ^* を伴って観測されるとする。 $\theta^* > 0$ の場合は X が Y に先行する状況、 $\theta^* < 0$ の場合がその反対の状況に対応する。

この時、Hoffmann et al. (2013) は、 X と Y の間の真のラグ係数 θ に対するラグ推定量

$$\hat{\theta} := \arg \max |U^n(\theta)_T| \quad (1)$$

を提案した。ここで、 $U^n(\theta)_t$ は θ 時間シフト済み HY 推定量

$$U^n(\theta)_T := \sum_{\substack{i,j=1 \\ S^i \leq T}}^{\infty} (\Delta_i X)(\Delta_j Y) K_{\theta}^{i,j}, \quad (2)$$

但し、 $\Delta_i X = X_{S^i}^1 - X_{S^{i-1}}^1$, $\Delta_j Y = X_{T^j}^2 - X_{T^{j-1}}^2$ である。また、 $K_{\theta}^{i,j} = 1_{[I^i \cap J_{-\theta}^j \neq \emptyset]}$ は時間 θ シフト済み“HY 係数”である;

これは2つの区間 $I^i = (S^{i-1}, S^i]$ と $J_{-\theta}^j$ がオーバーラップするか否かを判定する指標関数である。彼らは、 X と Y がセミマルチンゲール過程である時、ある正則条件の下で、 $\hat{\theta}$ が観測小区間幅の最大値 (r_n) がゼロに縮小する状況で、真のラグ・パラメータ θ^* に収束する、すなわち、一致推定量であることを示した。

ここで、(2) の代わりに、相関係数を測って

$$\rho(\theta) := \frac{\sum_{i,j} (\Delta_i X)(\Delta_j Y) K_{\theta}^{i,j}}{\sqrt{\sum_i (\Delta_i X)^2 \sum_j (\Delta_j Y)^2}} \quad (3)$$

に対する最大化を行ってもよい。この $\rho(\theta)$ は、標準的な時系列解析における、等間隔同期観測された2次元時系列データに対する相互相関係数に対応する。

この方法論のアイデアの概念図を図1(左)に示す。直感的に説明すると次の通りである。いま市場1で実現された株価が、市場2において (未知の値) ラグ θ 秒後に実現されているとする。この時、市場2の時系列データをラグ候補値 (図中 $l = \theta_i$) だけ過去時点に引き戻してから、2つの時系列間の相関係数 $\rho(\theta_i)$ を計測する。 l を動かして色々な θ_i を取りながら $\rho(\theta_i)$ を計算していき、その中の最大値を取るような θ_i を、未知のラグ・パラメータ θ の推定値 $\hat{\theta}$ と定義するのである。

Huth and Abergel (2014) は、(3) をベースにした先行遅行分析を、仏国株式データに対して行った。彼らは、先行遅行の程度を評価する指標として LLR (lead-lag ratio) を提案した:

$$LLR := \frac{\sum_{i=1}^p \rho(\theta_i)^2}{\sum_{i=1}^p \rho(-\theta_i)^2}. \quad (4)$$

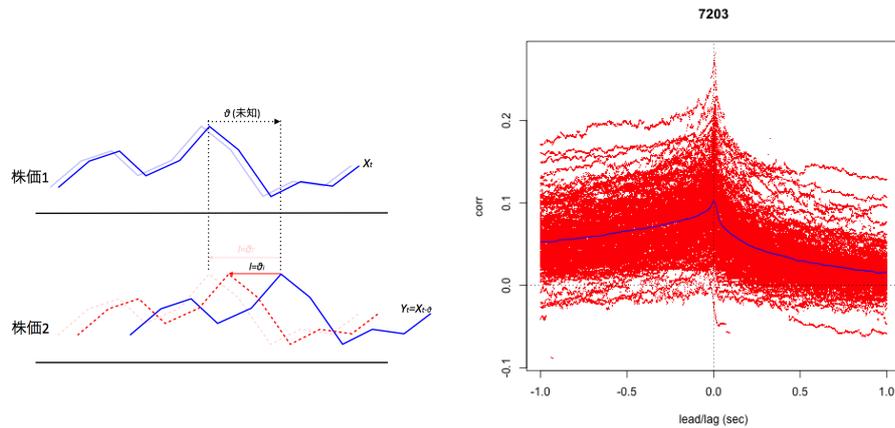


Figure1 (左) Hoffmann et al. (2013) の方法論のアイデア. (右) クロスコレログラムの例: トヨタ自動車 (7203), 東証-JNX, 2014年1月14日~同8月20日 (151営業日). 1日分が1本の点線に対応. 正の領域が東証先行, 負の領域がJNX先行. 実線は局所平均を取ったもの.

彼らは, (対数) 価格の増分に対する弱定常性の下では, X が Y に先行するための必要十分条件は $LLR > 1$ であると主張した. 背後にあるアイデアは, X が Y をその逆の場合よりもより正確に予測するならば, X は Y に先行するであろうというものである.

先行遅行指標として, $\hat{\theta}$, LLR には一長一短はある. Huth and Abergel (2014) らはより結果の安定的な LLR の利用を推奨している. 一方, LLR は物理的意味を持たずその使用は正当化されないと考える研究者もいる. 林 (2015a) において LLR (正確にはその対数值, $\ln(LLR)$) を用いた分析結果が示されたが, 本稿では $\hat{\theta}$ を用いた結果を掲載する.

Hoffmann et al. (2013) のフレームワークは提案されて日が浅く, 実際の市場データへの応用例は少ない. 今回のような同一銘柄の異市場間での先行遅行関係分析への応用は, 筆者の知る限り未だ行われていない. また, ベースとなる統計理論が未だ整備されていない状況の中, 本研究は実証的知見を得ることを目的としていることから, ここでは, X, Y に特定の確率過程モデルを仮定せず, 記述統計的立場を取りながら分析を進める.

以下においては, (東証, JNX, ChiX) の中から3組のペア $X-Y$ を作り, 各ペア毎に先行遅行分析を行う.

3.3 分析手順

HRY の方法論を適用するためには, 真の θ の推定値の候補となる有限個の離散集合を予め用意しなければならない. ここでは, 短時間での先行遅行関係に興味を持っていることから, ラグ-1秒から+1秒までの間の最小時間刻みのグリッド集合を用いる. すなわち, グリッド (先行遅行時間の候補値) の集合 $G = \{-1, -0.999, \dots, -0.001, 0, 0.001, \dots, 0.999, 1\}$ を取る (要素数 2,001). マイナス領域は Y の先行, プラス領域は X の先行を表す. 各点 $\theta_i \in G$ について, 相互相関係数 $\rho(\theta_i)$ を計算する. 参考までに, 図1(右)に, トヨタ自動車 (7203) について, 東証 (X)-JNX (Y) 間の $\rho(\theta_i)$ を計算し, 横軸 θ_i に対してプロットした例 (コレログラム) を示す. この例では, データ期間は2014年1月14日~8月20日 (151営業日), 下述のように1日を10個の時間帯に分割せずに, 1日分のデータを使って $\rho(\theta_i)$ を-1秒から1秒までの範囲で計算し, 一本の点線でつないで描いたものを, 全データ期間について重ねて表示している. 同図より, 各日とも概ねコレログラムのピーク, すなわち $\hat{\theta}$ が原点に近いところにあることが分かる. 両市場の先行遅行時間差は小さく, 極めて短時間で連動していることが伺える.

