

Cluster 分析の手法を用いた Probability of informed trading の推計方法に関する提案

長山博一

一橋大学経済学研究科金融工学教育センター所属

2014 年 11 月 21 日

目次

- 1 研究背景と目的
- 2 先行研究
- 3 提案手法
- 4 シミュレーション結果
- 5 まとめ
- 6 参考文献
- 7 Appendix

研究背景と目的

- ・ Market Microstructure とは:
情報等が投資家行動へ及ぼす影響を分析する Finance の学問分野
- ・ この分野で投資家間の情報の非対称性度合が資産価格に影響を与えることが指摘されている
- ・ EKOP(1996) は情報の非対称性度合を定量的に計測する方法を提案した→ Probability of informed trading(PIN)
- ・ しかし PIN 推計には以下の2つの困難が伴う
 - ① パラメータの推計誤差が大きい
 - ② 時間がかかる (例. DY(2009) では推計に PC100 台を使用)

< 目的 >

PIN 推計に付随する数値計算上の諸問題を解決してパラメータの頑健かつ速い推計を可能にする→実証分析に生かす

Probability of Informed Trading の推定モデル

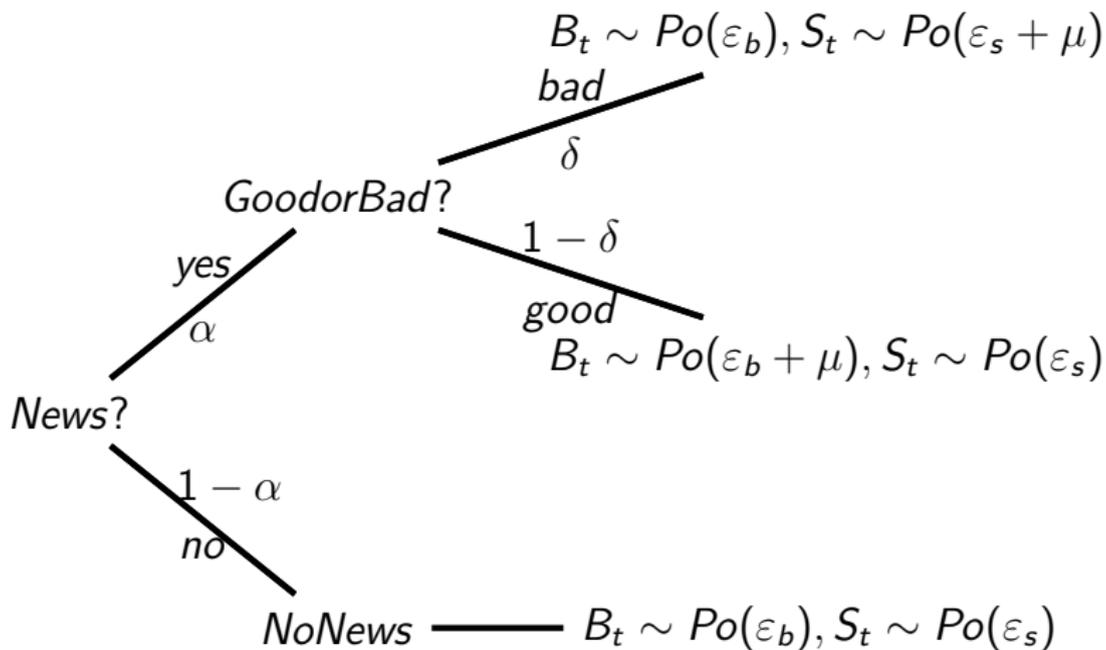
EHO(2002) は、EKOP(1996) が提案したモデルを少々拡張した PIN モデルを用いている。

