

# 高頻度注文板データの統計解析： 異市場・同一株式価格間の先行遅行関係

林 高樹<sup>1,2,3</sup>

(受付 2016 年 10 月 9 日；改訂 2017 年 3 月 15 日；採択 3 月 16 日)

## 要 旨

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大きさやその特徴を実証的に調査することを目的とする。

先行遅行時間の推定方法として、本稿は Dobrev and Schaumburg (2015) が最近提案した方法を採用する。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高い 100 銘柄から成る TOPIX100 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。まず、個別銘柄ごとに異取引市場間での市場価格の先行遅行時間を推定し、次に、これらの推定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。

キーワード：高頻度データ、高頻度トレード、先行遅行分析、Hoffman-Rosenbaum-Yoshida 推定量、Dobrev-Schaumburg 推定量、マーケット・マイクロストラクチャ。

## 1. 導入

世界の株式市場において近年取引所間の提携や合併が進み、また電子市場が相次いで設立されるなど市場間競争が激化する中、我が国では、現在、東京証券取引所(東証)、2 大私設取引所(PTS)—チャイエックス(運営会社チャイエックス・ジャパン)、ジャパンネクスト PTS (運営会社 SBI ジャパンネクスト証券)の 3 市場において、高速・高頻度の取引が行われている。

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高い「TOPIX100」構成銘柄である。

---

<sup>1</sup> 慶應義塾大学大学院 経営管理研究科：〒 223-8526 神奈川県横浜市港北区日吉 4-1-1 協生館；  
takaki@kbs.keio.ac.jp

<sup>2</sup> 首都大学東京大学院 社会科学部研究科経営学専攻

<sup>3</sup> 独立行政法人科学技術振興機構 戦略的創造事業(JST-CREST)

本研究では、特に 3 市場間の先行運行関係(異市場・同一銘柄間分析)に興味がある。ベースとなる統計的方法論は、Dobrev and Schaumburg (2015)による先行運行関係推定法である。実証分析の流れは林 (2016)で行ったものを概ね踏襲する。すなわち、まずデータ期間内一日毎・時間帯内毎、銘柄毎に、先行運行時間を推定し、推定値の日次推移や時間推移を図示し、要約統計量等を計算することで、市場ペア間の先行運行時間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。その際に採用する説明変数には、例えば、“売り方向”、“買い方向”の違いによる先行運行時間の相違など、林 (2016)とは異なる視点も導入する。

本研究は、林 (2016)と同様、データの解像度やデータの発生頻度(データ間隔)の制約により、ミリ秒以下の先行運行関係を定量化することはできない。また、各注文データには市場参加者を特定する情報が付与されていないため、近年市場での存在感を増している高頻度トレード(High-frequency Trading, HFT)の行動を直接分析することはできない。このようなデータ仕様に起因する制約はあるものの、分析を通じて HFT の注文行動を理解する一助となることが期待される。

筆者による 2 つの実証研究(林, 2015, 2016)からの分析の方法論上の主要な改良点としては、(i) Hoffmann et al. (2013)の代替的方法論として Dobrev and Schaumburg (2015)によるアプローチを採用して同一銘柄異市場間の先行運行時間を推定したこと、(ii)説明変数の見直しを行うことで先行運行時間の生成要因に関する回帰分析の精度の向上を図ったことが挙げられる。価格変動幅の情報を使わない後者のアプローチは価格変動に含まれるジャンプやマイクロストラクチャ・ノイズの直接の影響を受けないことから、よりロバストな推定となることが期待される。特に、林 (2015, 2016)で行った前者による推定結果はバラツキが非常に大きく、精度の向上が課題であった。

Dobrev らのアプローチは本稿で示すようにはるかに安定的な結果をもたらす。それでもなお、統計理論が未整備な状況の中、本稿の分析結果はあくまで暫定的なものであることに注意が必要である。<sup>1)</sup>

高頻度・高速市場での先行運行分析自体が近年始まった新しい研究領域であり、実証分析自体が殆ど行われていない。本分析では林 (2015, 2016)と同様、HFT を中心とする高頻度・高速で取引を行う市場参加者の行動やその帰結としての市場クオリティ等に関する特定の仮説を事前に持たず、探索的データ解析を行い、実証的知見を得ることが目的である。

## 2. 国内高速 3 市場の概要およびデータセット

### 2.1 3 市場と東証ティックサイズ変更

3 市場の概要については、宇野 (2012, 2014)、林 (2015)を参照せよ。2008 年 10 月のジャパンネクスト PTS (以下、“JNX”), 2010 年 7 月のチャイエックス(以下、“ChiX”)の開設以降、PTS 二市場の売買高シェアは次第に上昇し、本分析のデータ期間開始の 2013 年前半において 5-7% 程度を占めていた(大崎, 2014, 図表 2)。<sup>2)</sup>

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枝の広がりによる板での順番待ちの緩和”を図った。<sup>3)</sup> 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

表 1. 国内 3 市場のティックサイズ.

東証 (TOPIX100 構成銘柄)		ChiX	JNX
変更前	フェーズ 2 導入後		
≤3,000 円, 1 円	≤1,000 円, 0.1 円	≤5,000 円, 0.1 円	東証 (変更前) の 1/10
≤5,000 円, 5 円	≤5,000 円, 0.5 円	≤100,000 円, 1 円	
≤30,000 円, 10 円	≤10,000 円, 1 円	>100,000 円, 10 円	
...	...	...	

た. 表 1 に変更前後のティックサイズ的具体例を, PTS 二市場の値と共に示す.

東証ティックサイズ変更の影響については, 林 (2015) 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.

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, その他, Bellia et al. (2016), 太田 (2013, 2016), 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

3 市場から入手した高頻度注文板データを使用する. 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.

分析対象銘柄は, 東証の公表する「規模別株価指数・TOPIX ニューインデックスシリーズ」の一つ, “TOPIX 算出対象のうち, 時価総額および流動性の特に高い 30 銘柄”である「TOPIX Core30」と “TOPIX Core30 について時価総額, 流動性の高い 70 銘柄”である「TOPIX Large70」を合わせた「TOPIX100」の構成銘柄である (2013 年 10 月末選定分).<sup>4)</sup> これらは, 2014 年の東証ティックサイズ変更の対象銘柄群である. 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. なお, 分析対象銘柄はデータ期間の 2 年間固定し, 期間内に 2 回実施された「TOPIX100」の銘柄入替は反映させない.<sup>5)</sup>

PTS 二市場には昼休みによる中断時間がないが, PTS と東証間の同一銘柄価格の先行運行関係をみるため, 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 タイムスタンプの生成

本分析で採用する Dobrev and Schaumburg (2015) によるオリジナルのアプローチは, 2 つ価格系列のタイムスタンプ情報のみを使用する. 彼らは約定価格系列のタイムスタンプを使用した, 本分析では注文板上の価格形成に興味があることから, マイクロ・プライス系列のタイムスタンプを使用する.

まず, 筆者の行った実証分析 (林, 2016) にならう, 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).

$$(3.1) \quad X := \frac{B \cdot Q^A + A \cdot Q^B}{Q^A + Q^B},$$

但し,  $B$  ( $A$ ) は最良買 (売) 気配値,  $Q^B$  ( $Q^A$ ) は最良買 (売) 注文に関する注文枚数合計 (“デブス”) である.

ここでは、林 (2016) と同様に、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、約定はもちろん、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、本分析においても使用が妥当と考える。本稿全体を通して、マイクロ・プライス  $X$  の動きを“値動き”と呼ぶことにする。

各市場、各銘柄毎に、マイクロ・プライスが更新される時刻を集めてタイムスタンプ・データを作る。Dobrev and Schaumburg (2015) のオリジナルの提案では売買方向の区別のないタイムスタンプ系列が使われるが、本研究では、それに少し手を加え、売買の方向性によって二分割したタイムスタンプ系列を使用する(詳細は後述)。

タイムスタンプ・データ生成に使われる代替的な価格系列としては、最良買気配/最良売気配や、それらの平均値である仲値があるが、気配値水準の更新遅れ・鮮度劣化(staleness)は良く知られた現象であることから、これらの更新時点そのまま使用するのは適切でない可能性がある。<sup>6)</sup> 一方、約定値も代替的系列として考えられるが、気配更新データに比べデータ量が大幅に減少することや、約定価格の変化しないゼロ・リターンの割合も高いことから、今回は採用しない。なお、Dobrev and Schaumburg (2015) の実証分析においては、流動性の極めて高い E-Mini (S&P500 先物) や米国 10 年債 (T-Note) の現物や先物に関する約定系列のタイムスタンプを使用している。

### 3.2 時系列データ間の先行運行時間

2 変量の時系列データの先行・運行関係を分析するのに標準的な道具は、両者の相互共分散(cross-covariance)/相互相関係数(cross-correlation)を計測することである。Hoffmann et al. (2013) は、Hayashi and Yoshida (2005) によって提案された非同期共分散推定法(以下、“HY 推定法”)を応用して、離散観測される連続セミマルチンゲール過程間の先行運行関係を推定する方法を提案した。<sup>7)</sup> すなわち、2 つの対数価格系列  $X^1, X^2$  が、それぞれ不規則、不等間隔に並ぶ時点  $\{\tau_i^1\}, \{\tau_j^2\}$  において観測された時、予め定めた有限個の探索範囲  $\mathcal{G}$  に関して、対数収益率間の相互共分散の絶対値を最大化するようなラグ  $h$  の値

$$\hat{\theta} := \arg \max_{h \in \mathcal{G}} |CV(h)|$$

を先行運行推定量とした(以下、“HRY 推定量”)。ここで、

$$CV(h) := \sum_{i,j} (X_{\tau_i^1}^1 - X_{\tau_{i-1}^1}^1) (X_{\tau_j^2}^2 - X_{\tau_{j-1}^2}^2) \mathbf{1}_{\{(\tau_{i-1}^1, \tau_i^1] \cap (\tau_{j-1}^2 - h, \tau_j^2 - h] \neq \emptyset\}}$$

は HRY 推定法による相互共分散関数である。彼らは、正則条件の下で、サンプル観測区間幅および探索グリッド間隔がゼロに縮小する時に、 $\hat{\theta}$  が  $X^1, X^2$  間の真の先行運行時間パラメータ  $\theta$  に対する一致推定量であることを示した。

相互共分散・相互相関係数の大きさを評価しそれを先行運行分析に用いる方法論自体は時系列解析の古典的なアプローチでありマクロ経済分析やファイナンス分野などにおいて古くから行われているが、Hoffmann et al. (2013) はそれを近年の高頻度データに対する統計学の進展を踏まえて、連続時間セミマルチンゲールの離散高頻度観測の枠組の中での先行運行時間の統計的推測問題として定式化し、それに対する一致推定量を初めて提案したという点で、統計学における新たな研究テーマを開拓したと言えるだろう。

筆者は、林 (2015, 2016) において、HRY の枠組みを利用して、本稿と同様国内 3 市場のデータセットを用いて、マイクロ・プライス系列間のクロス相関を介した 3 市場間の先行運行時間の計測を試みてきた。しかしながら、得られた個別の HRY 推定値は、ばらつきが大きく、必ずしも信頼できる分析結果が得られたとは言い難かった。特に、前論文では価格系列を直接使

用した目的関数(相互相関の絶対値)を最大化するアプローチであったため、価格系列に含まれる種々の要因、とりわけ、ジャンプやマイクロストラクチャ効果が影響を及ぼしていた可能性が高い。