Hoffmann et al. (2013) によれば, データ点が多くなるにつれて統計量 $\hat{\theta}$ が真値 θ に近づく (一致性を持つ) ためには, グリッドの刻みの大きさを平均観測区間幅の大きさに比べて小さく取らねばならない. 一方, 彼らの Proposition 2 によれば, 一般には平均観測区間幅の大きさよりも精度良く推定することはできない. すなわち, 今回の高頻度データは, “高速” で連動する 3 市場間の先行運行時間を “精度良く” 計測するには十分に “高頻度” とは言えない. 誤差が大きく, そもそも難易度の高い “推定” の試みである. 今回の分析では, 記述的立場から次のような簡便的手法を取るが, サンプル数に制約がある中での “精度” の改善は今後の課題である.

筆者が予備的に行った単純なモデル (単一ラグ + ブラウン運動) を仮定したモンテカルロ・シミュレーション実験では, 両者のオーダーに相違がなければ, 独立なパスが多数得られる場合には, 各パス毎に得られる $\hat{\theta}$ の中心値を取るとそれが真の値に近いことが確認された. 今回の $\hat{\theta}$ は “ノイズ” レベルの高い (バラツキの大きい) 分布を持つことから, 標本平均の代わりに中央値 (メジアン) を用いることにする. もちろん, 現実の事象は一方の市場が他方の市場に対し一定時間で常に先行している等のような単純なものとは考えにくい, それでも今回のような記述統計量の大きさを報告することに意義はあると考える.

なお, HRY の方法論は提唱されて日が浅く, 今後理論的の整備がなされていくものである. 推定精度に関しては統計学上の研究課題であり, 今回の結果報告は暫定的なものである. 手法の妥当性も含めて今後検証が必要である.

ここでは異なる日のデータセットが互いに独立かつ同一にサンプルされたと見なす. これは, 市場参加者, 特に本研究で興味を持っている短期間で売買を繰り返す参加者の一日内の戦略が相場の (上下動に拘らず) 期間を通じて不変であるとの前提の下では, 妥当な仮定であろう. さらに, 短時間売買を繰り返す HFT はリスク回避のため一日の終了時には原則ポジションを取らないとされる. HFT などの短期売買主体の参加者の行動は 1 日内の時間帯によって変化することは想像に難くない. そこで, 1 日内の lead/lag 指数の変化を反映させるために, 東証の立会時間 (9:00-11:30, 12:30-15:00, 計 5 時間 = 300 分) を 30 分毎に 10 個の時間帯に分割し, 第 1 時間帯 (9-9:30), 第 2 時間帯 (9:30-10), ..., 第 10 時間帯 (2:30-3pm), のように設定する. それにともない, 表 1 で示したように PTS は昼休み等東証の立会時間外に取引可能な時間帯はあるが, 先行運行時間を見るという目的に照らし, 以上の 10 個の時間帯以外のレコードは全て分析対象から除去した.

その他に行った分析の前のデータ前処理は以下の通りである. まず, ティックデータのタイムスタンプはミリ秒刻みであるが, 同一タイムスタンプに複数の気配レコードある場合には, 最後尾のレコードのみを採用した. 次に, 同一市場内の直前の約定価格の上下 5% を超える気配値は, 市場実勢から乖離した注文であると見なし注文板から全て消去した. さらに, マイクロ・プライスの計算には最良気配が両サイドに提示されていることが必要なことから, 片側しか最良気配のないレコードは全て除去した.⁴⁶

実証分析は 2 つのステップから成る. ステップ 1 として, まず, 各銘柄別, 各日毎各時間帯毎に, 両市場のマイクロ・プライスより計算される対数収益率を用いて θ シフト済み相関係数 $\rho(\theta)$ を, G 上の全ての θ_i に対して計算する. 次に, 得られた $\rho(\theta_i)$ の値集合から, 各銘柄別, 各日毎各時間帯に $\hat{\theta}$ を計算し, 期間全体の分布の中央値を計算し, 散布図を作成することで全体的傾向を観察する. ステップ 2 として, ステップ 1 で得られた $\hat{\theta}$ を多変量時系列データ (銘柄 \times データ期間) として構成し, パネル回帰分析により, 先行運行時間の日次変動や銘柄固有の変動をコントロールしながら, 銘柄に共通な特徴や相違点を抽出し先行運行要因を探る.

上述のように HRY の方法論に関する統計理論は, 現時点において未整備の状況であるが, $\rho(\theta)$ の大きさの相对比较は可能であり, よって最大値を与える θ の値である lead/lag 指数 $\hat{\theta}$ について議論することは意味があると考

⁴⁶ したがって, 例えば, 直前の約定価格の 10% を超える幅のビット・アスク・スプレッドを持つレコードは, マイクロ・プライス系列の生成には使用されない. (ちなみに, 林 (2015b) においては, ビッド・アスク・スプレッドの幅自体に対して直接閾値を設定し “外れ値” の処理を行っていた.) フィルターを厳しく設定すると, 分析に使用可能な気配レコードが減少することになるが, 逆に緩過ぎると市場実勢から乖離したレコードが分析対象に含まれることになる. データ前処理の方法が分析結果に影響を与えることから, より適正な方法について今後検討の余地がある.

える。なお、分析データのタイムスタンプがミリ秒単位であることや、データのサンプル頻度 (1 秒当たりの件数) の制約から、ミリ秒のオーダーやそれ以下での高速・高頻度売買を行う HFT の行動に起因する先行運行関係を定量化は困難である。しかし、そのような“超”高速成分を除去した部分での、短時間の市場間の先行運行関係は捉えられることが期待される。

4 実証分析結果

4.1 先行運行時間の散布図

Core30 銘柄の各々に対して、データ期間内の各日毎、各時間帯毎に $\hat{\theta}$ を計算し、それらのデータ期間全体分布の中央値を計算した。図 2 は、横軸に $\hat{\theta}$ の 2013 年の期間内の中央値、縦軸に 2014 年の中央値を取り、30 銘柄について散布図を作成したものである (上段から順に第 8-10 時間帯)。各軸の目盛単位は秒である。その他の時間帯における散布図は省略する。^{*7}

市場ペア X-Y に対応する各図中 (左列, 中央列, 右列) において、正の領域 (水平軸 (2013 年) は右側, 垂直軸 (2014 年) は上側) は、X が Y に先行することを示す。各図中、45 度線より上側は 2013 年から 2014 年にかけて X の Y に対する先行度合が高まったことを、下側はその逆を示す。

図中の左列 (東証-JNX) は、期間内の平均的傾向として、東証が JNX に対して (若干の例外を除き) 先行していること、中央列 (東証-ChiX) は東証が ChiX に対して先行している銘柄が多数を占めること、そして、いずれの市場ペアにおいても、5 ミリ秒程度を中心に分布していることを示している。一方、右列 (JNX-ChiX) は両市場がほぼ先行運行のない関係 (時間差ゼロ) であることを示している。