EHO(2002) による PIN は以下の方法から導出される

- ・世の中の投資家を informed trader(情報優位の投資家)、uninformed trader(情報劣位の投資家) の 2 つに分類する
- ・取引数 (後述の B, S) は異なる 3 つのポアソン過程に従うと仮定する
- ・informed trader の取引数は、情報イベントが起こるかに依存する
- ・取引日ごとに 3 つに場合分けする
 - (1) 好材料の情報イベントが起こる
 - (2) 悪材料の情報イベントが起こる
 - (3) 情報イベントが起こらない

Probability of Informed Trading

Table: EHO(2002)PIN モデル



Probability of Informed Trading の推定モデル

- ・ α : ある取引日に情報イベントが起こる確率
- ・ δ : 情報イベントが悪材料である確率
- ・ μ : informed trader の情報に基づいた取引数の期待値
- ・ ε_b : uninformed trader の1日の買い主導約定件数の期待値
- ・ ε_s : uninformed trader の1日の売り主導約定件数の期待値
- ・ パラメータベクトル: $\eta \equiv (\alpha, \delta, \mu, \varepsilon_b, \varepsilon_s)'$
- ・ B_t, S_t : t 時点での買い主導約定件数、売り主導約定件数

パラメータの推計方法

過去の取引日数 T 日分のデータ ($\Omega \equiv ((B_1, S_1), \dots, (B_T, S_T))$) を所与とした時の対数尤度関数は、

$$\begin{aligned} \mathcal{L}(\eta|\Omega) = & \sum_{t=1}^T \ln \left[(1 - \alpha) e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} + \alpha \delta e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-(\mu + \varepsilon_s)} \frac{(\mu + \varepsilon_s)^{S_t}}{S_t!} \right. \\ & \left. + \alpha(1 - \delta) e^{-(\mu + \varepsilon_b)} \frac{(\mu + \varepsilon_b)^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} \right] \end{aligned}$$

- ・ PIN は下式で定義される出来高の期待値に占める私的情報に基づく出来高の期待値の比率となる。

$$PIN = \frac{\alpha \mu}{\alpha \mu + \varepsilon_b + \varepsilon_s}$$

- ・ 尤度関数を最大化するようなパラメータの値を求め、それを上の式に代入すれば PIN が求まる。

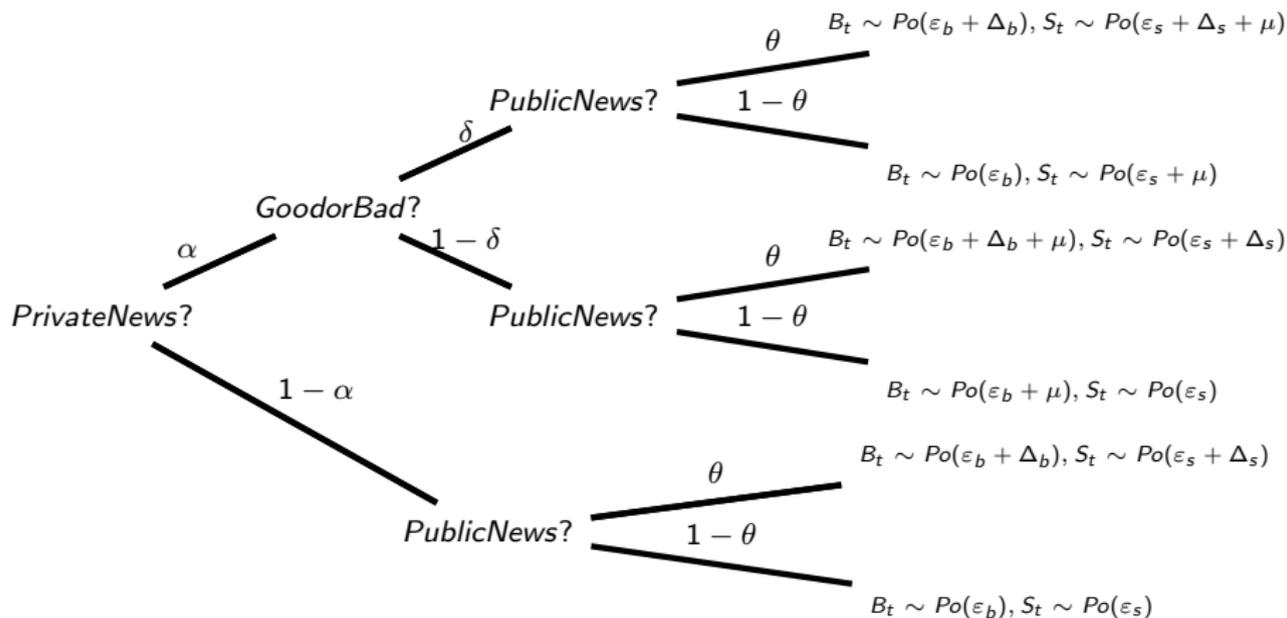
Probability of Informed trading(PIN) の計算方法