そこで、HRY 推定量の代替の方法として、以下で示すように、より安定的な推定をもたらすことが期待される Dobrev and Schaumburg (2015)の方法を採用することにする。

### 3.3 Dobrev and Schaumburg (2015)の方法とその改良

本稿では、マイクロ・プライスに変化をもたらすようなイベントを“注文”や“取引”等と呼ぶことにする。したがって、最良気配に関する新規注文やキャンセル、約定は全て含まれる。なお、Dobrev and Schaumburg (2015)においては、約定イベントのみを分析の対象としている。

以下に、必要な記号を導入する。対数価格(マイクロ・プライスの対数值)系列  $X^1, X^2$  のサンプル時点  $\{\tau_i^1\}, \{\tau_j^2\}$  に対応する計数過程(点過程)  $\mathcal{N}^1, \mathcal{N}^2$  を考える。一方、最小時間解像度  $\Delta > 0$  を単位とする等間隔離散グリッドを  $\mathcal{T} := \{t_i = i\Delta, i = 0, \dots, N\}$  とする。ここで、 $N$  は市場閉場時間に対応する最大グリッド点(定数)である(開場時間  $[0, T]$  として、 $N = \lceil T/\Delta \rceil$ )。次に  $\mathcal{T}$  の中から、各計数過程が変化した時点(注文が直前の  $\Delta$  間に到来した)のみを取り出したサンプル時点を  $\mathcal{T}^k := \{t_i \in \mathcal{T}; \Delta_i \mathcal{N}^k > 0\}$  ( $k = 1, 2$ )と書く。ここで、連続時間確率過程  $Z$  に対して、 $\Delta_i Z = Z_{t_i} - Z_{t_{i-1}}$  と書く。また、 $\mathcal{T}^k$  の全要素を一律に  $\theta \in \mathbb{R}$  だけ負方向へシフトした集合を  $\mathcal{T}_{-\theta}^k := \{t_i - \theta; t_i \in \mathcal{T}^k\}$  ( $k = 1, 2$ )と書く。さらに、集合  $A$  の要素数を  $\# \{A\}$  と書く。

この時、先行運行時間を探すための目的関数として、2つの点過程間の、ラグ  $h$  ( $\Delta$  の整数倍)だけずらした後の同時発生の相対頻度

$$(3.2) \quad A(h) := \frac{1}{\underline{A}} \# \{ \mathcal{T}^1 \cap \mathcal{T}_{-h}^2 \}$$

を考える。ここで、 $\underline{A} := \# \{ \mathcal{T}^1 \} \wedge \# \{ \mathcal{T}^2 \}$  である。定義より、離散観測グリッド間隔  $\Delta$  内 (Dobrev and Schaumburg, 2015 の実証分析例では1ミリ秒)に複数の注文が発生しても、1件とカウントされる。すなわち、本稿で用いるタイムスタンプのデータセットは、元々の  $\{\tau_i^1\}, \{\tau_j^2\}$  ではなく、それを区間幅  $\Delta$  で離散時間化し注文件数の情報も落とした  $\mathcal{T}^k$  である。本稿では、便宜上、 $A(h)$  を“相互共起強度関数”と呼ぶことにする。

Dobrev and Schaumburg (2015)は、偶然に発生した共起の割合をベースラインとして(3.2)より差引いた

$$(3.3) \quad \tilde{A}(h) := A(h) - \bar{A}$$

の使用を提案している。ここで、 $\bar{A} := \frac{1}{\# \{ \mathcal{G} \}} \sum_{h \in \mathcal{G}} A(h)$ 、予め定めた探索範囲  $\mathcal{G}$  における  $A(h)$  の平均値である。さらに、(3.2)を目的関数として  $\mathcal{G}$  上で最大化する値

$$\hat{\theta} := \arg \max_{h \in \mathcal{G}} \tilde{A}(h)$$

が Dobrev and Schaumburg (2015)の提案した先行運行推定量である。

本稿では、相互共起強度関数の絶対値の大きさ自体には関心を持たず、先行運行関係の計測にフォーカスすることから、(3.3)ではなく(3.2)の最大化を図る。以下では、データより計算された先行運行時間推定値  $\hat{\theta}$  を“DS 指標”と呼ぶことにする。定義から明らかなように、DS 指標においては、タイムスタンプのみを使用して目的関数  $A(h)$  を計算することから、価格情報に含まれるマイクロストラクチャ効果などの要因には直接は影響されず、HRY 指標に比してロバストな先行運行推定結果を与えることが期待される。

次に、本研究における工夫について述べる。先述のように Dobrev and Schaumburg (2015)

では全約定データのタイムスタンプを使用していたが、本研究は注文板データから計算されるマイクロ・プライスを使用することで、注文板の動き(マイクロ・プライスの変化時点)に関する先行運行時間を調べる。さらに、彼らの提案方法は相場の方向性に関する情報が欠落していることから、全タイムスタンプ  $\mathcal{T}^k$  を使うのではなく、価格変化が正であったもの  $\mathcal{T}_B^k := \{t_i \in \mathcal{T}^k; \Delta_i X^k > 0\}$ 、負であったもの  $\mathcal{T}_S^k := \{t_i \in \mathcal{T}^k; \Delta_i X^k < 0\}$  の 2 つのサブデータセットを抽出し、各々に対して DS 指標を計算することにする。ここで、マイクロ・プライスが上昇するケース( $t_i \in \mathcal{T}_B^k$ )とは、定義(3.1)より、最良買気配値または最良売気配値の少なくとも一方が上昇するか(例えば、ある時点の最良売気配がキャンセルまたは成行買約定により消滅することで、最良売気配値が上昇)、最良売気配の数量が減少するか、最良買気配の数量が増加するかのいずれかである。すなわち、相場における買い意欲が勝った“強気”の時点と解釈される。後者はその逆の場合である。このように相場の上下別に DS 指標を算出することで、先行運行関係の(相場の方向性に関する)非対称性についても調べることが可能になる。

なお、現在のところ、Dobrev and Schaumburg (2015)による推定量に関しては統計理論が整備されていない。本研究では、 $X^1, X^2, \tau^1, \tau^2$  に特定の確率過程モデルを仮定せず、記述統計の立場を取りながら分析を進める。

### 3.4 データ加工

先行運行分析は、3市場(東証, JNX, ChiX)の中から3組のペア  $X^1-X^2$  を作り、各ペア毎に行った。使用するデータの時間解像度が  $\Delta = 0.001$  秒であることから、 $N$  の大きさは、東証の立会時間長  $T = 5$  時間(9:00-11:30, 12:30-15:00)に相当する、 $N = 5 \times 60 \times 60 \times 1000 = 18,000,000$  である。一方、 $A(h)$  の最大値の探索範囲は、 $\mathcal{G} = \{-0.250, -0.249, \dots, 0, \dots, 0.249, 0.250\}$  (0.001秒刻み)と設定した。

1日内の先行運行時間が時間の推移と共に変化する現象は、林(2015, 2016)において指摘された。よって、そのような一日内変化を反映させるために、東証の立会時間計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(9:00-9:30)、第2時間帯(9:30-10:00)、...、第10時間帯(14:30-15:00)、のように設定する。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。

分析の前のデータ前処理として、元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てマイクロ・プライスを計算した(最小時間解像度内につき最大1件のデータを使用)。さらに、今回はマイクロ・プライスの更新時点をイベントの発生時点と捉えることから、マイクロ・プライスの計算不能な最良気配が売買両サイドにないレコードは排除した。

つぎに、上記のような改良型の DS 指標を計算するために、マイクロ・プライス系列の変化の方向によって、タイムスタンプのデータセットを、変化が正であったもの、負であったものの2つのデータセットに分割した。なお、変化を伴わない(ゼロ・リターン)のタイムスタンプは除去した。DS 指標は定義上タイムスタンプのみを使用して計算されるため、今回の分析では大きな変化(ジャンプ)の除去、マイクロ・ストラクチャの除去等は行わなかった。

また、回帰分析の際に共変量として用いる観測量の算出の際には、2013年時点の東証ティックサイズの定数倍(今回は20倍)以上のビッド・アスク・スプレッドを持つレコードは、市場実勢から乖離した注文であると見なし計算時のデータセットから除去した。<sup>8)</sup>

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・プライス系列を使って、売り買い2種類のタイムスタンプのデータセット  $\mathcal{T}_S, \mathcal{T}_B$  を生成し、各々より売り方向、買い方向の DS 指標  $\hat{\theta}$  を計算する。得られた、DS

指標についてプロットや要約統計量を計算して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×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。

#### 4. 実証分析結果

##### 4.1 先行運行時間推定値の時系列プロットおよび要約統計量

前節の手順に従って、TOPIX100構成銘柄について、売買方向別に、データ期間内の日別、時間帯別にDS指標 $\hat{\theta}$ を計算した。参考までに、図1は、2014年1月14日におけるみずほFG(8411)について計算された、東証( $X^1$ )-JNX( $X^2$ )間の相互共起強度関数 $A(h)$ を、 $h$ ( $= -0.050, \dots, 0.050; 0.001$ 秒刻み)に対してプロットしたものである。タイムスタンプ・データは買い方向 $T_B$ のみ使用し、見やすさを確保するため第2時間帯( $t_2$ )と第6時間帯( $t_6$ )のケースのみを一前者は実線で後者は破線で一表示する。2本の折れ線グラフとも周辺領域の凹凸の他、原点付近に鋭く尖った頂点(群)が見られる。時間帯2は最高点が負の領域に、(1番目とは高さ差のある)2番目の頂点が正の領域に存在する一方、第6時間帯はほぼ単峰が負領域に存在していると言ってよい。幾つかのデータセットについて目視した限りにおいては、ここで紹介するものと同様に、原点付近に急峻な峰が1個または若干個集中しているような形状となった。この中で最高点をもたらす $h$ の値がDS指標である。

TOPIX100の中でも特に流動性の高いCore30銘柄について、DS指標の日次での時系列推移を調べた。全データ期間485営業日毎に、売買方向別、銘柄別に一日内に(各時間帯毎に得られた)10個ある推定値のメジアンを計算して得られた日次時系列プロットを作成した。以下では東証を“TSE”と表記する。図2-4は、その中の6銘柄(三菱UFJFG(8306)-三井不動産(8801))についての、順に、TSE-JNX, TSE-ChiX, JNX-ChiXの市場ペアのものである。実線は買い方向、破線は売り方向を示す。さらに、2本の水平鎖線は下から0秒、0.004秒、2本の垂直点線は左からフェーズ1導入日(251日目)、フェーズ2導入日(380日目)を示す。売買両方向のプロットは重なりが大きく識別が難しい。

各図の縦軸はDSによる先行運行推定値(単位は秒)、横軸はデータ期間内の日付の通し番号である。各図内には、縦軸0秒には水平の実線、東証ティックサイズ変更の実施された日(1月