さらに、2013 年から 2014 年への変化として、銘柄によるばらつきはあるものの、全般に、東証-JNX においては幾つかの銘柄において東証が、東証-ChiX, JNX-ChiX においては ChiX が、相当数の銘柄において先行度を高めた様子が観察される。これらの $\hat{\theta}$ に関する 3 つの市場ペアの先行運行関係は、先に行われた LLR による分析結果とは異なったものになった (林 (2015a))。ここで、市場ペアによって、軸のスケールが異なることに注意されたい。特に、JNX-ChiX (右列) は 1 目盛がデータセットの時間解像度 (1 ミリ秒) と一致しており、各銘柄は -1 ミリ秒から +2 ミリ秒の格子点の上に分布している。

同図より、東証-JNX, 東証-ChiX の 2 ペアと、JNX-ChiX の散布図の様子の違いは明らかである。市場間を結ぶネットワークのレイテンシー (遅延) に起因するののか、市場参加者の構成割合や取引戦略の相違に起因するののか現時点では不明であるが、何らかの構造的な違いを表していることが察せられる。参考までに、比較対象銘柄群 JAFEE30 に関する散布図を図 3 に示す。先の Core30 銘柄と概ね類似したものとなった。JNX-ChiX 間において殆ど全ての銘柄に先行運行時間の差がない (観測限界=1 ミリ秒以下) という結果は説得的であり、本手法の有用性を示唆するものであると筆者は考える。

なお、本論文では省略するが $\hat{\theta}$ と $\ln(LLR)$ (LLR の対数值) を縦軸横軸に取り散布図を作成すると、互いに異符号を持つケースが散見された。これは、一方の市場での価格形成が他方よりも常に先行していると言った単純な図式ではなく、2 つの市場のそれぞれにおいて他方に先行するような変動要因を持っていることを示唆しているのかもしれない。このような実証的現象は、加藤他 (2011) による“マルチラグモデル”によって再現できることがモンテカルロ・シミュレーション実験により確認されているが、より詳細な理論研究については今後の課題である。

^{*7} 時間帯による変動は見られたものの、概ね同様な傾向を示した。

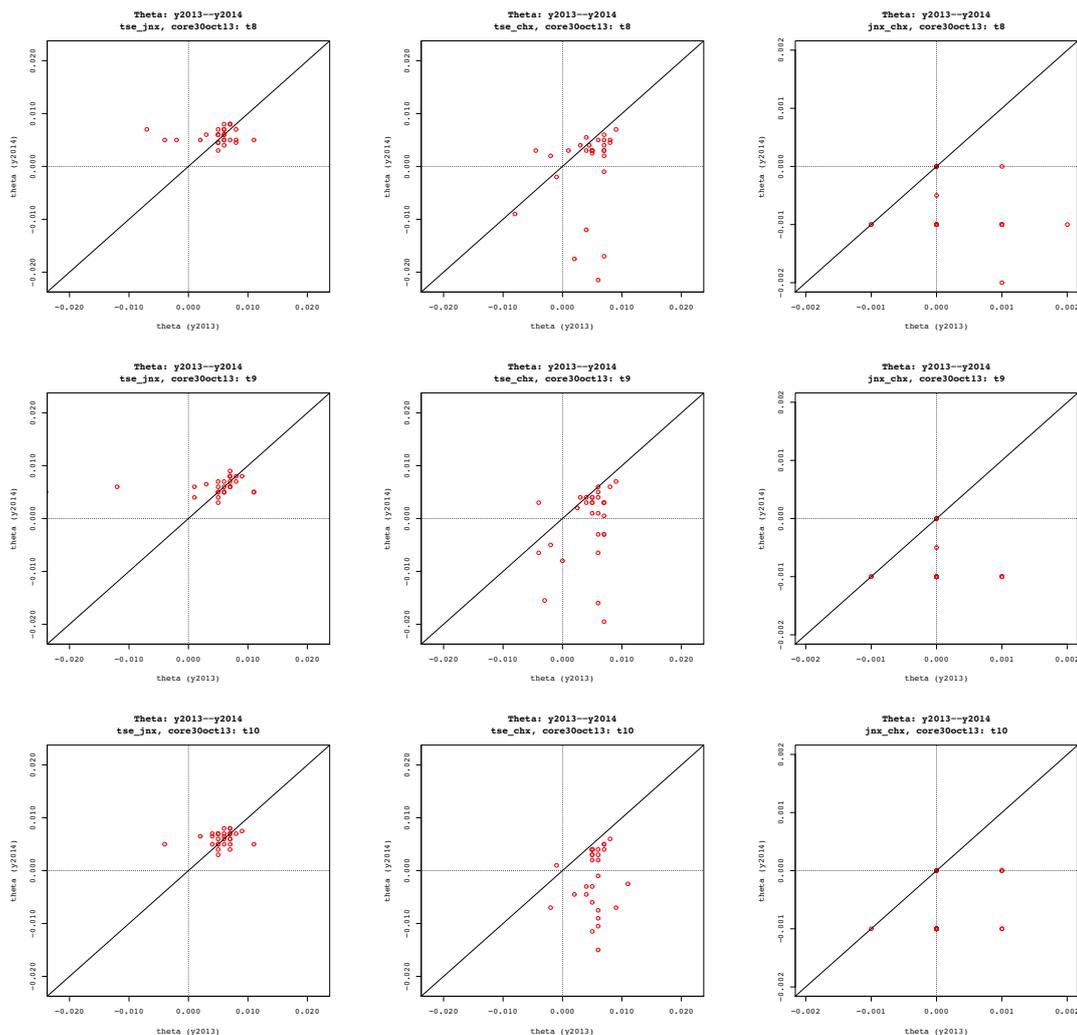


Figure2 $\hat{\theta}$ の 2013 年データ期間内中央値 (横軸), 同 2014 年データ期間内中央値 (縦軸) の散布図 (Core30 銘柄): 東証 (TSE)-JNX (左列), 東証 (TSE)-ChiX (中央列), JNX-ChiX (右列). 第 8 時間帯 (上段), 第 9 時間帯 (中段), 第 10 時間帯 (下段). 各軸の目盛単位は秒.

4.2 パネルデータ分析

前節ではデータ期間内の中央値を使って, 市場間の平均的な先行遅行関係について観察したが, 次に, 個々の $\hat{\theta}$ 計測値をそのまま被説明変数に用い, 市場間の先行遅行指標の銘柄間の特徴の類似性や相違性をパネルデータ分析によって定量的に評価する. 先述のように $\hat{\theta}$ はデータ解像度ならびにサンプル頻度の観点からは, 現状は“ノイズ”のレベルが高く個々の計測値の信頼性には課題があるものの, ここでは, $\hat{\theta}$ を説明変数に対して標準的な回帰分析の手法を適用することによって $\hat{\theta}$ の変動についての特徴把握を試みる. その際, 銘柄固有の要因や日次の要因による変動は除去して考える.