DY(2009) は、EHO(2002)PIN モデルの B と S が同時に増える現象を織り込んでいないという欠点を修正したモデルを提案した。

- ・ 私的情報イベントと公的情報イベントの区別
- ・ EHO(2002)PIN モデルの 3 つの場合について、symmetric order flow shock が起きたかどうかで場合分けをする
- ・ symmetric order flow shock(SOFS) とは、買い主導、売り主導の取引が双方とも増大するような公的情報が生起すること
- ・ Δ_b : SOFS に対応した 1 日の買い主導約定件数の期待値
- ・ Δ_s : SOFS に対応した 1 日の売り主導約定件数の期待値
- ・ θ : 公的情報イベントが起こる確率
- ・ パラメータベクトル: $\eta \equiv (\alpha, \delta, \mu_b, \mu_s, \varepsilon_b, \varepsilon_s, \Delta_b, \Delta_s, \theta)'$

Probability of Informed Trading の推定モデル

Table: DY(2009)APIN モデル



APIN の推計方法

対数尤度関数は、

$$\begin{aligned}
 \mathcal{L}(\eta|\Omega) = & \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta) e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} \right. \\
 & + (1-\alpha)\theta e^{-(\Delta_b+\varepsilon_b)} \frac{(\Delta_b+\varepsilon_b)^{B_t}}{B_t!} e^{-(\Delta_s+\varepsilon_s)} \frac{(\Delta_s+\varepsilon_s)^{S_t}}{S_t!} \\
 & + \alpha(1-\theta)(1-\delta) e^{-(\mu_b+\varepsilon_b)} \frac{(\mu_b+\varepsilon_b)^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} \\
 & + \alpha\theta(1-\delta) e^{-(\mu_b+\Delta_b+\varepsilon_b)} \frac{(\mu_b+\Delta_b+\varepsilon_b)^{B_t}}{B_t!} e^{-(\Delta_s+\varepsilon_s)} \frac{(\Delta_s+\varepsilon_s)^{S_t}}{S_t!} \\
 & + \alpha(1-\theta)\delta e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-(\mu_s+\varepsilon_s)} \frac{(\mu_s+\varepsilon_s)^{S_t}}{S_t!} \\
 & \left. + \alpha\theta\delta e^{-(\mu_b+\Delta_b+\varepsilon_b)} \frac{(\mu_b+\Delta_b+\varepsilon_b)^{B_t}}{B_t!} e^{-(\Delta_s+\varepsilon_s)} \frac{(\Delta_s+\varepsilon_s)^{S_t}}{S_t!} \right]
 \end{aligned}$$

下の制約条件付き最大化問題を解くことで η の推定値が得られる。

$$\underset{\eta \in \text{BFS}}{\text{Maximize}} \mathcal{L}(\eta | \Omega)$$

$$\text{BFS} \equiv \{\eta | \alpha, \delta, \theta \in [0, 1], \mu_b, \mu_s, \varepsilon_b, \varepsilon_s, \Delta_b, \Delta_s \in [0, \infty)\}$$