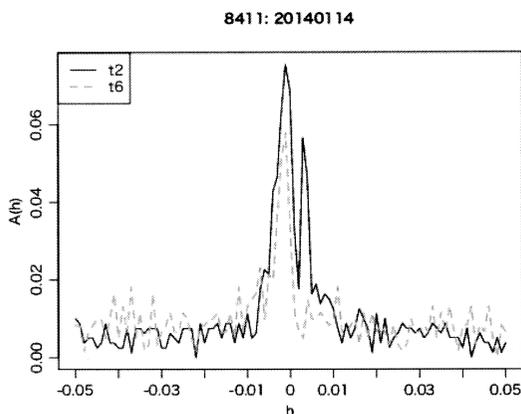


図1. 相互共起強度関数 $A(h)$ の計算例。第2時間帯(実線)と第6時間帯(破線)。

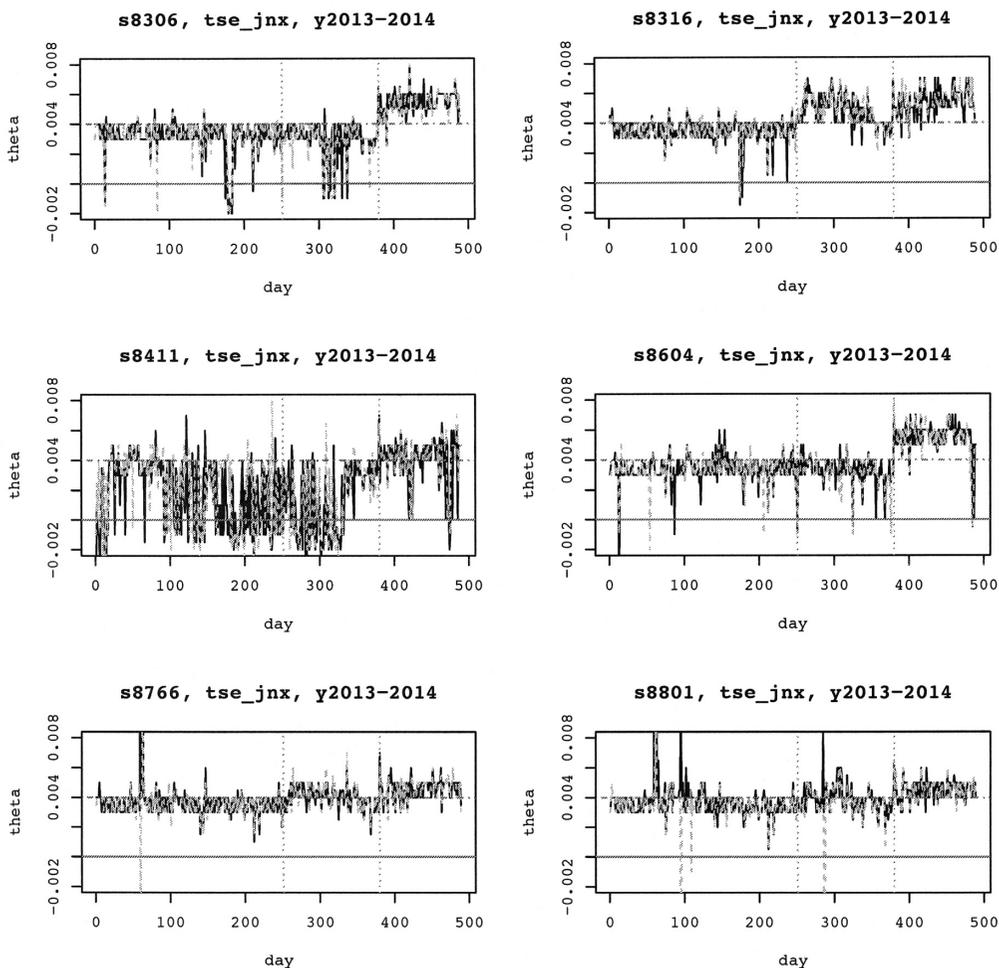


図 2. DS 指標の日々推移(実線は買い方向, 破線は売り方向): 三菱 UFJ フィナンシャルグループ(8306)-三井不動産(8801): TSE-JNX. 2本の水平鎖線は下から 0 秒, 0.004 秒, 2本の垂直点線は左からフェーズ 1 導入日(251 日目), フェーズ 2 導入日(380 日目)を示す.

14 日, 7 月 22 日)に垂直の点線を付している. さらに, 図 2 と図 3 においては, 0.004 秒の位置には水平の点線を付している.

両図より, いずれの銘柄とも, 4 ミリ秒付近を中心に変動していることが観察される. これは, TSE-JNX, TSE-ChiX のペアにおいて, 東証が PTS に対しておおよそ 4 ミリ秒先行していたことを示している. また, 林 (2015, 2016)で報告したのと同様に, フェーズ 1 以前(-2014 年 1 月 10 日), フェーズ 1 適用期間(2014 年 1 月 14 日-7 月 18 日), フェーズ 2 適用期間(7 月 22 日-12 月 30 日)と, 変動の様子に変化の見えることが確認される.

さらに, みずほ FG (8411)の変動の大きさは他銘柄と比べて圧倒的に大きい. 特に, フェーズ 2 導入以前においては, 推定値が負の値を取る日が頻繁に観測される. これは, データ期間中 100 円台で推移していたみずほ FG 株のティックサイズが, フェーズ 2 が実施される前まで

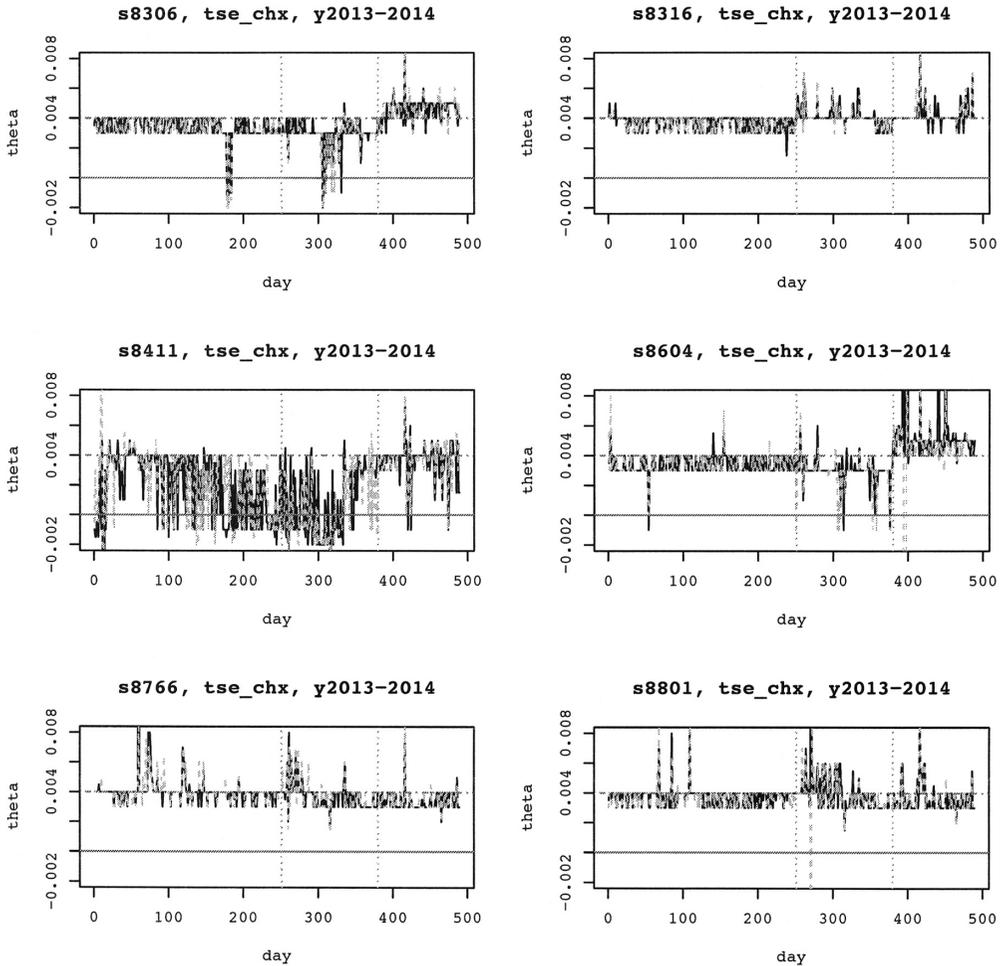


図 3. DS 指標の日々推移(実線は買い方向, 破線は売り方向)：三菱 UFJ フィナンシャルグループ(8306)–三井不動産(8801)：TSE-ChiX. 2本の水平鎖線は下から0秒, 0.004秒, 2本の垂直点線は左からフェーズ1導入日(251日目), フェーズ2導入日(380日目)を示す。

1円に設定されていたことに呼応している。

一方, 図4のJNX-ChiXの組み合わせにおいては, より小さな変動幅で, おおよそ0.000秒(水平実線)の周りを変動している。さらに, よく見ると, フェーズ1ないしはフェーズ2開始後に変動の中心が0.000秒から,  $-0.001$ 秒へ変化していることも分かる。すなわち, 東証のティックサイズ縮小により, ChiXはJNXよりも値動きが速くなったことを表している。

表2-4に, スペース節約のためCore30構成銘柄に限定して, 時間帯や売買方向性の区別をせずに, 各銘柄について先行運行推定値のデータ期間内平均と標準偏差を計算した結果を示す。

フェーズ1期間は過渡期のためか結果は必ずしも安定していない。銘柄間の相違はあるが, TSEティックサイズ変更前後(フェーズ1導入前とフェーズ2導入以降)を比較すると, 上の結果は大雑把に次のように要約できる。

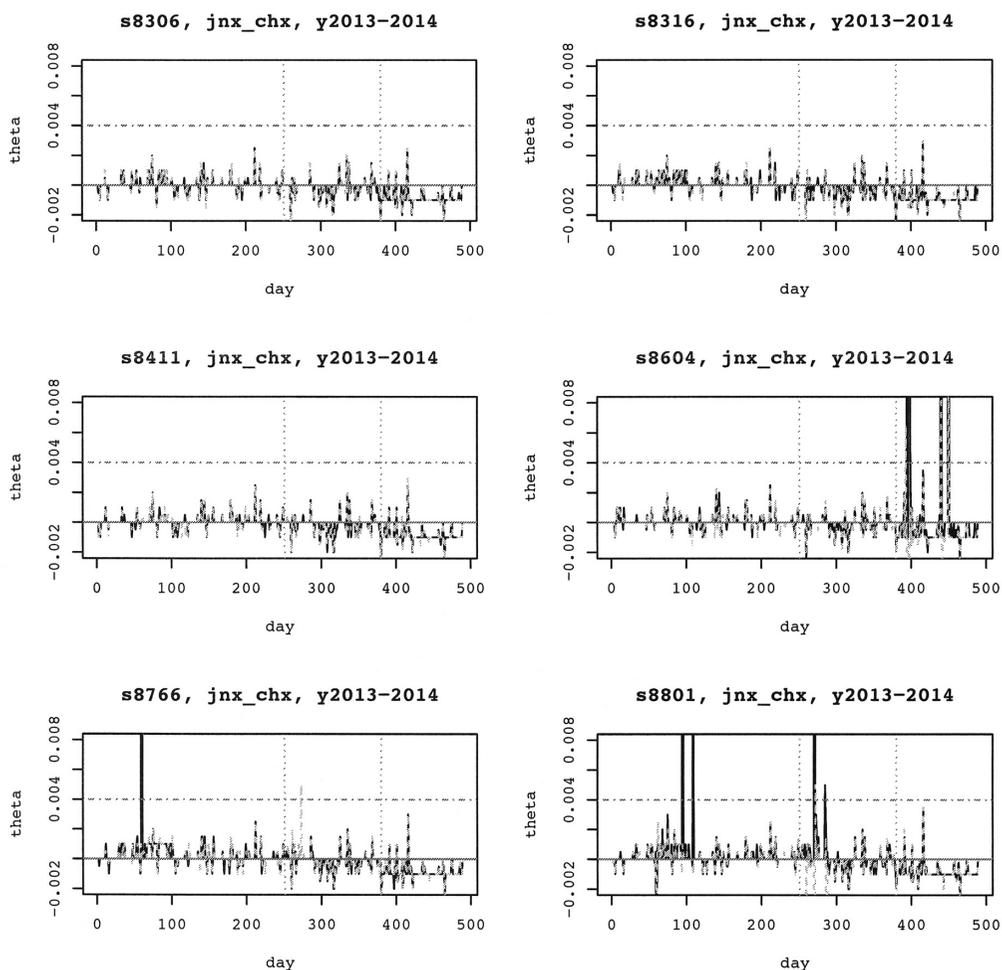


図 4. DS 指標の日次推移(実線は買い方向, 破線は売り方向): 三菱 UFJ フィナンシャルグループ(8306)-三井不動産(8801): JNX-ChiX. 2本の水平鎖線は下から 0 秒, 0.004 秒, 2本の垂点線は左からフェーズ 1 導入日(251 日目), フェーズ 2 導入日(380 日目)を示す.