先に, データ期間内中央値に集約した先行遅行関係において 1 日内変化が見られたことから, ここでは, やはり東証の立会時間全 5 時間を 30 分毎に 10 個の時間帯に分割したデータを使って計算された 1 日 10 個の先行遅行時間を多変量時系列 (パネルデータ) に配置した. 銘柄のカバレッジとして, 興味の対象である Core30 銘柄のほか

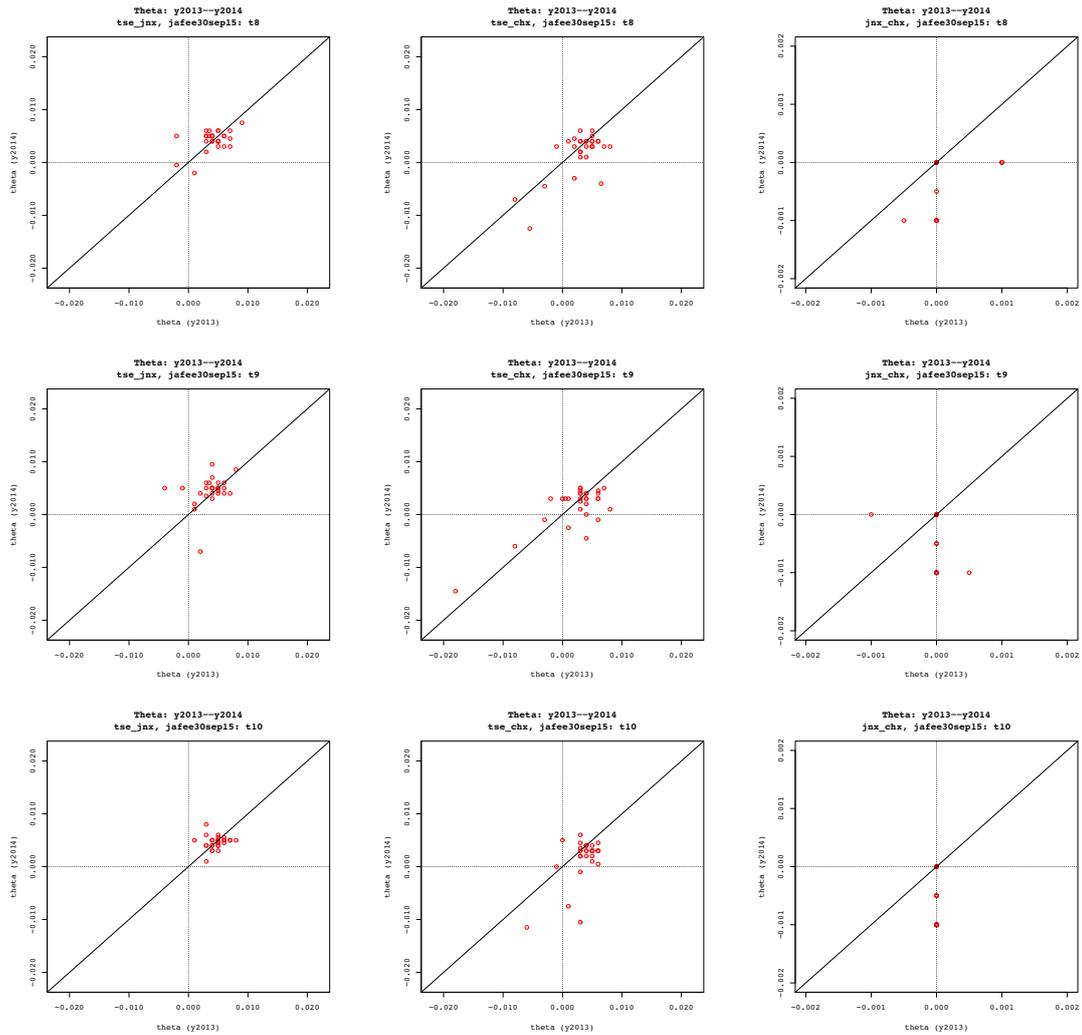


Figure3 $\hat{\theta}$ の 2013 年データ期間内中央値 (横軸), 同 2014 年データ期間内中央値 (縦軸) の散布図 (JAFEE30 銘柄): 東証 (TSE)-JNX (左列), 東証 (TSE)-ChiX (中央列), JNX-ChiX (右列). 第 8 時間帯 (上段), 第 9 時間帯 (中段), 第 10 時間帯 (下段). 各軸の目盛単位は秒.

に, 比較対照銘柄である JAFEE30 銘柄を加えた計 60 銘柄から成るデータセットを作成した. 前者は全て 2014 年の東証ティックサイズ変更の対象であるが, 後者は対象ではない. 一つのデータセットとして同時に分析することで両群の差異について調べる.

パネルデータ分析のためのアプローチとして, 銘柄固有の要因や日次要因に起因する $\hat{\theta}$ 間の相関の影響を考慮するために, 今回は線形混合効果モデル (linear mixed effects model) を用いることにした.*8

本稿では, 被説明変数として, 第 i 銘柄 ($i = 1, \dots, 60$), 第 j 日 ($j = 1, \dots, 245$), 第 t 時間帯 ($t = 1, \dots, 10$) における $\hat{\theta}$ の値を使用し, $y_{ijt} = \hat{\theta}_{ijt}$ と書く. $y_{ijt} > 0$ は, 市場 X が市場 Y に先行することを, $y_{ijt} < 0$ は逆を示す.

説明変数については, ボラティリティや約定件数など個別銘柄の市場特徴量を, データ期間内の各日において, ペアを組む 2 つの市場について計算し固定効果として与えた. また, 2014 年の東証ティックサイズ変更の効果を捉えるための 2 水準固定因子, 1 日内時間変動を捉えるための 10 水準因子を加えた. さらに, 日次効果を表す変数と, 銘

*8 R 関数のパッケージ “lme4” 内にある関数 lmer() を使用した. ロバスト回帰等, 推定方法の改善については今後の課題である.

銘柄の違いを表現するため銘柄固有要因を表す変数の2変数を変量効果として導入することにした。さらに、銘柄がCore30に(すなわち東証ティックサイズ変更対象銘柄群に)属するか否か(JAFEE30に属するか)を示す2水準固定因子を加えた。

分析目的上、二つの市場間の相対的な特徴の差異を見ることが必要なことから、今回の分析では両市場特徴量の対数比を取ることで新たな変数を定義した。例えば、第*i*銘柄、第*j*日の第*t*時間帯におけるX,Y二つの市場における実現ボラティリティが RV_{ijt}^X, RV_{ijt}^Y であれば、 $RV_{ijt} = \ln(RV_{ijt}^X/RV_{ijt}^Y)$ と言った具合である。すなわち、 $RV_{ijt} > 0$ (< 0)であれば、市場Xで計測された実現ボラティリティが、市場Yよりも大きかった(小さかった)ことを示す。また、被説明変数への影響の大きさを説明変数間で比較できるように、全ての共変量は標準偏差が1となるようにスケール変換を施した。よって、回帰係数の値は、 RV_{ijt} が1標準偏差分増加した時の、先行運行時間 y_{ijt} の増加時間(単位は秒)と解釈される。一方、 y -切片の値の解釈は、全ての共変量の値が同時にゼロとなる(市場1と市場2の観測値が等しい)ような仮定の銘柄に対するベースライン効果(第1時間帯、非Core30銘柄、東証ティックサイズ変更前に同時に属する)を秒単位で示したものであり、さらに、各固定因子の効果はベースラインからの変化量として捉えることができる。⁹