PIN のパラメータの最尤推定量は明示的な式で書くことができないので、最尤推定値を求めるには数値最適化手法を用いる。

・ AdjustedPIN (APIN) は以下で定義される

$$APIN = \frac{\alpha(\delta\mu_b + (1 - \delta)\mu_s)}{\alpha(\delta\mu_b + (1 - \delta)\mu_s) + \varepsilon_b + \varepsilon_s + \theta(\Delta_b + \Delta_s)}$$

・ Probability of a symmetric order flow (PSOS) は以下で定義され、流動性不足度合の代理変数であるとされている

$$PSOS = \frac{\theta(\Delta_b + \Delta_s)}{\alpha(\delta\mu_b + (1 - \delta)\mu_s) + \varepsilon_b + \varepsilon_s + \theta(\Delta_b + \Delta_s)}$$

PIN 推計上の問題点

EHO, DY どちらのモデルにしる、前頁のような数値最適化問題を解く際には、2つの問題が生じる。

1. 初期値選択の問題

初期値が真のパラメータから乖離していると、目的関数が最大値ではなく極大値となるようなパラメータを、最尤推定値としてしまう可能性が高まる。

2. Computational overflow, underflow の問題

例えば、Rでは、 e^{710} と e^{-746} より極端な値が計算中に現れると計算不能となる。

問題点の対策

1. 初期値選択問題の対策

- ① Yan and Zhang(2012) は初期値の組 125 組、DY(2009) は 10 組に対してそれぞれ数値最適化問題を解く手法を提案
- ② Gan et.al(2013) は Cluster 分析の手法を用いてパラメータを推定し、それを初期値として数値最適化問題を解く方法を提案

2. Computational overflow, underflow の問題

- ① EHO(2010), LK(2011) は overflow, underflow が起こりにくいような EHO(2002) モデルの対数尤度関数の分解方法を提案
- ② Jackson(2013) は取引数 (B,S) を定数倍で割り、推定値が求まった後にパラメータに同じ数を掛け直す方法を検証している

CPIN 推計方法

- (1) 注文不均衡 $X_t = B_t - S_t (t = 1, 2, \dots, T)$ を計算する
- (2) X_t をクラスタリングの手法を用いて、A. 値が高いクラスター、B. 値が低いクラスター、C. 中間のクラスター、の3つのクラスターに分ける
- (3) クラスターの分けられ方と $\{X_t\}_{t=1}^T, \{B_t\}_{t=1}^T, \{S_t\}_{t=1}^T$ からパラメータの推定値を計算する (例 α の推定値はサンプル T 個の内、クラスター A, B に入った個数で与えられる)
- (4) 最後に、PIN の推定値 CPIN は、

$$\widehat{CPIN} = \frac{\hat{\alpha}\hat{\mu}}{\hat{\alpha}\hat{\mu} + \hat{\varepsilon}_b + \hat{\varepsilon}_s}$$

で表される。

修正 CPIN

・ただし、Gan et.al(2013)によるCPINには(A)tieの存在(B)端点解に頑健でない、という問題がある。それを以下のように改善を図った。

(A) tieに対応したCPIN

tieが起こらないようなデータを構築するために、 X_t の代わりに出来高($V_t = Bt + St$ とおく)に対する注文不均衡 X_t の比率 $=X_t/V_t$ を計算した。

(B) 端点解に対応したCPIN

サンプルにはない極端な値をデータに含めることで、真のクラスター数が1or2個の場合にも3つのクラスターに適切に分けられる。

以降この修正を加えた推定をModified CPIN(MCPIN)と呼ぶことにする。

シミュレーションを用いてこの手法と先行研究の手法とを比較する。
EHO(2002) モデル:

- ① EHO-YZ-PIN: EHO(2010) の対数尤度関数分解を用いて YZ の方法で初期値 125 組に対して最適化する方法
- ② LK-YZ-PIN: LK(2011) の対数尤度関数分解を用いて YZ の方法で初期値 125 組に対して最適化する方法
- ③ LK-CPIN(ext): LK(2011) の対数尤度関数分解を用いて CPIN による初期値 1 組とその他 45 組に対して最適化する方法
- ④ LK-CPIN(MCPIN): LK(2011) の対数尤度関数分解を用いて CPIN(MCPIN) による初期値 1 組のみに対して最適化する方法

DY(2009) モデル

- ① DY-APIN: p31,32 の対数尤度関数分解を用い、DY(2009) を真似た 10 組の初期値に対し最適化する方法
- ② MCAPIN: MCPIN の考え方を APIN モデルに当てはめた方法
- ③ LK-MACPIN: p31,32 の対数尤度関数分解を用い、MCAPIN による 1 組の初期値に対してのみ最適化する方法

(a)EHO(2002)PIN モデルのパラメータ推定誤差

$\alpha = a, \delta = b, \mu = ck, \varepsilon_b = (1 - c)k/2, \varepsilon_s = (1 - c)k/2, a, b, c \sim U[0,1],$
 $k = 2,500$ と設定したもとで、 $\{B_t, S_t\}_{t=1}^{60}$ を含む 1000 組のデータの乱数を発生させ、MSE と MAE を基準に推定パラメータと真のパラメータの誤差を測った。

$k = 2,500$						
Mean Squared Error	PIN	α	δ	μ	ε_b	ε_s
EHO-YZ-PIN	0.15721	0.02504	0.05444	1,406,096.0	143,424.80	173,902.30
LK-YZ-PIN	0.00203	0.01127	0.02048	24,080.4	7,905.09	31.30
CPIN	0.00658	0.03755	0.05722	166,566.5	13,766.27	7,188.97
LK-CPIN	0.00408	0.01726	0.02443	23,987.9	13,142.89	205.71
LK-CPIN(ext)	0.00193	0.01064	0.01955	19309.1	7,258.68	158.79
MCPIN	0.00079	0.00908	0.01969	18,035.8	19.02	18.88
LK-MCPIN	0.00079	0.00783	0.01742	16,418.6	23.05	21.30
Mean Absolute Error						
EHO-YZ-PIN	0.28243	0.09457	0.15780	928.3	234.69	255.11
LK-YZ-PIN	0.01827	0.05212	0.08723	23.88	7.22	3.33
CPIN	0.03069	0.10022	0.13563	122.05	19.16	14.08
LK-CPIN	0.02225	0.06032	0.09257	26.12	10.37	3.91
LK-CPIN(ext)	0.01853	0.05161	0.08622	20.95	6.93	3.62
MCPIN	0.01744	0.05549	0.08829	29.15	2.82	2.91
LK-MCPIN	0.01671	0.04929	0.08387	19.89	3.17	3.11

(b)EHO(2002)PIN モデルのパラメータ推定誤差

$\alpha = a, \delta = b, \mu = ck, \varepsilon_b = (1 - c)k/2, \varepsilon_s = (1 - c)k/2, a=0.3, b=0.5, c=0.4$
 $k \sim U[50, 5000]$ と設定したもとの、 $\{B_t, S_t\}_{t=1}^{60}$ を含む 1000 組のデータの乱数を発生させ、MSE と MAE を基準に推定パラメータと真のパラメータの誤差を測った。