- TSE-JNX, TSE-ChiX においては, TSE が PTS 二市場に対して, おおよそ 4 ミリ秒先行している.
- TSE のティックサイズ変更により, 特に TSE-JNX において, TSE の先行度合いが増した.
- JNX-ChiX においては, 先行運行時間の差は 1 ミリ秒未満(計測限界以下)である.
- TSE のティックサイズ変更により, JNX-ChiX は JNX の先行から ChiX の先行へと変化した.

これらは, 筆者が先に林(2016)で示した HRY 指標の計測結果とおおむね整合的である. 東証が国内市場シェアの 9 割超を占める中, おそらく全ての市場参加者が東証の値動きを見ながら行動しているであろうことから, 東証が PTS 二市場に先行するという今回の結果は自然であ

表 2. DS 指標の要約統計量(Core30 構成銘柄別, 単位: 秒): TSE-JNX.

code	Period 1 20130104–20140110		Period 2 20140114–20140718		Period 3 20140722–20141230	
	mean	stdev	mean	stdev	mean	stdev
s2914	0.0036	0.0040	0.0045	0.0063	0.0050	0.0011
s3382	0.0036	0.0082	0.0044	0.0066	0.0047	0.0011
s4063	0.0033	0.0013	0.0043	0.0043	0.0046	0.0010
s4502	0.0034	0.0016	0.0046	0.0068	0.0052	0.0013
s4503	0.0033	0.0011	0.0037	0.0061	0.0044	0.0011
s5401	0.0032	0.0019	0.0032	0.0033	0.0043	0.0012
s6301	0.0040	0.0062	0.0038	0.0010	0.0050	0.0012
s6501	0.0033	0.0010	0.0030	0.0015	0.0048	0.0011
s6752	0.0032	0.0081	0.0037	0.0011	0.0047	0.0011
s6758	0.0038	0.0028	0.0039	0.0012	0.0052	0.0011
s6902	0.0033	0.0043	0.0042	0.0012	0.0046	0.0010
s6954	0.0046	0.0208	0.0043	0.0082	0.0045	0.0011
s7201	0.0036	0.0010	0.0033	0.0013	0.0048	0.0016
s7203	0.0031	0.0016	0.0051	0.0022	0.0054	0.0016
s7267	0.0033	0.0015	0.0046	0.0014	0.0055	0.0012
s7751	0.0035	0.0014	0.0046	0.0089	0.0054	0.0013
s8031	0.0037	0.0010	0.0036	0.0015	0.0050	0.0011
s8058	0.0039	0.0088	0.0039	0.0012	0.0052	0.0011
s8306	0.0029	0.0020	0.0027	0.0018	0.0053	0.0010
s8316	0.0034	0.0012	0.0049	0.0066	0.0054	0.0011
s8411	0.0019	0.0090	0.0014	0.0099	0.0039	0.0021
s8604	0.0033	0.0025	0.0029	0.0028	0.0054	0.0018
s8766	0.0035	0.0209	0.0037	0.0102	0.0045	0.0010
s8801	0.0031	0.0276	0.0018	0.0289	0.0044	0.0014
s8802	0.0040	0.0221	0.0037	0.0011	0.0044	0.0009
s9020	0.0036	0.0024	0.0044	0.0071	0.0047	0.0010
s9432	0.0034	0.0014	0.0044	0.0053	0.0049	0.0010
s9433	0.0033	0.0022	0.0047	0.0070	0.0045	0.0010
s9437	0.0041	0.0158	0.0038	0.0014	0.0048	0.0011
s9984	0.0035	0.0011	0.0061	0.0021	0.0059	0.0014

る。一方、今回計測された市場間の全般的な時間差(東証対 PTS は 4 ミリ秒差, PTS 間は 0 ミリ秒差)は何か物理的な大きさを表しているのだろうか。

Dobrev and Schaumburg (2015)において、実証分析の事例として米国 10 年国債現物と S&P500 指数先物(E-Mini)は、約 5 ミリ秒後者が前者に対して先行していると報告している。前者は BrokerTech (New Jersey 州 Secaucus)の、後者は CME (Illinois 州 Aurora)のデータである。光速で 4.7 ミリ秒かかる距離(理論的な下限値)であり、彼らの計測値がそれに近いことを示している。一方、国内 3 市場のマッチング・エンジンは、所在地はセキュリティ上公開されていないが、地理的に互いに近い場所に設置されており、広大な国土を有する米国市場とは状況は異なる。よって恐らく、ブローカーが TSE のマッチング・エンジンより出入するのに経路せねばならない東証のネットワーク(arrownet)に起因する遅延が主要因ではないかと筆者は考える。一方、PTS 二市場間にはそのようなネットワーク上の制約がない。また、東証に届く多

表 3. DS 指標の要約統計量(Core30 構成銘柄別, 単位: 秒): TSE-ChiX.

code	Period 1 20130104–20140110		Period 2 20140114–20140718		Period 3 20140722–20141230	
	mean	stdev	mean	stdev	mean	stdev
s2914	0.0039	0.0009	0.0042	0.0047	0.0040	0.0009
s3382	0.0039	0.0065	0.0038	0.0043	0.0038	0.0009
s4063	0.0033	0.0010	0.0040	0.0060	0.0036	0.0009
s4502	0.0037	0.0011	0.0043	0.0015	0.0043	0.0012
s4503	0.0035	0.0013	0.0036	0.0016	0.0034	0.0008
s5401	0.0034	0.0019	0.0033	0.0069	0.0034	0.0040
s6301	0.0041	0.0040	0.0037	0.0017	0.0040	0.0009
s6501	0.0035	0.0006	0.0030	0.0013	0.0039	0.0012
s6752	0.0035	0.0081	0.0034	0.0009	0.0038	0.0009
s6758	0.0040	0.0043	0.0037	0.0009	0.0043	0.0009
s6902	0.0037	0.0019	0.0039	0.0031	0.0037	0.0009
s6954	0.0041	0.0101	0.0041	0.0047	0.0035	0.0012
s7201	0.0038	0.0007	0.0031	0.0011	0.0040	0.0011
s7203	0.0034	0.0012	0.0047	0.0015	0.0045	0.0019
s7267	0.0037	0.0012	0.0041	0.0014	0.0044	0.0009
s7751	0.0039	0.0009	0.0042	0.0015	0.0044	0.0009
s8031	0.0039	0.0005	0.0035	0.0014	0.0041	0.0009
s8058	0.0041	0.0049	0.0037	0.0010	0.0042	0.0008
s8306	0.0032	0.0014	0.0026	0.0017	0.0045	0.0009
s8316	0.0036	0.0023	0.0042	0.0010	0.0043	0.0009
s8411	0.0019	0.0081	0.0016	0.0154	0.0030	0.0055
s8604	0.0034	0.0020	0.0027	0.0057	0.0059	0.0437
s8766	0.0042	0.0079	0.0038	0.0023	0.0034	0.0008
s8801	0.0039	0.0060	0.0039	0.0164	0.0038	0.0047
s8802	0.0040	0.0131	0.0036	0.0011	0.0034	0.0049
s9020	0.0035	0.0006	0.0043	0.0060	0.0037	0.0013
s9432	0.0036	0.0012	0.0044	0.0096	0.0040	0.0009
s9433	0.0035	0.0009	0.0042	0.0013	0.0035	0.0008
s9437	0.0043	0.0199	0.0035	0.0020	0.0036	0.0009
s9984	0.0037	0.0009	0.0056	0.0016	0.0051	0.0014

種・大量の注文を処理する際に発生する待ち行列の処理時間も遅延要因になっているかもしれない。

次に、表 5 は、TOPIX100 の個別銘柄毎に(時間帯区別なく)計算されたデータ期間内平均値を使って、さらに 100 銘柄全体の平均値と標準偏差を計算したものである(単位ミリ秒)。

さらに、売買方向の相違による先行運行推定時間の差異を、1 日以内での時間帯別に調べる。図 5 は、売買方向別に、各時間帯毎に、DS 指標推定値の全平均(各銘柄の期間内平均の、全銘柄平均)をプロットしたものである。同図内は、上から順に、TSE-JNX, TSE-ChiX, JNX-ChiX, 買い方向は左列、売り方向は右列に配置されている。各プロット内で、3 つのデータ期間の平均値を重ねて表示している。点線がフェーズ 1 導入前(“prd1”), 破線がフェーズ 1 導入期間中(“prd2”), 実線がフェーズ 2 導入後(“prd3”)である。

同図から読み取ることのできる全体的な傾向は次の通りである。

表 4. DS 指標の要約統計量(Core30 構成銘柄別, 単位: 秒): JNX-ChiX.

code	Period 1 20130104–20140110		Period 2 20140114–20140718		Period 3 20140722–20141230	
	mean	stdev	mean	stdev	mean	stdev
s2914	0.0002	0.0028	0.0003	0.0075	-0.0008	0.0009
s3382	0.0004	0.0112	0.0004	0.0085	-0.0008	0.0009
s4063	0.0002	0.0009	0.0003	0.0099	-0.0008	0.0009
s4502	0.0002	0.0009	0.0003	0.0087	-0.0008	0.0010
s4503	0.0001	0.0009	0.0001	0.0081	-0.0008	0.0009
s5401	0.0001	0.0009	-0.0001	0.0009	-0.0007	0.0024
s6301	0.0004	0.0085	-0.0001	0.0009	-0.0008	0.0009
s6501	0.0001	0.0008	-0.0001	0.0009	-0.0006	0.0037
s6752	-0.0001	0.0087	-0.0001	0.0009	-0.0008	0.0009
s6758	0.0001	0.0047	-0.0001	0.0009	-0.0008	0.0009
s6902	0.0004	0.0053	0.0001	0.0051	-0.0008	0.0011
s6954	0.0010	0.0219	0.0004	0.0085	-0.0009	0.0030
s7201	0.0001	0.0009	-0.0001	0.0009	-0.0008	0.0009
s7203	0.0001	0.0009	0.0000	0.0022	-0.0008	0.0009
s7267	0.0001	0.0009	0.0001	0.0049	-0.0008	0.0009
s7751	0.0001	0.0009	0.0008	0.0137	-0.0008	0.0009
s8031	0.0002	0.0008	-0.0001	0.0009	-0.0008	0.0009
s8058	0.0004	0.0095	-0.0001	0.0010	-0.0008	0.0009
s8306	0.0001	0.0008	-0.0001	0.0009	-0.0008	0.0009
s8316	0.0002	0.0009	0.0003	0.0102	-0.0008	0.0009
s8411	0.0002	0.0026	0.0000	0.0022	-0.0008	0.0009
s8604	0.0003	0.0058	0.0001	0.0065	0.0009	0.0452
s8766	0.0001	0.0213	0.0002	0.0132	-0.0008	0.0009
s8801	0.0002	0.0291	-0.0017	0.0334	-0.0002	0.0112
s8802	0.0004	0.0242	0.0000	0.0013	-0.0008	0.0009
s9020	0.0002	0.0009	0.0000	0.0094	-0.0008	0.0009
s9432	0.0002	0.0009	0.0003	0.0107	-0.0008	0.0009
s9433	0.0002	0.0021	0.0002	0.0091	-0.0008	0.0009
s9437	0.0035	0.0325	-0.0001	0.0010	-0.0008	0.0009
s9984	0.0002	0.0009	-0.0001	0.0011	-0.0007	0.0036

表 5. DS 指標の要約統計量(TOPIX100 構成銘柄全体の要約, 単位: ミリ秒).