複数の候補変数群の中から、AICや回帰係数の有意性を確認しながら、また解釈容易性も勘案しつつ、モデル選択、変数選択を行った。その結果、次のモデルを選択し、パラメータの推定を行った。

$$y_{ijt} \sim 1 + \underbrace{Code_i + Ymd_j}_{\text{random effects}} + \underbrace{Tck_j}_{\text{2-level factor}} \times \left\{ \underbrace{Idx_i}_{\text{2-level factor}} + \underbrace{T_t}_{\text{10-level factor}} + \underbrace{RV_{ijt} + nQ_{ijt} + Vol_{ijt} + Spr_{ijt} + Dep_{ijt}}_{\text{covariates}} \right\} \quad (5)$$

ただし、1は*y*-切片、 $RV_{ijt}, \dots, Dep_{ijt}$ は、固定効果(fixed effects)を、 $Code_i$ は個別銘柄効果、 Ymd_j は日次効果であり、共に変量効果(random effects)を表す。ここで、 RV_{ijt} は仲値より計算される実現ボラティリティの対数比、 nQ_{ijt} は気配更新数の対数比、 Vol_{ijt} は約定枚数の対数比、 Spr_{ijt} はスプレッド率(ビッド・アスク・スプレッド÷仲値)の対数比、 Dep_{ijt} は5本気配値までの売買気配数量合計の対数比である。¹⁰ Idx_i は、Core30に属するかJAFEE30に属するかを表す2水準固定因子(水準0は後者)、 Tck_j は、東証ティックサイズ変更の前後を表す2水準固定因子(水準0は変更前)、 T_t は時間帯を表す10水準固定因子(水準0は第1時間帯(9時-9時半))である。なお、これらの説明変数のうち、実現ボラティリティ、気配更新数、約定枚数は属する時間帯の累積値、他方、スプレッド率、および5水準デプスは時間帯内平均値(次の気配更新までの経過時間による加重平均)であり、いずれも、各銘柄毎かつ各時間帯毎に算出した。

推定された回帰係数(および95%信頼区間)を図4-7に集約して表示する。まず、図4は推定された*y*-切片および共変量の回帰係数を、図5は1日内効果の推定値をそれぞれの95%信頼区間と同時にプロットした図である。上述のように、被説明変数 $\hat{\theta}_{ijt}$ は、コレログラムの最大値を達成する値であり、ばらつきは大きい。このような中、ここで示す推定係数の有意性についての評価の是非について十分留意しつつ、今回得られた幾つかの特徴的な観察結

⁹ 分母や分子の値にゼロが入ることで対数比が発散したり不定となった説明変数を持つレコードは、パネル回帰から除去した。例えば、みずほFG(8411)は2014年のティックサイズ変更以前は、東証において取引は活発だったものの、最良気配数量が極めて多く(板が厚く)仲値の変化が乏しい日が多く、仲値ベースの実現ボラティリティの値がゼロとなるケースが多発した。なお、林(2015a,b)では、 $RV_{ijt} = \ln(RV_{ijt}^Y/RV_{ijt}^X)$ と定義していたが、解釈の容易性に鑑み、本稿では対数内の分母分子を入れ替えることにした。

¹⁰ 各変数の入力となる市場観測量として、3.3で述べたデータ前加工処理後の計測量を使用した。

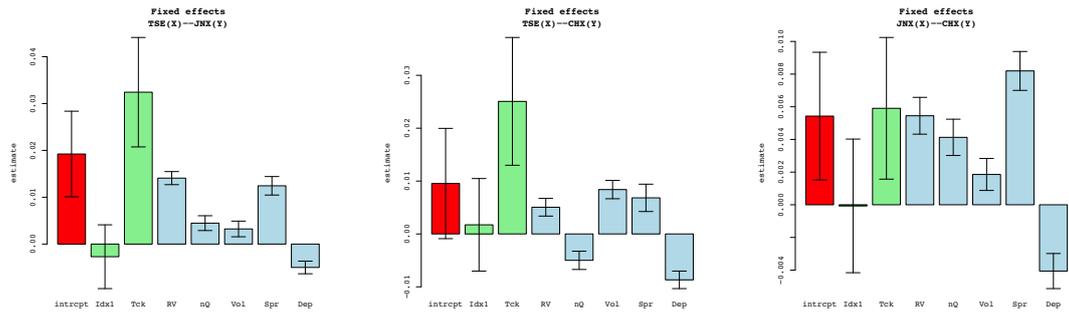


Figure4 回帰係数推定値および95%信頼区間-1: 左から、Y-切片, *Idx* (Core30 銘柄効果), *Tck* (ティックサイズ変更効果), *RV*, *nQ*, *Vol*, *Spr*, *Dep* の各共変量。東証 (TSE)-JNX (左図), 東証 (TSE)-ChiX (中央図), JNX-ChiX (右図)。

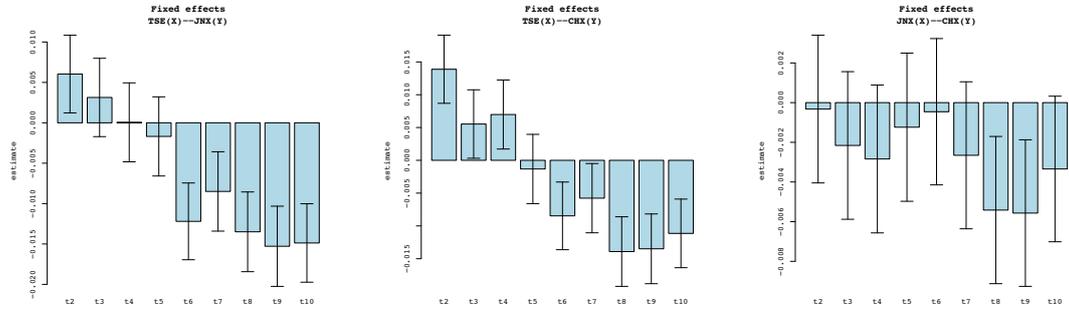


Figure5 回帰係数推定値および95%信頼区間-2: 時間帯効果. 左から、第2時間帯 (*t2*) から第10時間帯 (*t10*) まで。東証 (TSE)-JNX (左図), 東証 (TSE)-ChiX (中央図), JNX-ChiX (右図)。

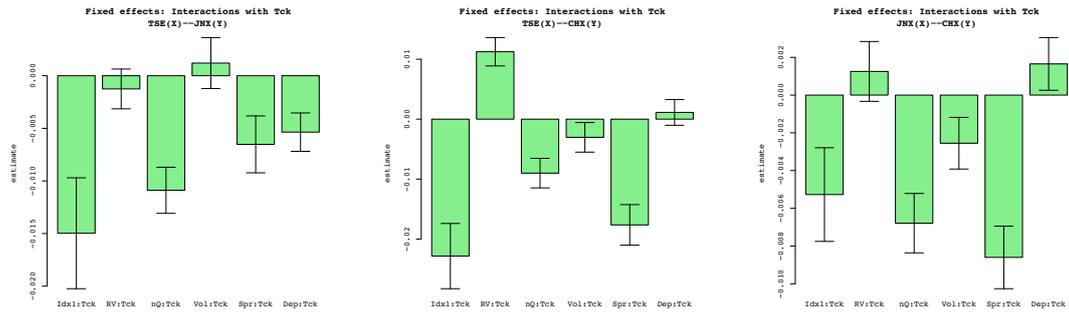


Figure6 回帰係数推定値および95%信頼区間-4: ティックサイズ変更効果と各共変量との交互作用。左から、Core30 銘柄効果 (*Idx*) \times ティックサイズ変更効果 (*Tck*), *RV* \times *Tck*, ..., *Dep* \times *Tck* まで。東証 (TSE)-JNX (左図), 東証 (TSE)-ChiX (中央図), JNX-ChiX (右図)。