Mean Squared Error	PIN	α	δ	μ	ε_b	ε_s
EHO-YZ-PIN	0.00851	0.00727	0.02739	649,830.60000	14,917.42000	14,949.81000
LK-YZ-PIN	0.00082	0.00379	0.01499	107.79310	13.71013	14.51627
CPIN	0.00082	0.00402	0.01525	138.59510	12.38383	12.65753
LK-CPIN	0.00082	0.00380	0.01495	108.40790	13.62612	14.49546
LK-CPIN(ext)	0.00082	0.00379	0.01479	108.07870	13.60788	14.54331
MCPIN	0.00097	0.00456	0.01600	314.23690	12.45708	12.69598
LK-MCPIN	0.00082	0.00378	0.01516	108.67290	13.67153	14.64632
Mean Absolute Error						
EHO-YZ-PIN	0.08041	0.06567	0.13239	617.91760	91.66209	92.02741
LK-YZ-PIN	0.02294	0.04911	0.09681	7.98800	2.81792	2.88567
CPIN	0.02293	0.04952	0.09706	8.94978	2.66873	2.73077
LK-CPIN	0.02295	0.04915	0.09680	8.00943	2.80718	2.88200
LK-CPIN(ext)	0.02292	0.04917	0.09578	8.01533	2.80819	2.88703
MCPIN	0.02418	0.05197	0.09942	13.31811	2.68285	2.74266
LK-MCPIN	0.02296	0.04919	0.09718	8.04802	2.81470	2.89348

DY(2009)PIN モデルのパラメータ推定誤差

$\{B_t, S_t\}_{t=1}^{60}$ を含む 1000 個のデータセットの乱数を発生させ、MSE と MAE を基準に推定パラメータと真のパラメータの誤差を測っている。パラメータは

$$\alpha = a, \delta = b, \theta = d, \mu = ek, \varepsilon_b = ck, \varepsilon_s = ck, \Delta_b = fk, \Delta_s = fk,$$

と設定した。ただし $a, b, c, d, e, f \sim U[0,1]$ で、 $k = 2500$ である。

	MSE			MAE		
	DY-APIN	CAPIN	LK-CAPIN	DY-APIN	CAPIN	LK-CAPIN
APIN	0.00355	0.00062	0.00062	0.02418	0.01604	0.01586
PSOS	0.00708	0.00257	0.00123	0.03478	0.02630	0.02173
α	0.03139	0.00808	0.00645	0.08229	0.05466	0.04708
δ	0.04920	0.02093	0.02584	0.12147	0.08917	0.09392
θ	0.05021	0.02816	0.02495	0.12491	0.10526	0.09220
μ	93, 184.6	48, 164.9	48, 734.5	83.13	40.70	40.86
ε_b	12, 474.4	6, 798.7	1, 291.5	21.12	15.74	6.99
ε_s	24, 106.2	6, 081.9	1, 104.7	33.28	16.04	7.36
Δ_b	142, 174.2	118, 250.6	115, 630.8	121.75	105.21	100.46
Δ_s	142, 610.3	119, 519.1	115, 801.7	124.59	106.70	100.85

シミュレーション結果考察

- ① 全体的な取引量を固定して他のパラメータを動かした (a) の場合、どちらのモデルでも MCPIN を用いることで、パラメータの推計誤差を小さくすることができた。
- ② 確率等に関係するパラメータを一定とし取引量に関してのみ動かした (b) の場合、どちらのモデルも (特に EHO(2002) で明白)、サンプルにないデータを加えて推定を行っているため MCPIN は多少バイアスのかかった推定となっているが、それを用いて 1 回最適化すればよい推定となることがわかる。
- ③ シミュレーション結果のプロットを通じて、出来高が小さい時に CPIN による推定が比較的悪くなることがわかるが、1 回の最適化によって問題なく改善している。(Appendix 参照)