Period	TSE vs JNX		TSE vs ChiX		JNX vs ChiX	
	Buy orders	Sell orders	Buy orders	Sell orders	Buy orders	Sell orders
Period 1						
mean	4.320	2.990	4.410	3.880	1.590	-0.470
stdev	2.150	2.010	2.430	1.510	2.650	1.610
Period 2						
mean	3.860	3.360	3.680	3.570	0.360	-0.360
stdev	0.860	1.650	0.720	0.590	0.980	1.600
Period 3						
mean	4.530	4.560	3.690	3.600	-0.630	-0.820
stdev	0.530	0.520	0.810	0.490	0.600	0.350

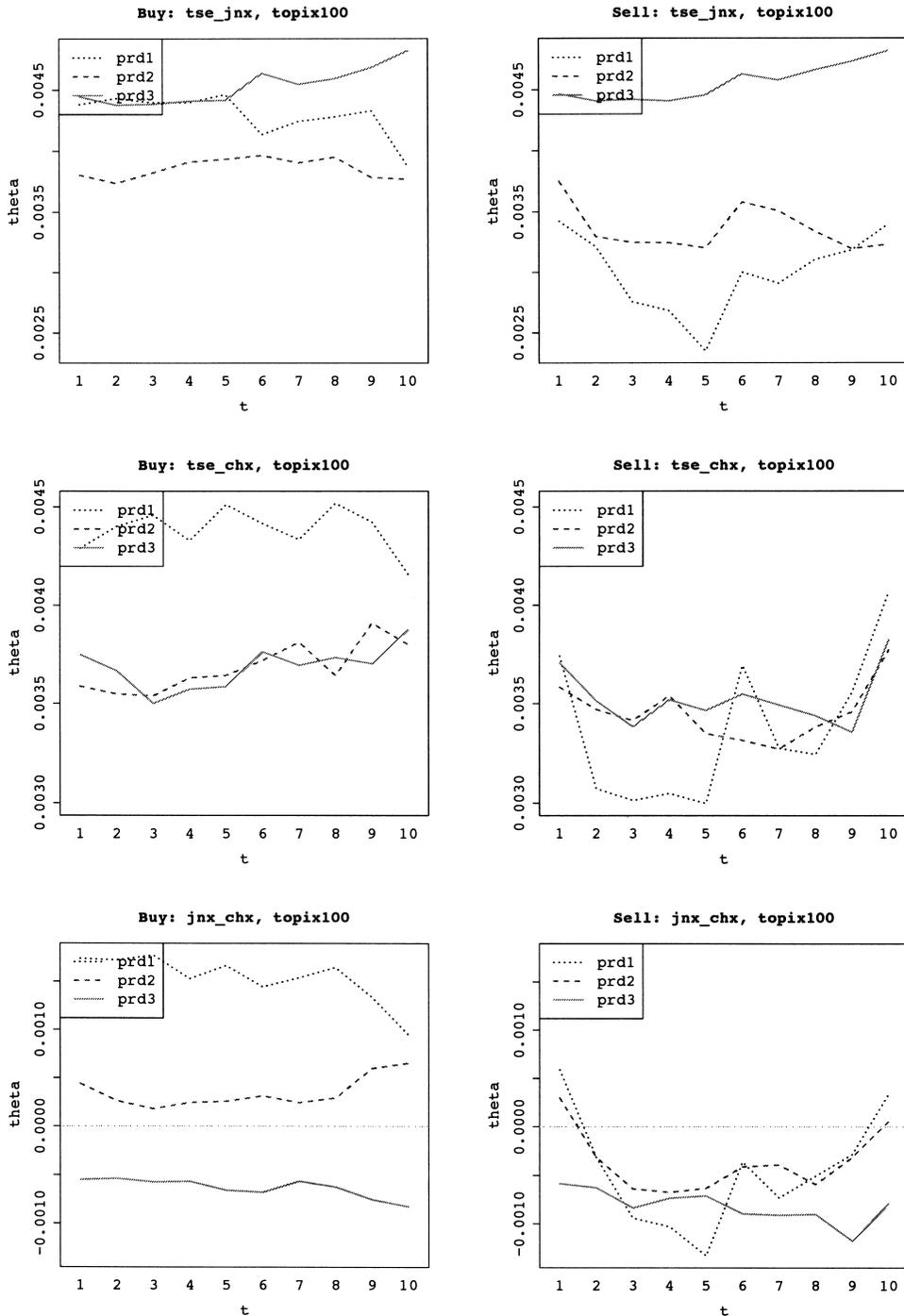


図5. 個別銘柄 DS 指標の、各時間帯(横軸)におけるデータ期間内の銘柄全体平均値(縦軸)(TOPIX100 銘柄): TSE-JNX(上段), TSE-ChiX(中段), JNX-ChiX(下段). 買い方向(左列), 売り方向(右列).

- フェーズ 1 導入前は、売買方向間の非対称性が大きい。特に、売り方向は買い方向に比べて先行度合いがマイナス方向に位置していた。
- フェーズ 2 導入後は、先行度合いの売買方向差は縮小し、ほぼ類似の 1 日内推移パターンを示すようになった。

また、

- TSE-JNX では、TSE のティックサイズ変更後、TSE の先行度合いが、特に売り方向側で大幅に増加した。
- TSE-ChiX では、TSE のティックサイズ変更後、TSE の先行度合いが、特に買い方向側で減少した(または、ChiX の運行度合いが、買い方向側で減少した)。
- JNX-ChiX では、TSE のティックサイズ変更後、買い方向では ChiX が運行から先行に変化し、売り方向では ChiX の先行度合いが増加した。

さらに、一日内推移を見ると、

- 売り方向では“U字型”が見られる(寄付後と大引け前において、TSE-PTS 間は TSE の先行度大、JNX-ChiX 間は両者の差小)。フェーズ 1 導入以前は時間帯による差異がより顕著だった。
- 買い方向は、売り方向ほど顕著ではない。フェーズ 1 導入以前は、上下しながら右肩下がりの“逆 U 字(または逆 J 字)型”を示していた(3つのペアとも、大引けにかけて先行市場側の先行度縮小)。一方、フェーズ 2 以降は、TSE-PTS 間において右肩上がり、すなわち、大引けに向かいむしろ TSE の先行度が増加する様子が見られる。

確認のため、売買方向の違いによる先行度合いの相違に関して、市場ペア別/データ期間別/時間帯別に、平均値の差に関する Wilcoxon 符号順位和検定(ペア検定)を行ったところ(全部で  $3 \times 3 \times 10 = 90$  通り)、有意に差が見られるケースが随所に見られた(ここでは省略)。

なお、図 2-4 でも現れたような、稀に発生する極端な値(外れ値)が平均値に大きな影響を与えている可能性がある。実際、平均値の代わりにメジアンを使ったプロットではここで観察されたような時間帯推移は殆ど見られなかった。一方、ロットサイズの大きな注文の執行には時間がかかるはずで、そのような先行運行時間が適切に反映された数値の評価も意義がある。このような極端な値を考慮したきめ細かい分析は今後の課題であるが、東証のフェーズ 2 導入によって、先行運行時間に関する売買の方向性間の非対称性は大きく減少したとは言えそうだ。

## 4.2 回帰分析

### 変数とモデル

前節では DS 指標による市場ペア間の先行運行時間の推定結果について報告した。DS 指標はジャンプやマイクロストラクチャ・ノイズを除去しないプレーンな HRY 指標に比べて安定した計測結果を与えることが確認された。

本節では、DS 指標を被説明変数、各銘柄の取引に関する市場観測量を説明変数とする回帰分析を行うことで、市場ペア間での価格形成の先行運行関係の要因を調べることにする。その際、銘柄固有の要因や日次の要因による影響は除去して考える。

銘柄のカバレッジとして、上記 Core30 銘柄を包含する TOPIX100 構成銘柄(2013 年末)を用いる。これらは、全て 2014 年の東証ティックサイズ変更の対象である。

先に、データ期間内中央値に集約した先行運行関係において 1 日内変化が見られたことから、ここでも、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セットを

用いる。

また、上述のように、売り方向(“S”)と買い方向(“B”)での先行運行推定時間の差が見られることから、2つのDS指標データセットを縦方向に並べ、相場の方向性を識別する2水準の固定因子を導入する。

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DS指標間の相関の影響可能性を考慮するために、林(2015, 2016)と同様に線形混合効果モデル(linear mixed effects model)を用いることにする。

被説明変数として、第*i*銘柄( $i = 1, \dots, 100$ ), 第*j*日( $j = 1, \dots, 489$ ), 第*t*時間帯( $t = 1, \dots, 10$ )におけるDS指標を使用し、 $y_{ijt}$ と書く。 $y_{ijt} > 0$ は、市場 $X^1$ が市場 $X^2$ に先行することを、 $y_{ijt} < 0$ は逆を示す。今回の報告では、東証の開場閉場前後の“非定常”な変動を除くため、全10個のうち、前場後場の寄付後30分および引け前30分の時間帯を除去し、それ以外の6個の時間帯のデータのみを使用する( $t = 2, 3, 4, 7, 8, 9$ )。従って、BまたはSの一方方向のデータをパネルデータとして、全期間分を行方向に時間、列方向に銘柄を(仮想的に)配置すると、行数 $2934 (= 1 \text{日} \times 6 \text{時間帯} \times 489 \text{営業日}) \times \text{列数} \times 100$ の大きさを持つ行列データとなる。

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る。林(2015, 2016)で報告したように、2014年の東証ティックサイズ変更は対象群であるTOPIX100銘柄を中心に個別銘柄の挙動に大きな変化をもたらしたことから、図5と同様に、データセットを(I)期間1:フェーズ1適用前(20130104-20140110, 250営業日), (II)期間2:フェーズ1適用後フェーズ2適用前(20140114-20140722, 129営業日), (III)期間3:フェーズ2適用後(20140722-20141230, 110営業日)の3つの期間に分割して分析を行う。

質的説明変数(固定因子)としては、1日内時間変化を捉えるための6水準因子、売買の方向性を示す2水準因子を加える。さらに、日次効果を表す変数と、銘柄間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(random effect)として導入することにする。

量的説明変数としては、各個別銘柄の最良気配値、最良気配枚数、最良気配更新時刻のタイムスタンプ、約定価格、約定枚数、約定時刻のタイムスタンプから計算される、市場のクオリティや流動性を表すと考えられる代表的な指標を選ぶ。次に説明するように、2種類の量的変数のグループを導入する。