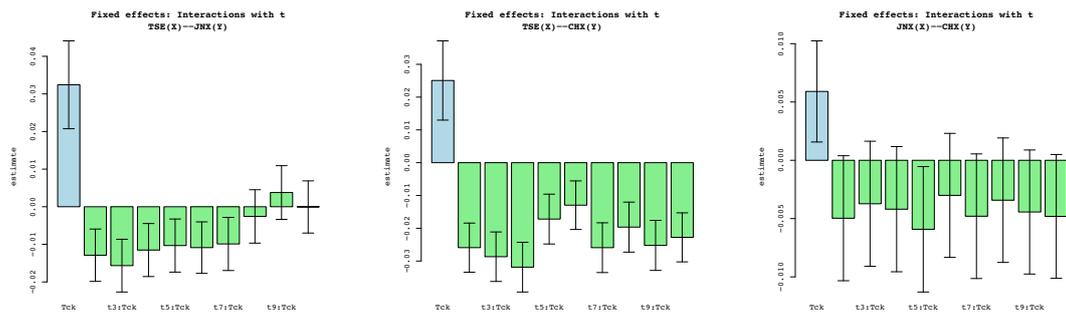


Figure7 回帰係数推定値および95%信頼区間-3: ティックサイズ変更効果と時間帯効果との交互作用。左から、第2時間帯(t_2)×ティックサイズ変更効果(Tck), ..., 第10時間帯(t_{10})× Tck まで。東証(TSE)-JNX(左図), 東証(TSE)-ChiX(中央図), JNX-ChiX(右図)。

果について言及したい。

図4によれば、市場ペアの3つの組み合わせにおいて、共変量の回帰係数の符号に共通のパターンが観測される。すなわち、実現ボラティリティ比 RV_{ijt} , およびスプレッド率比 Spr_{ijt} は有意(両側5%)に正の係数を持つ。これは林(2015a)で紹介したLLRによる分析結果と同じである。さらに、今回の分析では、約定枚数比 Vol_{ijt} は正に有意、デプス比 Dep_{ijt} が負に有意となった。

最近のHFTの市場クオリティへの影響に関する実証研究によれば、HFT(特に“メイカー戦略”を取る業者)は流動性供給を通じて市場ボラティリティを下げる効果があるとされる(Brogaard et al. (2014); Hasbrouck and Saar (2013))。一方、今回得られた結果は、一つの市場の実現ボラティリティがもう一つの市場に対して相対的に大きくなるのに連動して、その市場の相対的な先行度合が増すことを示している。筆者の理解では、HFTの功罪の文脈で語られる“ボラティリティ”は市場価格の上下動、すなわち不確実性の大きさであり一般には回避すべき対象であると考えられる。他方で、ここで用いているのは仲値系列より計算された実現ボラティリティであり、最良気配値の一方が動くたびにその変化が反映される量である。ボラティリティは元来当該企業のファンダメンタル価値の変動を示す値であり3市場間で同一値を取るべきものとの視点もある一方で、仲値系列ベースの実現ボラティリティは、最良気配枚数(待ち行列の長さ)や注文到着頻度、ひいてはティックサイズの大きさ等に依存する観測量であり、要すれば各市場における“マイクロストラクチャ”および当該銘柄の人気度・取引活況度に依存する量である。^{*11} すなわち、定義式は同一ながら標準的な実現ボラティリティとは異なる使い方をしていると言えよう。しかも、ここでは説明変数としてボラティリティそのものではなく2市場のボラティリティ比を使っていることから、仮に分析に使用したデータの背後にHFTの取引量の増大があるとしても、今回の結果は上記HFTの功罪に関する実証研究におけるコンセンサスと必ずしも矛盾するものではないと考える。

スプレッド率比 Spr_{ijt} は、流動性供給側による指値注文キャンセルにしろ、流動性需要側による成行注文・即時約定にしろ、最良気配の一方が消滅することにより高まる量であることから、相場が急変する時や、需給が一方に一時的に偏る時に高まるであろう。相場動向により迅速に注文行動を行える市場参加者の割合の大きい市場の方がより高い先行度合を持つとすれば、自然な結果ではある。その一方で、一般に流動性の高い市場はスプレッド率が低いことが期待される。本分析で扱っている市場間の先行遅行度と相対的流動性の関係は明白でないものの、 Spr_{ijt} の符号は負であるべきとの見方も可能であろう。今後さらなる検証や考察が必要である。

同様に、約定枚数比 Vol_{ijt} が正に有意との結果も、テイカー側の能動的な売買行動の帰結と考えれば自然である。

^{*11} このような視点では、“マイクロストラクチャ・ノイズ”の除去は不要どころか、活用すべき情報源と言える。

さらに、デプス比 Dep_{ijt} の結果は、一方の市場の内側 5 本値の注文枚数合計 (デプス) がもう一方の市場に対して相対的に増すと、その市場の先行度が相対的に低下する傾向を示している。5 本値を考慮することで、デプスの厚さは、市場参加者の多さはもちろん、彼らの売買スタイルの多様性を表しているとも考えれば、推定係数が負であるとの結果はデプスが相対的に大きくなるほどその中には高速での取引を行わない参加者からの注文がより多く含まれやすい等の傾向を反映したものかもしれない。

図 5 に示す時間帯効果は、東証-JNX, 東証-ChiX とともに、第 1 時間帯 (ベースライン) と比べて、午前には東証が先行し、午後は PTS が先行する様子が観察される。特に、これらの市場ペアでは、最後の 3 つの時間帯において、第 1 時間帯に比して 10 ミリ秒以上 PTS の東証に対する先行度が増加していることが分かる。PTS を利用しているのは、HFT はもとより、機関投資家や証券会社の自己売買部門等のプロの参加者であり、彼らが主市場の大引けが近づくにつれて、1 日の売買目標や終了時点での目標ポジションに向け積極姿勢に転じている様子を反映したものであると考えれば合点がいく。

次に、東証ティックサイズ変更の影響を調べる。先の 2 つの図 4-5 に加えて、ティックサイズ変更による各変量の持つ回帰係数の変化 (Tck_j と各説明変数との交互作用) を示す図 6 と、一日内各時間帯別の東証ティックサイズ変更の影響度合 (Tck_j と T_t の交互作用) を示す図 7 を同時に眺める。なお、上述のように、ここで示されている効果の大きさは、全ての共変量において値がゼロである (共変量を構成する対数内の分母分子の市場観測量が 2 つの市場間で等しい) 仮想的銘柄に対してのものであると解釈される。実際、分析の過程では、データ処理の方法により、今回報告するものとは大きさや符号の異なる推定値が得られている。^{*12}

