

# 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 (要約版)

林 高樹

慶應義塾大学, 首都大学東京, CREST JST

2017年4月28日

## 概要

### 研究の目的

- 高頻度領域における証券価格間の先行遅行 (lead-lag) 時間の推定.
- 新手法の提案
- 高頻度データによる実証分析. 先行遅行関係の発生要因の理解, 知識発見

### 先行遅行時間推定の方法論

- Hoffman, Rosenbaum and Yoshida (2013)
- Dobrev and Schaumburg (2015)
- H. and Koike (2016)

### 実証分析

- 異市場同一銘柄間分析 (Cross-market, single-asset analysis)
  - 国内3市場 (東証, Japannext, Chi-X Japan)
  - 林 (2015, 2016, 2017)
- 先行遅行時間推定
- パネル回帰分析

## Hoffmann-Rosenbaum-Yoshida's method

Hoffmann, Rosenbaum and Yoshida(2013)

- Formulated a lead/lag estimation problem and proposed an estimator based on H. and Yoshida (2005)'s covariance estimator ("HY estimator")
  - no "synchronization" needed
- Two continuous semimartingales:  $X, Y$
- Shift  $J^i$  (interval for  $Y$ ) to the left (past) by  $\theta$ :  
 $J_{-\theta}^i = (T^{i-1} - \theta, T^i - \theta]$

- $\theta$ -shifted HY estimator:

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

where  $K_{\theta}^{i,j} = 1_{\{I^i \cap J_{-\theta}^j \neq \emptyset\}}$ :  $\theta$ -shifted "HY factor"

## HRY's method

Hoffmann, Rosenbaum and Yoshida(2013)

**Assumption:** There exists a latent process  $\tilde{Y}$  s.t.  $Y_t = \tilde{Y}_{t-\theta^*}$ .  $\theta^*$  is an unknown constant.

- $\tilde{Y}$  is not observed in real time, but is observed with (constant) delay  $\theta^*$ 
  - $\theta^* > 0$ :  $X$  leads  $Y$
  - $\theta^* < 0$ :  $Y$  leads  $X$

# HRY's method

Hoffmann, Rosenbaum and Yoshida(2013)

HRY(2013)'s proposal:

- “HRY estimator” for the unknown true lag parameter  $\theta^*$

$$\hat{\theta}_{HRY} := \hat{\theta}_n := \arg \max_{\theta \in \mathcal{G}^n} |U^n(\theta)_T| \quad (2)$$

where  $\mathcal{G}^n$  is a finite grid satisfying certain regularity conditions.

- In the high-frequency setting,  $\hat{\theta}_{HRY}$  is consistent for  $\theta^*$  (Theorem 1 of Hoffman, et al(13)).

## Estimated HRY measures: micro price, TSE vs JNX

time series plot of  $\hat{\theta}$

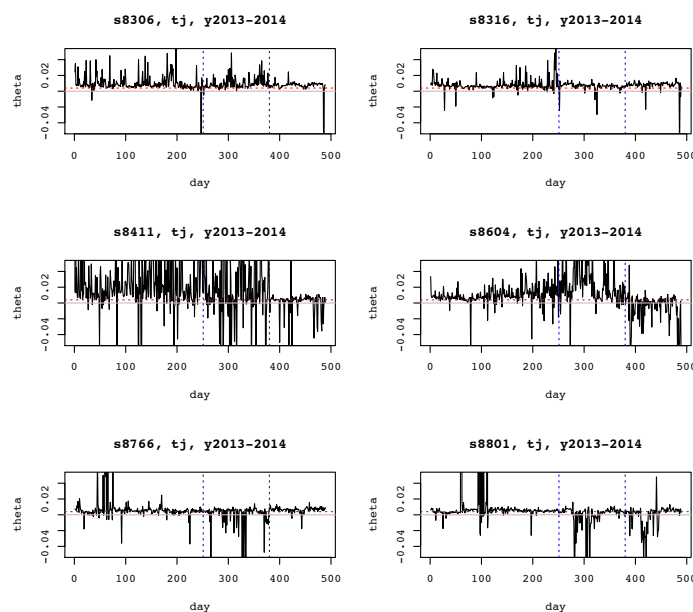


Figure 1 : Daily HRY measures, Mitsubishi UFJ FG(8306)–Mitsui Real Estate (8801): TSE–JNX.