グループ1として、林(2015, 2016)と同様に、今回の分析では、市場間の先行運行時間は二つの市場間の相対的な特徴の差異によって生成されると見る立場から、両市場の特徴量の対数比を取ることで新たな変数を定義する。例えば、第*i*銘柄、第*j*日の第*t*時間帯における二つの市場 $X^1, X^2$ における実現ボラティリティが $RV_{ijt}^1, RV_{ijt}^2$ であれば、 $RVr_{ijt} = \ln(RV_{ijt}^1/RV_{ijt}^2)$ と言った具合である。すなわち、 $RVr_{ijt} > 0$  ( $< 0$ )であれば、市場 $X^1$ で計測された実現ボラティリティが、市場 $X^2$ よりも大きかった(小さかった)ことを示す。なお、グループ1の変数群は対数比を取っていることから、かりに市場ペア $X^1$ と $X^2$ の役割を入れ替えて変数を作り回帰を行っても、結果は変わらない。

一方、このような市場ペア間の相対的な特徴量以外の、その時点におけるマクロ的な、市場全体の要因が先行運行時間に影響を与えていることも想定される。そこで、それらをコントロールするために、グループ2として各銘柄の3市場横断的な相場状況を表す指標を導入する。

本稿では、これら全ての量的変数に対し、期間毎、銘柄毎に標準化を行ったものを報告する。これは、期間毎、銘柄毎に各変数の中心の値やバラツキの大きさに差異が見られたことに対する措置であるが、異なる標準化処理を施したものは、ここでの報告とは異なる回帰分析結果となっており、異なる結果の解釈が必要となる。ここでの中心化の処理により、*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ような仮定の銘柄に対するベースライン効果(第

2時間帯, 買い方向)を秒単位で示したものである(共変量の値がゼロとなるとは, グループ1の量的変数群においては, 市場1と市場2の観測量の比が, グループ2の量的変数群においては, 3市場全体あるいは東証の観測量が, 各々の平均値に等しい). この時, 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.<sup>9)</sup>

分析の実行には, 統計パッケージRのパッケージ‘lme4’内にある関数lmer()を使用した. 複数の候補変数群の中から, 変数間の相関係数の大きさ, 回帰係数の有意性やAICの値を確認しつつ, さらに解釈容易性も勘案しながら, 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.<sup>10)</sup>

$$(4.1) \quad y \sim \underbrace{(1|Code) + (1|Ymd)}_{\text{random effects}} + \underbrace{T}_{\text{6-level factor}} + \underbrace{Dir}_{\text{2-level factor}} \\ * \underbrace{(RVr + NQr + LTSr + SPRr + QBr + QAr + RVag + RETts + TVag)}_{\text{covariates}}$$

ここで,  $RVr, \dots, TVag$ は固定効果(fixed effects),  $Code$ は個別銘柄効果,  $Ymd$ は日次効果であり共に変量効果(random effects)である. 但し,  $RVr$ は, 各市場の仲値より計算される実現ボラティリティの(市場ペア間の)対数比,  $NQr$ は最良気配更新数(ミリ秒単位に集約後の)対数比,  $LTSr$ は平均約定枚数(約定枚数÷約定回数)の対数比,  $SPRr$ はスプレッド率(ビッド・アスク・スプレッド÷仲値)の対数比,  $QBr$  ( $QAr$ )は, 最良買気配(売気配)の数量(“デブス”)の対数比である. これらはグループ1の量的変数群である.

また, 市場ペア間の相対比を取らない各個別銘柄の量的変数(グループ2の変数群)として,  $RVag$ は3市場を一つの市場と見た場合の最良売買気配値(買いサイドは3市場の最良買気配値の最大値, 売りサイドは最良売気配値の最小値)間の仲値より計算された実現ボラティリティ,  $RETts$ は東証仲値の各時間帯始値と終値を使って計算される対数収益率,  $TVag$ は3市場全体の約定枚数合計の対数値である.<sup>11)</sup> なお, スプレッド率比( $SPRr$ )および2つのデブス比( $QBr, QAr$ )については, 対数比を取るのに使用した各市場でのスプレッドやデブスの値は, 当該時間帯内における時間加重平均値(次の気配更新までの経過時間がウェイト)である.<sup>12)</sup>

$T$ は時間帯を表す6水準固定因子( $T = 2, 3, 4, 7, 8, 9$ . 例えば, 水準2は, 第2時間帯(9時半-10時))である.  $Dir$ は売買の方向性を示す2水準因子( $Dir = B, S$ )である. さらに, 相場方向性による回帰係数の違いの有無を捉えるため,  $Dir$ と各共変量の交互作用項(interaction terms)もモデルに加えた. その他の交互作用項はモデル構築の段階で除去した.<sup>13)</sup>

### 分析結果

回帰分析の結果は, 表6-8に記載する. 表6は, データ期間1(20130104-20140110), 表7は, データ期間2(20140114-20140720), 表8はデータ期間3(20140722-20141230)の推定結果である. 得られた幾つの特徴的な観察結果について言及したい. 全般に, 3市場ペア, TSE-JNX(左列), TSE-ChiX(中央列), JNX-ChiX(右列)に共通する幾つかのパターンが見られる. 但し, 林(2015, 2016)でも指摘したように, PTS同士であるJNX-ChiXペアにおける推定結果は, 他の市場ペアとはやや乖離している. また, 例えば, データ期間3におけるTSE-JNXの共変量の回帰係数の値は, 前2期間に比べ桁が一つ小さくなっているなど, 同じ市場ペアであってもデータ期間の相違による変化も認められる.

先述のように,  $y$ -切片の値は, 全ての共変量の値が同時にゼロとなるような仮想銘柄に關する, ベースライン(「第2時間帯( $T = 2$ )」かつ「買い方向( $Dir = B$ )」)のデータ群に対する先行運行時間を示している. 表内の「 $DirS$ 」は, 「売り方向( $Dir = S$ )」の固定因子効果( $y$ -切片の変化値), 「 $t3$ 」-「 $t9$ 」は第3時間帯( $T = 3$ )から第9時間帯( $T = 9$ )までの各固定因子効果の大

表 6. 回帰分析結果：期間 1 (20130104–20140110).

Period 1	(I)	TSE-JNX	(II)	TSE-ChiX	(III)	JNX-ChiX
Fixed effects	Estimate	t value	Estimate	t value	Estimate	t value
(Intercept)	0.00341	20.722	0.00373	18.420	0.00090	5.719
RVr	0.00083	26.724	0.00100	32.399	0.00036	9.479
NQr	0.00001	0.302	-0.00012	-3.839	-0.00050	-12.881
LTSr	-0.00033	-9.412	-0.00029	-8.769	0.00007	2.185
SPRr	-0.00068	-18.460	-0.00076	-20.547	-0.00047	-12.004
QBr	0.00039	13.166	0.00012	3.815	-0.00036	-10.317
QAr	0.00050	16.254	0.00086	26.802	0.00044	12.208
RVag	0.00041	8.253	0.00051	10.417	0.00042	7.635
RETts	-0.00003	-1.136	-0.00008	-2.691	-0.00003	-0.842
TVag	-0.00081	-18.605	-0.00112	-26.327	-0.00052	-11.412
DirS	0.00000	0.055	-0.00004	-1.035	-0.00006	-1.381
t3	0.00008	1.313	0.00002	0.304	-0.00012	-1.661
t4	0.00008	1.246	-0.00002	-0.372	-0.00024	-3.173
t7	0.00022	3.317	0.00018	2.765	-0.00002	-0.264
t8	0.00032	4.817	0.00041	6.325	0.00003	0.400
t9	0.00074	10.930	0.00069	10.558	0.00007	0.959
RVr:DirS	-0.00002	-0.568	-0.00014	-3.351	-0.00022	-4.153
NQr:DirS	0.00000	0.064	0.00005	1.098	-0.00004	-0.666
LTSr:DirS	-0.00001	-0.192	0.00010	2.156	-0.00006	-1.395
SPRr:DirS	0.00004	0.836	0.00010	1.927	0.00011	2.096
QBr:DirS	0.00008	1.837	0.00005	1.179	-0.00002	-0.450
QAr:DirS	-0.00016	-3.766	-0.00020	-4.521	0.00001	0.280
RVag:DirS	0.00003	0.408	-0.00005	-0.778	-0.00003	-0.438
RETts:DirS	0.00016	3.884	0.00013	3.349	0.00010	2.021
TVag:DirS	0.00001	0.187	0.00010	1.752	0.00008	1.290
Random effects	Std.Dev.		Std.Dev.		Std.Dev.	
ynd	0.00075		0.00065		0.00081	
code	0.00149		0.00192		0.00136	
Resid.	0.00973		0.00949		0.01148	
Model fitness						
AIC	-1829041.04		-1851469.30		-1759206.56	
BIC	-1828745.38		-1851173.51		-1758910.51	
logLik	914548.52		925762.65		879631.28	
deviance	-1829097.04		-1851525.30		-1759262.56	
df.resid	284665		285999		288660	

きさ(ベースラインとの差)を示している。時間効果については、期間 1, 期間 2 においては、TSE-JNX, TSE-ChiX とともに、後場の時間経過( $T = 7, 8, 9$ )と共に、推定値(ベースラインとの差)が正に拡大し、かつ有意性も高まる。TSE の PTS に対する先行度合いが第 2 時間帯と比べて高まることを示している。一方、データ期間 3 においては、TSE-JNX はデータ期間 1, 2 に見られるパターンを維持したが、TSE-ChiX では消滅した。期間 3 において、ChiX の市場参加者の行動、すなわち注文板形成が大きく変化した可能性がある。売買方向性に関しては、有意なケースは期間 3 の TSE-JNX のみであった(売り ( $Dir = S$ )) の場合が、買いの場合に比べて

表 7. 回帰分析結果：期間 2 (20140114-20140720).

Period 2	(I) TSE-JNX		(II) TSE-ChiX		(III) JNX-ChiX	
Fixed effects	Estimate	t value	Estimate	t value	Estimate	t value
(Intercept)	0.00359	36.315	0.00345	41.671	0.00019	2.148
RVr	0.00048	23.521	0.00046	21.170	0.00022	7.235
NQr	0.00022	10.074	0.00000	-0.181	0.00007	2.291
LTSr	-0.00009	-3.807	-0.00011	-4.886	-0.00004	-1.515
SPRr	-0.00037	-15.019	-0.00038	-15.201	-0.00023	-6.895
QBr	0.00024	11.847	-0.00002	-1.184	-0.00031	-11.762
QAr	0.00021	10.228	0.00030	14.119	0.00030	10.982
RVag	0.00004	1.071	-0.00004	-1.031	-0.00004	-0.871
RETts	0.00000	-0.058	-0.00006	-2.656	-0.00005	-1.706
TVag	-0.00023	-7.145	-0.00011	-3.435	0.00004	1.101
DirS	0.00002	0.675	0.00000	-0.047	-0.00005	-1.370
t3	0.00004	0.823	-0.00002	-0.406	-0.00009	-1.561
t4	0.00009	2.154	0.00001	0.216	-0.00009	-1.586
t7	0.00009	2.005	0.00007	1.676	-0.00001	-0.149
t8	0.00018	4.030	0.00009	2.054	0.00001	0.084
t9	0.00017	3.683	0.00021	4.688	0.00018	2.927
RVr:DirS	-0.00004	-1.447	0.00004	1.371	0.00005	1.210
NQr:DirS	-0.00014	-4.733	-0.00004	-1.314	0.00003	0.799
LTSr:DirS	0.00008	2.278	0.00004	1.106	-0.00002	-0.703
SPRr:DirS	0.00004	1.046	0.00001	0.310	0.00007	1.571
QBr:DirS	-0.00005	-1.720	0.00000	-0.136	0.00003	0.786
QAr:DirS	-0.00010	-3.536	0.00006	1.904	0.00002	0.440
RVag:DirS	0.00000	-0.058	-0.00005	-1.104	0.00002	0.339
RETts:DirS	0.00004	1.505	0.00008	2.726	0.00011	2.783
TVag:DirS	0.00005	1.127	-0.00001	-0.127	0.00000	-0.047
Random effects	Std.Dev.		Std.Dev.		Std.Dev.	
ynd	0.00062		0.00056		0.00068	
code	0.00075		0.00057		0.0005	
Resid.	0.00474		0.00471		0.00641	
Model fitness						
AIC	-1158093.51		-1160800.89		-1089535.44	
BIC	-1157816.30		-1160523.65		-1089257.71	
logLik	579074.76		580428.44		544795.72	
deviance	-1158149.51		-1160856.89		-1089591.44	
df.resid	147313		147416		150087	

TSE の対 JNX 先行度が微かながら (0.05 ミリ秒) 大きい。但し、他のデータ期間や他の組み合わせでは有意な結果が得られている訳ではなく、解釈やモデルの改良も含め追加の調査検証が必要である。

共変量に対する推定結果は、ベースライン(特に、買いサイド)のデータ群についてのものである。細部に差異はあるものの、データ期間の違いによらず、概ね回帰係数の符号や大きさ、 $t$  値の大きさに類似性が見られる。特にデータ期間 1 とデータ期間 2 の類似性は高く、データ期間 3 はこれらとは若干の乖離が認められる。

表 8. 回帰分析結果：期間 3 (20140722-20141230).