まず、東証-JNX において、図 4 で示される固定因子の主効果 (JAFEE30 銘柄、東証ティックサイズ変更前、第 1 時間帯がベースライン) を見ると、東証ティックサイズ変更により JAFEE30 銘柄では 30 ミリ秒強東証が JNX に対して先行度を高めた (Tck が有意に正) が、これを図 6 の交互作用効果を見比べると、東証の先行度増は Core30 銘柄では 15 ミリ秒程度であることが読み取れる ($Idx1 : Tck$ が有意に負)。次に、東証-ChiX においては、ティックサイズ変更により JAFEE30 銘柄が 24-25 ミリ秒程度東証での先行度を増した (図 4) が、Core30 銘柄では JAFEE30 銘柄より ChiX での先行度が同程度増しており (図 6) ティックサイズ変更の影響がほぼ相殺されていることが分かる。他方、JNX-ChiX では、ティックサイズ変更は JAFEE30 銘柄の JNX の先行度を 6 ミリ程度押し上げたが (図 4)、逆に Core30 銘柄においては 5 ミリ秒程度 JAFEE30 銘柄より ChiX 側の先行度が高くやはりティックサイズ変更の影響がほぼ相殺されているのが分かる (図 6)。ちなみに、東証ティックサイズ変更前においては、Core30 銘柄と JAFEE30 銘柄間に有意な差は見られなかった (図 4)。

さらに、ティックサイズ変更による回帰係数の変化としては、図 6 より、気配件数比 nQ_{ijt} 、スプレッド率比 Spr_{ijt} において、有意な負の変化が全市場ペアに共通なパターンとなった。また、図 7 より、ティックサイズ変更により、東証-JNX においては第 8-10 時間帯を除く全時間帯、東証-ChiX においては全時間帯において、JNX および ChiX の対東証の先行度が第 1 時間帯 (ベースライン) に比して有意に増したことが分かる。^{*13}

全般を通して、PTS 同士である JNX-ChiX ペアにおける推定結果が、他の 2 ペアとはやや乖離したものとなった。

今回の分析では、最良気配値の、最良気配数量の逆数をウェイトに用いた加重平均をマイクロ・プライスとして分析に使用した。一方、林 (2015a,b) でのマイクロ・プライスは、ウェイトとして内側 3 枚の数量合計を使用していた。今回得られた結果は、先に行った分析と全般には類似の傾向を持つが、細部では異なる箇所も見られる。マイクロ・プライスの作り方の違いにより、分析用データに反映される参加者の種類が異なってくる可能性もある。

^{*12} 例えば、本論文の旧バージョン (林 (2015b)) においては、共変量を平均ゼロに中心化して分析を行った。固定因子の効果としてここに掲載するものとは異なる推定結果が得られた。

^{*13} 本論文とは別に、林 (2015a) でも使用した比較対照銘柄群 ("FIX30") を用いて同じモデル (5) で分析を行ったところ、同様な傾向が得られた。なお、林 (2015a) における分析対象銘柄は、2012 年 10 月時点での TOPIX Core30 であった。2013 年 10 月末に 2 銘柄の入替えが行われた (花王 (4452)、東芝 (6502) がはずれ、デンソー (6902)、三井不動産 (8801) が入った)。

今後の課題

今回分析に使用した $\hat{\theta}$ は、推定精度について統計学上の課題であり、方法論の未確立な統計量を使用した今回の結果報告は時期尚早かもしれない。もとより、分析データのタイムスタンプがミリ秒単位であることや、データの発生頻度（データ間隔）制約から、ミリ秒のオーダーやそれ以下での“超”高速・高頻度売買を行う HFT の行動に起因する先行遅延関係を定量化できていないことには留意が必要である。また、生データを分析可能な形式にするために数段階の前処理が施されるが、分析前のデータ前処理段階における“外れ値”に対するフィルターの設定方法が分析結果に影響を与えることが本研究の過程で確認されている。妥当な前処理方法については今後検証が必要である。また、マイクロ・プライスのウェイトの取り方の違いにより、分析用データに反映される市場参加者の種類が異なってくる可能性もある。これについても、今後検証を進めねばならない。

本研究の過程で、2つの先行遅延指標 ($\hat{\theta}$, $\ln(LLR)$) の銘柄別時間帯別の期間内中央値を同時に計算したところ、異符号を持つケースが多数見られた。このことは2つの時系列が単一の固定されたラグ・パラメータ θ を持っているのではなく、符号の異なる複数のラグ・パラメータを持っていることを示唆しているとも考えられ、更なる検証が必要である。そのような場合には、一つのラグのみ計測する今回のアプローチは適切ではない。“マルチラグ・モデル”は(加藤他(2011))によって提案されているが、推定方法については課題である。

今回の分析結果を実際の市場参加者の行動と関連付ける作業は残されている。今後既存のファイナンス/マイクロストラクチャ分野の先行研究との関連性についても調査せねばならない。

Michael Lewis 著のベストセラー“Flash Boys”でも描かれているように、地理的にも離れた場所に多数の取引所が乱立し、我が国とは異なる最良執行ルールが適用される中で“レイテンシー・アービトラージ”の余地が存在する米国株式市場と比べると、今回の分析結果は、国内市場において直ちにそのようなアービトラージ(倫理的課題については脇へ置くとしても)を実現するような取引手法に応用できる可能性が低いものと考えられるが、いずれにせよ実務への応用も研究上の課題である。

5 結言

本研究は、東京証券取引所(主市場)と2つの私設証券取引所-チャイェックス(ChiX)、ジャパンネクストPTS(Japannext PTS)-の3市場にて同時に取引されている国内株式について、注文板形成の先行遅延関係の存在やその大きさを実証分析した。

まず、各市場の注文板データを使ってマイクロ・プライス系列を生成し、Hoffmann et al. (2013)による先行遅延関係推定法を用いて、銘柄別時間帯別に先行遅延時間を計測した。個別の計測値より、データ期間内中央値に要約し散布図を観察した。次に、個々に計測された先行遅延時間を多変量時系列データ(銘柄×データ期間)として構成し、パネル回帰分析を行い、個別銘柄要因や日次要因を除去しながら、銘柄に共通な特徴や相違点を抽出し先行遅延要因を探った。分析対象銘柄は、東京証券取引所において時価総額および流動性の特に高い30銘柄から成るTOPIX Core30構成銘柄と比較対照銘柄(JAFEE30)の計60銘柄、分析期間は2013年1月4日~2014年12月30日の489日間、使用データの時間解像度はミリ秒単位である。

主要な実証的発見は次の通りである。まず、データ期間内中央値で観察する限り、銘柄や時間帯による違いはあるものの、3市場間の先行遅延関係として、一般的に東証は他二市場に5ミリ程度前後の大きさで先行する傾向が見られた。東証とJapannextの関係は、林(2015a)におけるLLR指標によるものと異なった。2014年の東証におけるティックサイズ変更は、全般にChiXが他二市場に対して先行度合を若干高めた様子が観察された。