Estimated HRV measures:

- Unstable
- Affected by microstructure noise
- Sensitive to jumps

An alternative approach by Dobrev and Schaumburg:

- “High-Frequency Cross-Market Trading: Model Free Measurement and Applications.” Working paper, 2016.
- Utilizes not price changes but timestamps of the trading activities
- Is not (directly) influenced by “microstructure noise” pertaining to the behavior of the observed, “inefficient” prices

## DS's method

- $I_t^j = 1$  if traded at  $t$ ; 0 otherwise.  $t = 0, 1, \dots, N$ ,  $j = 1, 2$ .
  - regular spacing with  $\Delta = \frac{T}{N}$ .
- Objective function to be maximized:

$$A(h) := \frac{1}{\underline{A}} \sum_{t=1}^{N-h} I_t^1 I_{t+h}^2$$

- where  $\underline{A} := \min(\sum_t I_t^1, \sum_t I_t^2)$ .
- Modified objective function:

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

- where  $\bar{A} := \frac{1}{2H+1} \sum_{h=-H}^H A(h)$ .

本研究における相互共起強度関数  $A(h)$  の計算:

- 約定系列に付随するタイムスタンプ・データではなく、マイクロプライスの変化する時点から成るタイムスタンプ・データを使用する.
- さらに、マイクロプライスの変化の正負に応じて、タイムスタンプ・データを、“買い方向,” “売り方向” に2分割して使用する.

## 相互共起関数

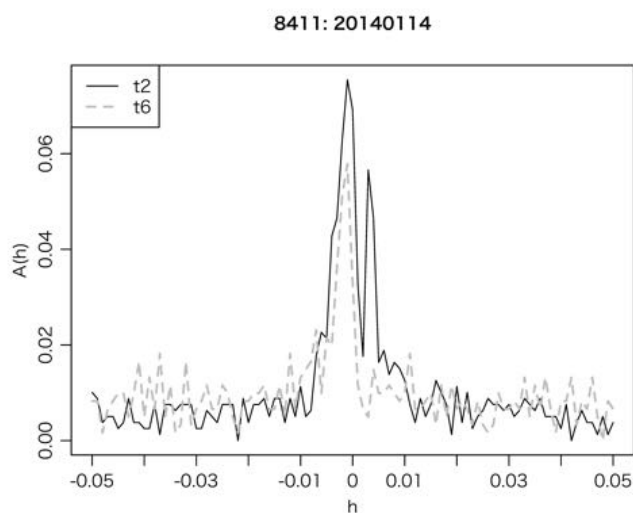


Figure 2 : 相互共起強度関数  $A(h)$  の計算例 (みずほ FG (8411), 2014 年 1 月 14 日). 第 2 時間帯 (実線) と第 6 時間帯 (破線).

## Estimated DS measures: micro price, TSE vs JNX

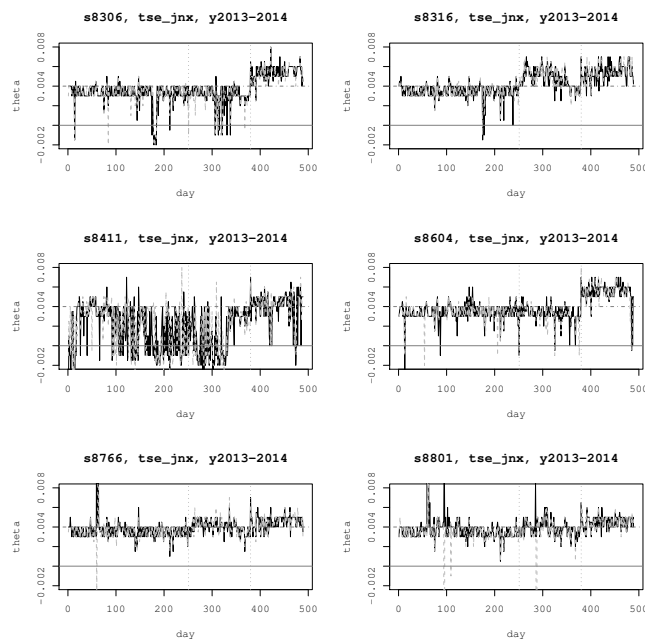


Figure 3 : DS 指標の日次推移 (実線は買い方向, 破線は売り方向): 三菱 UFJ フィナンシャルグループ (8306)–三井不動産 (8801): TSE–JNX.