Period 3	(I)	TSE-JNX	(II)	TSE-ChiX	(III)	JNX-ChiX
Fixed effects	Estimate	t value	Estimate	t value	Estimate	t value
(Intercept)	0.00436	57.539	0.00357	43.634	-0.00058	-7.113
RVr	0.00002	2.069	0.00029	16.534	0.00028	14.527
NQr	-0.00007	-7.245	0.00024	12.706	0.00023	11.331
LTSr	0.00001	0.742	-0.00007	-3.549	-0.00004	-2.204
SPRr	-0.00002	-1.980	-0.00012	-5.451	-0.00010	-4.229
QBr	0.00007	8.673	-0.00001	-0.436	0.00001	0.755
QAr	0.00007	8.002	0.00039	22.371	0.00027	14.960
RVag	0.00015	10.606	0.00013	4.656	0.00004	1.492
RETts	0.00000	-0.249	-0.00004	-2.201	-0.00003	-1.683
TVag	0.00001	0.902	-0.00014	-5.147	-0.00002	-0.575
DirS	0.00005	3.867	0.00002	0.664	-0.00004	-1.718
t3	0.00007	4.021	-0.00003	-0.751	-0.00002	-0.649
t4	0.00012	6.913	0.00000	0.061	-0.00005	-1.334
t7	0.00028	15.498	-0.00001	-0.411	-0.00018	-4.654
t8	0.00034	18.516	0.00004	1.040	-0.00017	-4.417
t9	0.00039	21.275	0.00001	0.352	-0.00026	-6.838
RVr:DirS	-0.00002	-1.115	-0.00014	-5.809	-0.00007	-2.492
NQr:DirS	0.00002	1.889	-0.00010	-4.068	-0.00013	-4.783
LTSr:DirS	0.00003	1.900	0.00003	0.947	0.00002	0.705
SPRr:DirS	0.00000	-0.251	0.00001	0.197	-0.00005	-1.379
QBr:DirS	-0.00004	-3.578	0.00008	3.235	0.00001	0.537
QAr:DirS	-0.00001	-1.151	-0.00024	-9.728	-0.00004	-1.682
RVag:DirS	-0.00007	-3.590	-0.00006	-1.689	0.00001	0.147
RETts:DirS	0.00002	1.734	0.00008	3.380	0.00003	1.193
TVag:DirS	0.00005	2.451	0.00010	2.804	0.00000	0.048
Random effects	Std.Dev.		Std.Dev.		Std.Dev.	
ynd	0.00055		0.00053		0.00067	
code	0.00053		0.00058		0.00040	
Resid.	0.00177		0.00342		0.00374	
Model fitness						
AIC	-1236026.56		-1069524.74		-1047396.14	
BIC	-1235753.79		-1069251.98		-1047123.38	
logLik	618041.28		534790.37		523726.07	
deviance	-1236082.56		-1069580.74		-1047452.14	
df.resid	125684		125644		125656	

回帰係数の絶対値が大きい(すなわち被説明変数への影響度の大きい),あるいは,有意性の高い変数としてまず主効果の中から幾つか挙げてみよう. まず, 実現ボラティリティ比(RVr)は, いずれの期間, いずれの市場ペアの組み合わせ(全9通り)においても, 有意に正である. すなわち, 一方の市場でのボラティリティが相対的に高まる状況では, その市場の板形成の速さ(価格変化スピード)がもう一つの市場に対して相対的に高まることを示唆している. 例えば, 期間 1, TSE(市場  $X^1$ )-ChiX(市場  $X^2$ )のペアのケースにおいて, RVr の値が  $+1\sigma$  だけ平均値より大きくなると,  $+0.001$  秒(1 ミリ秒),  $X^1$  の  $X^2$  に対する先行時間が増す(または, 運行

表 9. 売買の方向 ( $Dir = B, S$ ) 別, 東証リターン ( $RETts$ ) と被説明変数  $y$  の変化の関係。

	買い方向 (“B”)	売り方向 (“S”)
東証上昇時間帯 ( $RETts > 0$ )	$\Delta y < 0$	$\Delta y > 0$
東証下落時間帯 ( $RETts < 0$ )	$\Delta y > 0$	$\Delta y < 0$

時間が減る)。上述のように、市場ペア  $X^1$  と  $X^2$  の役割を入れ替えても解釈は変わらない。

ところで、今回は、実現ボラティリティに関する変数として、 $RVr$  と  $RVag$  の 2 つの変数を採用した。もし実現ボラティリティが潜在的な企業価値変化の大きさを表す量 (の推定値) であるとの立場にたてば、市場に依らない共通の値 (後者) のみの採用が適切であろう。<sup>14)</sup> しかしながら、今回の分析では、注文板形成の背後にある注文フローの先行運行関係に興味があり、マイクロストラクチャ・プロセス自体が重要である。よって、筆者による先の分析 (林, 2015, 2016) と同様に、市場間の実現ボラティリティの比  $RVr$  を変数に加えるのは適切であると考えられる。実際、今回の分析でも  $RVr$  は先行運行時間に対して有意に影響していることが示された。

アスク側デプス比 ( $QAr$ ) も、全てのケースにおいて、正に有意であり、かつ、影響度の大きさは  $RVr$  と同程度である。一方、ビッド側デプス比 ( $QBr$ ) の符号や有意性は組み合わせによって安定していない。相対的にアスク側、すなわち売り希望の指値注文量 (デプス) が多い方が先行度合いが高い (運行度合いが低い) と言える。

1 取引当り約定枚数 (ロットサイズ) 比 ( $LTSr$ ) は、1 つの組み合わせ (期間 3, TSE-JNX) を除き、いずれも負に有意である。即時執行取引の注文サイズが相対的に大きい方が先行度合いが低い (運行度合いが高い) ことを示す。影響度の大きさは  $RVr$  の 1/3-1/7 程度である。

スプレッド率比 ( $SPRr$ ) は、1 ケースを除き、負に有意である。スプレッドが相対的に広い方が先行度合いが低い (運行度合いが高い) ことを示す。インパクトの大きさは  $RVr$  と同程度である。

グループ 2 の変数は、 $X^1$  と  $X^2$  の指定順序に係数が依存する変数であり解釈に留意が必要であるが、9 ケースを通じて、係数の符号や大きさが安定しているとは言いがたい。データ期間 1, 2, TSE-JNX, TSE-ChiX において、3 市場全体の約定枚数 ( $TVr$ ) の係数が有意に負である。実質的に、東証の約定枚数が大きい状況では、TSE の先行度合いが減少すること (すなわち、PTS の価格形成が通常に比べて早まる) を示していると思われる。時間帯内の東証リターン ( $RETts$ ) は、有意なのは TSE-ChiX の 3 ケースのみであるが、全てのケースにおいて負の係数を持つ。すなわち、買い方向 (ベースライン) に関しては、TSE の上昇率が高いほど、TSE-JNX, TSE-ChiX 間では PTS 側の、他方 JNX-ChiX 間では ChiX 側の先行度合いが増す (または、運行度合いが減る) 傾向を示していると考えられる。

次に、相場方向性 ( $Dir$ ) と共変量間の交互作用項であるが、主効果に比べ、安定的なパターンは見い出せない。そもそも DS 指標の、買い方向 (“B”) と売り方向 (“S”) の値の差が小さいために、売買方向性 ( $Dir$ ) の違いが回帰係数の変化に現れにくかったのかもしれない。全般に、主効果に (売買方向性との) 交互作用効果を加えて得られる売りサイドの係数の解釈は、買いサイドのものと同様と言える。

唯一特筆すべき共通パターンとして、東証リターン ( $RETts$ ) との交互作用項は 9 ケースとも正 (内 6 ケースで有意) である。交互作用の絶対値の大きさを考慮すると、売り方向の  $RETts$  の回帰係数 (主効果) は、1 ケース (JNX-ChiX, 期間 3) を除き全て正に反転することが確認される。先の買い方向の結果と合わせると、表 9 のようにまとめられる。

すなわち、“順張り” 方向 (相場上昇時間帯の買い方向、相場下落時間帯の売り方向) では、 $X^1$  の先行度合い減 (または運行度合い増) の、“逆張り” 方向 (相場上昇時間帯の売り方向、相場下

落時間帯の買い方向)では、 $X^1$ の先行度合い増(または遅行度合い減)の、ある意味対称な現象が観察される。相対的に注文板の薄いPTS市場においては、よりマーケット・メイキングを行うHFTの注文の割合が高く、よって、相場の方向性に対するマイクロ・プライスの感応度が高いことを表しているのかもしれない。

その他の有意な交互作用効果も含め、解釈は今後の課題である。

以上の結果は、林(2015, 2016)と係数の符号が異なるものも含まれる。例えば、林(2016)ではスプレッド率比の回帰係数が正であった。同論文では留意点として、“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、 $S_{pr}$ の符号は負であるべきとの見方も可能であろう”と述べた。

もとよりモデルが違うのであり回帰係数の符号が異なること自体は不自然ではない。そもそも、被説明変数は先の報告では異なる方法が採用されていた。先の報告で採用したHoffmann et al. (2013)の方法論によって計算されるHRY指標は大きな価格変動(ジャンプ)に影響されやすいことが分かっている。しかしながら、林(2015, 2016)においては、HRY指標を計測するために計算したマイクロ・プライス系列に対して、ジャンプを除去する処理は行っていなかった。DS指標においても、データ前処理方法に対する依存性は依然として課題ではあるが、その影響はHRY指標よりも小さいと思われる。

いずれにせよ、今回の分析における変数選択は最終的なものではない。例えば、他変数との相関性等を考慮して採用しなかったが、市場間の“ティックサイズ対仲値比”や“出来高比”は、結果の解釈に有用な変数とも考えられる。<sup>15)</sup>

本論文の結果に関する検証作業や、分析方法の見直し・改善作業は、今後とも継続的に行う必要がある。

#### 今後の課題

現実には、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と想定される。このような“マルチラグ・モデル”は加藤 他(2011)においてHRYの枠組みの拡張として検討された。筆者は、目下、ウェーブレット分析を応用した、異なる時間解像度に同時に存在する複数の先行遅行時間を同時推定する新たな手法を開発中である(Hayashi and Koike, 2016)。