次に、パネル回帰分析により、先行遅延関係を説明する変数として、ボラティリティやスプレッド率、約定枚数が相対的に高まる市場はより先行度合が高まる傾向が見られた。他方、気配数量(デプス)が相対的に高まる市場は逆

に先行度合が低下(運行度合が高まる)傾向が見られた。先行運行時間は一日の中で変化することも確認された。特に、東証の立会終了前の最後の90分間は、前場寄付後30分と比べて、10数ミリ秒程度PTSが東証に対して先行度を増すことが確認された。これらの分析結果は、日々の相場変動によらず一日内の時間帯に応じて売買行動をシステムティックに変化させるような、高速の取引戦略の存在を示唆しているものと思われる。

さらに、東証ティックサイズの影響による先行運行度の変化は、Core30銘柄と非Core30銘柄間で違いが見られた。市場ペアに共通する回帰係数の変化も観察された。

分析の限界および課題として、分析データのタイムスタンプがミリ秒単位であることや、データのサンプル頻度の制約から、ミリ秒未満での高速・高頻度売買を行う“超”高速のHFTの行動に起因する先行運行関係は捉えられてはいない。データ前処理方法に対する今回の分析結果の頑強性も課題点として残されており、適切なデータ前処理方法についても検証・検討を重ねる必要がある。さらに、採用した先行運行推定に関する統計的方法論は提唱されて日が浅く理論的に未整備な状況にある。“ノイズ”のレベルが高くしかも推定精度に関する理論の確立していない指標 $\hat{\theta}$ を用いた今回の分析結果は報告としてはあくまで暫定的なものに過ぎず、結果の解釈には十分注意されなければならない。今後検証が必要である。本報告の知見を深化・発展させるためには、分析手法の改良や代替的手法の開発が必要である。

今回得られた結果は特定の期間、特定の銘柄群を分析して得られた暫定的なものに過ぎないことから、これを直ちに一般化することはできない。他の期間や銘柄データを用いた実証分析が行われねばならない。得られた結果をマイクロストラクチャ研究と関連付ける作業も推し進めねばならない。

謝辞

本研究遂行にあたり、東京大学吉田朋広教授、大阪大学太田亘教授、統計数理研究所小池祐太氏、パリ第6大学Mathieu Rosenbaum教授、株式会社東京証券取引所近藤真史氏、同保坂豪氏、チャイエックス・ジャパン株式会社永堀真氏より有益なコメントを頂いた。日本ファイナンス学会第21回大会(2013年6月2日、於武蔵大学)、同25回大会(2015年6月7日、於東京大学)、2013年度統計関連学会連合大会(2013年9月9日、於大阪大学)、研究集会“金融工学からERMへ”(2014年3月6日、於一橋大学)、研究集会“Asymptotic Statistics and Computations 2014”(2014年3月11日、於統計数理研究所)、“The 5th Annual on High-Frequency Data in Finance Conference”(2013年10月26日、Stevens工科大学)、“Market Microstructure and High-frequency Data”(2014年5月31日、於シカゴ大学)、“Non- and Semiparametric Volatility and Correlation Models—Economic Sources of Volatility, Risk Decomposition and Financial Crises”(2014年7月25日、於パーダーボルン大学)、“Statistics for Stochastic Processes and Analysis of High Frequency Data IV”(2015年3月23日、於パリ第6大学)、2015年度統計関連学会連合大会(2015年9月7日、於岡山大学)、人工知能学会第15回金融情報学研究会(2015年9月26日、於東京大学)における口頭発表において、参加者より有意義なコメントを頂いた。株式会社日本取引所グループ、チャイエックス・ジャパン株式会社、SBI ジャパンネクスト証券株式会社よりデータをご提供頂いた。本研究は、JST(科学技術振興機構)CRESTプロジェクト『先端的確率統計学が開く大規模従属性モデリング』(研究代表者吉田朋広教授)、全国銀行学術研究振興財団研究資金(2013年9月–2014年8月)、および慶應義塾大学学事振興資金(2012–2014年度)の研究資金援助により行われた。ここに謝意を表します。当然ながら、論文中に含まれる誤りは全て筆者に帰属します。

References

Avellaneda, Marco, Josh Reed, and Sasha Stoikov (2011) “Forecasting prices from level-I quotes in the presence of hidden liquidity,” *Algorithmic Finance*, Vol. 1, No. 1, pp. 35–43.

- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014) "High-Frequency Trading and Price Discovery," *Review of Financial Studies*, Vol. 27, No. 8, pp. 2267–2306.
- Gatheral, Jim and Roel C.A. Oomen (2010) "Zero-intelligence realized variance estimation," *Finance Stoch.*, Vol. 14, pp. 249–283.
- Hasbrouck, Joel and Gideon Saar (2013) "Low-latency trading," *Journal of Financial Markets*, Vol. 16, No. 4, pp. 646–679.
- 林高樹 (2015a) 「高頻度注文板データによる 2014 年東証ティックサイズ変更の国内株式市場への影響分析」, 『証券アナリストジャーナル』, 第 53 巻, 第 4 号, 29–39 頁.
- (2015b) 「国内高速 3 株式市場間の注文板形成の先行遅延関係」. 日本ファイナンス学会第 23 回大会予稿集 CD-ROM.
- Hayashi, Takaki and Nakahiro Yoshida (2005) "On Covariance Estimation of Non-synchronously Observed Diffusion Processes," *Bernoulli*, Vol. 11, No. 2, pp. 359–379.
- (2008) "Asymptotic Normality of a Covariance Estimator for Nonsynchronously Observed Diffusion Processes," *Ann. Inst. Statist. Math.*, Vol. 60, No. 2, pp. 357–396.
- (2011) "Nonsynchronous covariation process and limit theorems," *Stoch. Process. Appl.*, Vol. 21, No. 10, pp. 2416–2454.
- Hoffmann, Marc, M. Rosenbaum, and Nakahiro Yoshida (2013) "Estimation of the lead-lag parameter from non-synchronous data," *Bernoulli*, Vol. 19, No. 2, pp. 426–461.
- 保坂豪 (2014) 「東京証券取引所における High-Frequency Trading の分析」, 『証券アナリストジャーナル』, 第 52 巻, 第 6 号, 73–82 頁.
- Huth, Nicolas and Frédéric Abergel (2014) "High Frequency Lead/lag Relationships: Empirical facts," *Journal of Empirical Finance*, Vol. 26, pp. 41–58. arXiv preprint arXiv:1111.7103.v1.
- 加藤宏典・佐藤整尚・吉田朋広 (2011) 「Lead-lag 推定量を用いた為替データの分析」. 2011 年度統計関連学会連合大会予稿集.
- 近藤真史 (2015) 「東証立会市場における呼値の単位の変更の影響」. JPX ワーキング・ペーパー, Vol.7.
- 太田亘 (2013) 「取引システム高速化の流動性に対する長期的影響」. 日本ファイナンス学会第 21 回大会予稿集 CD-ROM.
- 大埴剛士 (2014) 「米国市場の複雑性と HFT を巡る議論」. JPX ワーキング・ペーパー特別レポート.
- 宇野淳 (2012) 「株式取引の市場間競争-上場株取引の市場分散と価格形成-」, 『証券アナリストジャーナル』, 第 50 巻, 第 9 号, 6–16 頁.