## Summary statistics of estimated DS measures

summary over Period 7/22/14–12/30/14: Micro price, TSE vs JNX

code	Period 1 20130104–20140110		Period 2 20140114–20140718		Period 3 20140722–20141230	
	mean	stdev	mean	stdev	mean	stdev
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

Table 1 : DS 指標の要約統計量 (単位: 秒): TSE–JNX, 売買・時間帯区別せず.

## 観察結果の要約-1

(TSE-ChiX, JNX-ChX の結果については, 論文内の図表を参照.)

フェーズ 1 期間は結果が安定的でない. 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, おおよそ次のように要約可能:

- TSE-JNX, TSE-ChiX: TSE が PTS 二市場に対して, 約 4 ミリ秒先行.
  - フェーズ 2 以降, 特に TSE-JNX は, TSE の先行度増.
- JNX-ChiX: 先行遅行時間の差は 1 ミリ秒未満 (計測限界以下).
  - フェーズ 2 以降, JNX-ChiX は, JNX の先行から ChiX の先行へと変化.

## DS 指標の要約統計量 (売買方向別)

	TSE vs JNX		TSE vs ChiX		JNX vs ChiX	
	Buy	Sell	Buy	Sell	Buy	Sell
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

Table 2: DS 指標の要約統計量 (TOPIX100 構成銘柄全体の要約, 単位: ミリ秒).

## DS 指標 (銘柄・期間全体平均): 売買方向別・時間帯別

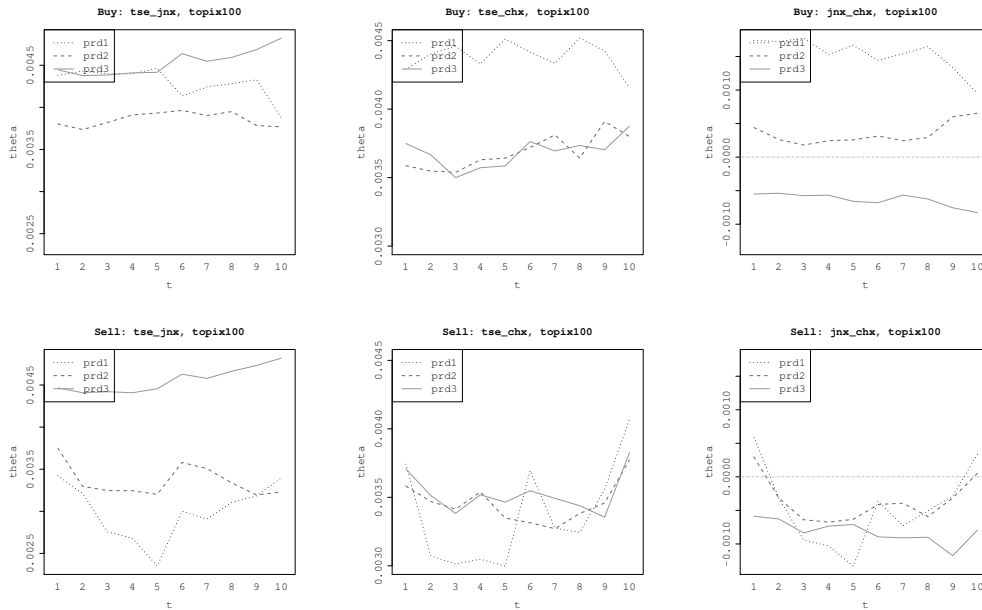


Figure 4: 個別銘柄 DS 指標の, 各時間帯 (横軸) におけるデータ期間内の銘柄全体平均値 (縦軸) (TOPIX100 銘柄): TSE-JNX (上段), TSE-ChiX (中段), JNX-ChiX (下段). 買い方向 (左列), 売り方向 (右列).

## 観察結果の要約-2

全般的な傾向:

- フェーズ 1 導入前, 売買方向間の非対称性大. 特に, 売り側は買い側に比し先行度がマイナス方向に位置.
- フェーズ 2 導入後, 先行度の売買方向間の差異は縮小.