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。特に、今回、売買注文の方向性による先行遅行時間の非対称性が観察された。売買注文の偏りは、市場の価格発見機能、情報トレーダーとノイズトレーダーの割合(PIN)、市場の効率性やクオリティ等に関わる重要な概念であり、注文板上の注文枚数の不均衡(quote imbalance)、約定の不均衡(trade imbalance)等の数量を用いた理論研究や実証研究が盛んに行われている(例、Chung and Hrazdil, 2010; Chen et al., 2013; Lipton et al., 2013; Bechler and Ludkovski, 2014)。今回の観察は、これらの分野における本研究で採用した手法の適用可能性を示していると考えられる。

今回の分析では、マイクロ・プライスの更新時間から作られるタイムスタンプ・データを使って先行遅行時間を推定した。一方、筆者の行った予備分析では、約定時点のみから成るタイムスタンプ・データを使って計算された先行遅行時間は、今回のものとは大きく異なるものとなった。これについては目下分析作業中であり、別の機会に報告することとしたい。

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、注文板形成の先行運行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、Dobrev and Schaumburg (2015)による先行運行関係推定法の改良版を用いて、銘柄別時間帯別に先行運行時間を推定し、時系列プロットを作成すると共に、要約統計量を計算した。次に、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)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運行要因を探った。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高い100銘柄から成るTOPIX100構成銘柄、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。

主な実証的発見として、平均的には、銘柄や時間帯による違いはあるものの、3市場間の先行運行関係として、全般に東証はPTS二市場に4ミリ秒程度前後の大きさで先行する傾向が見られた。2014年の東証における二度のティックサイズ変更は、特に東証-JNXにおいて東証の先行度合いを若干高めた。またJNX-ChiX間ではJNXの先行からChiXの先行へと平均で1ミリ秒未満だが先行運行関係が変化した。

パネル回帰分析により、先行運行関係を説明する要因として、ボラティリティ比やアスク側のデプス比が有意に正の係数、スプレッド率比が有意に負な係数を持つ等の結果が得られた。これらは、今回の回帰分析の結果は先に林(2015, 2016)で報告したものとは必ずしも整合的ではなかった。JNX-ChiXペアは他の市場ペアから、東証ティックサイズ変更フェーズII以降(データ期間3)は、それまでの期間から乖離した特徴を示した。

課題点として、データ前処理方法に対する分析結果の頑強性の問題が挙げられる。適切なデータ前処理方法や分析方法について検証・検討を重ねる必要がある。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。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、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の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、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。

先述の通り、図2-3において、みずほFG(8411)がフェーズ2導入以前にPTSがTSEに先行する状況が観察されたように、取引ルールその他のマイクロストラクチャ要因が、異市場・同一銘柄間の先行運行時間の決定に重要な役割を果たしているのは明らかである。Renò(2003)は、高頻度データの実証的特徴(stylized facts)の一つである“Epps効果”<sup>16)</sup>を説明する要因として、証券価格間の非同期性に加えて先行運行関係を挙げている。今後さらに売買の高速化、市場データの高頻度化、データ量の増大化が進む中、推定方法に改良を重ねることで証券間の先行運行時間がより正確に推定できるようになれば、高頻度データより計算される先行運行時間推定値の大きさやその時間変動が、高速化の進む今日の注文板市場のマイクロストラクチャ効果の評価に活用できるようになることも期待される。

### 注.

<sup>1)</sup> 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いと市場“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、今回の先行遅延度合い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) 本稿執筆時点(2017年1月)の市場規模として2017年1月10-13日(4営業日)の実績を見ると、週間売買代金は東証11兆433億円、JNX2570億円、ChiX1353億円、一方、取引件数は東証816万件、JNX150万件、ChiXは40万件である。地方三市場(名証、札証、福証)も含めた国内売買代金シェアでは、東証96.5%、JNX2.3%、ChiX1.2%である。(出所：<http://fragmentation.fidessa.com/jp/>)。
- 3) 東証作成リーフレット(2013年12月1日付)より。
- 4) <http://www.jpx.co.jp/markets/indices/line-up/index.html>。
- 5) 2013年10月末の入れ替えでは、富士重工(7270)が入り、ヤマダ電機(9831)が外れ、2014年10月末には、楽天(4755)、マツダ(7261)が入り、旭硝子(5201)、ニコン(7731)が外れた。
- 6) 鮮度劣化した気配値(stale quote)とは、最新の情報を反映していない、更新されずに注文板に残っている(発注されてから時間の経過した)古い気配値を指す。参考文献としては、例えば、Budish et al. (2015)を参照。
- 7) なお、セミマルチングールの枠組ではないが、非同期観測された2つの高頻度時系列データ間の相互相関係数を計測し、先行遅延関係を調べた先駆的研究としてde Jong and Nijman (1997)がある。
- 8) 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は今後検討の余地がある。
- 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。
- 10) ここでは、 $\mathbf{R}$  にならない、例えば被説明変数  $y$  の説明変数  $x$  と  $z$  に対する回帰モデルを  $y \sim x+z$  と表現する。また、 $x$  と  $y$  の主効果に加えて、 $x$  と  $z$  との交互作用項を変数に加えたモデルを  $y \sim x*z$  または  $y \sim x+z+x:z$  と表現する。ここで、 $x:z$  が交互作用項である。表記上は質的変数と量的変数は区別しない。さらに、今回使用した `lmer()` にならない、変数効果項  $r$  が加わった混合モデルは、 $y \sim x*z+(1|r)$  等と表現する。詳しくは、 $\mathbf{R}$  やパッケージ '`lme4`' のマニュアルを参照されたい。
- 11) もとより、東証の出来高シェアは日経平均採用銘柄を中心に概ね全体の9割を超えているともされていることから、 $TVag$  の値は東証の約定枚数に置き換えても以下の結果は本質的に変わらないと思われる。
- 12) 各変数の入力となる市場観測量として、3.4で述べたデータ加工処理後の計測量を使用した。
- 13) 例えば、 $Dir$  と  $T$  との交互作用項はモデルに加えた方が図5とは整合的であろう。今後の課題としたい。
- 14) 匿名レフリーからコメントを頂戴した。林(2016), p.148も参照せよ。
- 15) 匿名レフリーから指摘を頂戴した。
- 16) Epps (1979)により報告された、高頻度の証券リターン間の相関係数の絶対値の大きさが、リターン計測間隔を小さくするにつれ、低減していく現象。

## 謝 辞

本研究遂行にあたり、Dobrislav Dobrev 氏および Ernst Schaumburg 氏より、未刊行のワーキングペーパーを送付して頂いた。吉田朋広教授(東京大学)、小池祐太氏(首都大学東京)、Frédéric Abergel 教授(CentraleSupélec)、Marc Hoffmann 教授(Université Paris IV)、Jean Jacod 教授(同)、Mathieu Rosenbaum 教授(同)より貴重なコメントを頂いた。匿名の二名のレフェリーには原稿を丁寧に読んで頂き、有益なコメントを多数頂いた。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、日本学術振興会科学研究費(基盤研究(c)、課題番号:16K03601、研究代表者:林 高樹)、ならびに科学技術振興機構(JST)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)による研究成果の一部である。ここに謝意を表します。当然ながら、論文に含まれる誤りは全て筆者に帰属します。

## 参 考 文 献

- Bechler, K. and Ludkovski, M. (2014). Optimal execution with dynamic order flow imbalance, arXiv:1409.2618v2.
- Bellia, M., Subrahmanyam, M. G., Yuferova, D. and Uno, J. (2016). Low-latency trading and price discovery without trading: Evidence from the Tokyo Stock Exchange pre-opening period, available at SSRN: <http://dx.doi.org/10.2139/ssrn.2705962>.
- Budish, E., Cramton, P. and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics*, **130**(4), 1547–1621.
- Chen, Z., Lin, W. T., Ma, C. and Zheng, Z. (2013). The impact of individual investor trading on stock returns, *Emerging Markets Finance & Trade*, **49**, Supplement 3, 62–69.
- Chung, D. and Hrazdil, K. (2010). Liquidity and market efficiency: A large sample study, *Journal of Banking & Finance*, **34**, 2346–2357.
- de Jong, F. and Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets, *Journal of Empirical Finance*, **4**(2-3), 259–277.
- Dobrev, D. and Schaumburg, E. (2015). High-frequency cross-market trading: Model free measurement and applications, unpublished manuscript.
- Epps, T. W. (1979). Comovements in stock prices in the very short run, *Journal of the American Statistical Association*, **74**(366), 291–298.
- Gatheral, J. and Oomen, R. C. (2010). Zero-intelligence realized variance estimation, *Finance and Stochastics*, **14**, 249–283.
- 林高樹 (2015). 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*, **53**(4), 29–39.
- 林高樹 (2016). 国内高速 3 株式市場間の注文板形成の先行遅延関係分析, 『ファイナンスにおける数値計算法の新展開(ジャフィー・ジャーナル: 金融工学と市場計量分析)』(日本金融・証券計量・工学会 編), 128–155, 朝倉書店, 東京.
- Hayashi, T. and Koike, Y. (2016). Wavelet-based methods for high-frequency lead-lag analysis, arXiv:1612.01232v1.
- Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli*, **11**(2), 359–379.
- Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data, *Bernoulli*, **19**(2), 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*, **52**(6), 73–82.

- 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, **7**, 日本取引所グループ, 東京.
- Lipton, A., Pesavento, U. and Sotiropoulos, M. G. (2013). Trade arrival dynamics and quote imbalance in a limit order book, arXiv:1312.0514v1.
- 大崎貞和 (2014). グローバルな市場間競争と日本取引所グループ, 証券アナリストジャーナル, **52**(2), 6-15.
- 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.
- 太田亘 (2016). 取引システム高速化とティックサイズの制約, 現代ファイナンス, No.38, 27-59.
- Renò, R. (2003). A closer look at the Epps effect, *International Journal of Theoretical and Applied Finance*, **6**(1), 87-102.
- 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, **50**(9), 6-16.
- 宇野淳 (2014). 株式市場統合の現状と課題, 証券アナリストジャーナル, **52**(2), 16-23.

Statistical Analysis of High-frequency Limit-order Book Data:  
 On Cross-market, Single-asset Lead-lag Relationships  
 in the Japanese Stock Market

Takaki Hayashi<sup>1,2,3</sup>

<sup>1</sup>Graduate School of Business Administration, Keio University

<sup>2</sup>Graduate School of Social Sciences, Tokyo Metropolitan University

<sup>3</sup>JST-CREST

We are concerned with very short-term, lead-lag relationships between market prices of identical stocks traded concurrently on multiple trading venues in Japan, specifically the Tokyo Stock Exchange and two Proprietary Trading Systems, namely Japannext PTS and Chi-X Japan. In this paper, we conduct an empirical analysis with a modified version of the methodology recently proposed by Dobrev and Schaumburg (2015). This methodology focuses on the arrival times of the “events” for the paired point processes. That is, it utilizes (irregularly-spaced) timestamp records of the trading activities alone, and hence is not (directly) influenced by “microstructure noise” pertaining to the behavior of the observed, “inefficient” prices. As in our previous work (Hayashi, 2015, 2016) based on the methodology by Hoffmann et al. (2013), we empirically measure the magnitudes of the lead-lag times using high-frequency, limit-order book data for major Japanese stocks with milli-second time resolution obtained from the three venues.