さらに, TSE ティックサイズ変更により,

- TSE-JNX, TSE の先行度, 特に売り側で大幅増.
- TSE-ChiX, TSE の先行度, 特に買い側で減少.
- JNX-ChiX, 買い側は ChiX が遅行から先行に変化, 売り側は ChiX の先行度増.

1 日内推移:

- 売り方向に “U 字型”. フェーズ 1 導入前に顕著.
- 買い方向, フェーズ 1 導入前, 若干の “逆 U 字 (または逆 J 字) 型.”



## Panel regression analysis

- Want to explain how the lead/lag indices are generated in terms of observable characteristics of the stocks.
  - $i$ -th stock,  $j$ -th day,  $t$ -th time bin
- Covariates (group 1): Ratios between market pairs
  - $RVr_{ijt} = \log(RV_{ijt}^X/RV_{ijt}^Y)$ : Realized volatility ratio
  - $NQr_{ijt} = \log(NQ_{ijt}^X/NQ_{ijt}^Y)$ : # quote revisions ratio
  - $LTSr_{ijt} = \log(LTS_{ijt}^X/LTS_{ijt}^Y)$ : average lot size ratio
  - $Spr_{ijt} = \log(Spr_{ijt}^X/Spr_{ijt}^Y)$ : Spread ratio (to mid price) ratio
  - $QAr_{ijt} = \log(QA_{ijt}^X/QA_{ijt}^Y)$ : Depth (on Best Ask) ratio.
  - $QBr_{ijt} = \log(QB_{ijt}^X/QB_{ijt}^Y)$ : Depth (on Best Bid) ratio
- Covariates (group 2): Aggregate values of the three markets
  - $RVag_{ijt}$ : RV of the aggregate market
  - $RETts_{ijt}$ : Log return on the TSE
  - $TVag_{ijt}$ : Total share volume of the aggregate market
- All the variables are standardized.

## Panel regression analysis

- Fit a linear mixed effects model to the panel dataset.
- $y_{ijt} = \hat{\theta}_{ijt}$  :  $i$ -th stock,  $j$ -th day,  $t$ -th time bin
- $T$ : time-of-the-day effect ( $t = 1, \dots, 10$ ), 6-level factor
  - Split the trading hours (300 min.) into ten 30 min. intervals.
  - Bin 1: 9-9:30, Bin 2: 9:30-10:00, ..., Bin 10: 14:30-15:00
  - Use only Bins 2, 3, 4, 7, 8, 9.
- $Dir$ : Buy direction/ Sell direction, 2-level factor

$$y \sim \underbrace{(1|Code) + (1|Ymd)}_{\text{random effects}} + \underbrace{T}_{\text{6-level factor}} + \underbrace{Dir}_{\text{2-level factor}}$$

$$* \left( \underbrace{RVr + NQr + LTSr + SPRr + QBr + QAr}_{\text{covariates (group 1)}} \right)$$

$$\left. \underbrace{+RVag + RETts + TVag}_{\text{covariates (group 2)}} \right)$$

## Panel regression

- データセットを期間別に 3 分割し, 市場ペア毎に回帰.
- 回帰分析結果. 有意な係数の例:
  - 正: ボラティリティ比, アスク側デプス比, 等.
  - 負: スプレッド率比.
- 結果の詳細は, 本文を参照.
- 林 (2015, 2016) の結果とは必ずしも整合的でない.
- 推定結果の頑強性は依然として課題.

## Acknowledgement

- Grant-in-Aid for Scientific Research (C), Japan Society for the Promotion of Science (JSPS), no.16K03601.
- CREST JST Project, Graduate School of Mathematical Sciences, University of Tokyo.
- Japan Exchange Group, Inc.
- SBI Japannext Co., Ltd
- Chi-X Japan Ltd

## References

- 林 (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 (2016). “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.” ジャファイア・ジャーナル, 128–155.
- Dobrev and Schaumburg. (2015). “High-Frequency Cross-Market Trading: Model Free Measurement and Applications.” Unpublished manuscript.
- H. and Koike (2016). “Wavelet-based methods for high-frequency lead-lag analysis.” arXiv:1612.01232.
- Hoffmann, Rosenbaum, and Yoshida (2013). “Estimation of the lead-lag parameter from non-synchronous data,” *Bernoulli*, **19-2**, 426–461.