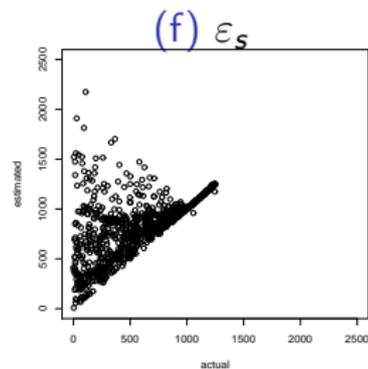
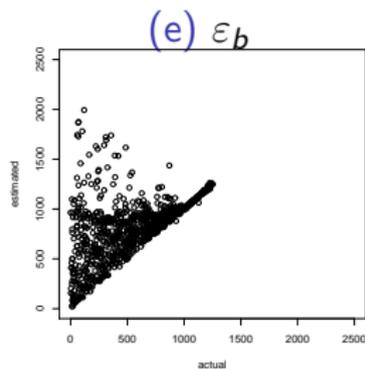
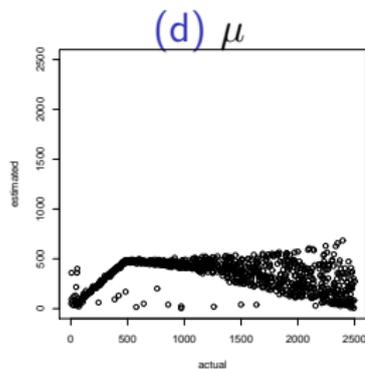
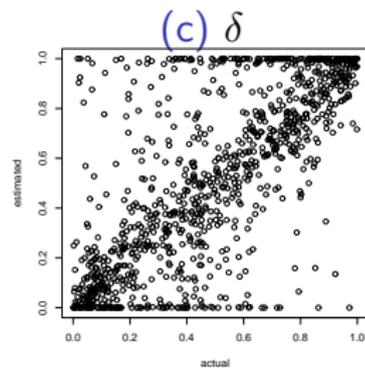
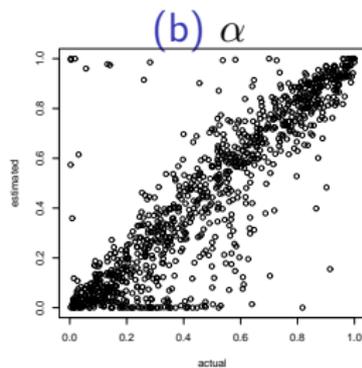
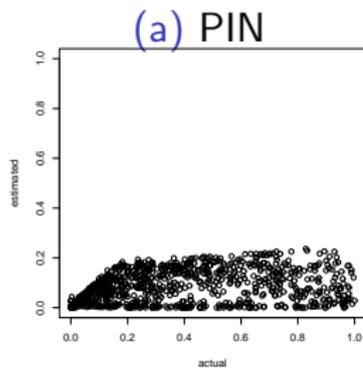
まとめ

- ① Cluster 分析を応用して様々な PIN モデルのパラメータ推計に関する初期値の問題を解決する framework を示すことができた。
- ② 数値最適化の回数 (ほとんどこれが実行時間に比例する) を $1/125$ や $1/10$ などに削減してもなお頑健でより良い推定となっている
- ③ 紙面の都合上示せなかったが、対数尤度関数の適切な分解 (p31,32 で新たに提案) と Jackson(2013) の方法も上手く取り入れることで近年の出来高が急増している環境にあっても computational overflow, underflow の問題を概ね解消した推定ができた
- ④ PIN や APIN といった値だけでなく他のパラメータ (Kubota and Takehara(2009): α , Brennan et.al.(2014): δ) を実証分析に利用したい場合に有用であると考えられる
- ⑤ 多くの PIN に関する文献では 1 年分のデータが推定に用いられているが、この手法を用いることでより、高頻度な推定 (四半期など) に対する推定が正当化されると思われる

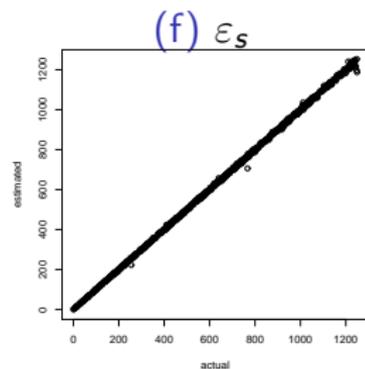
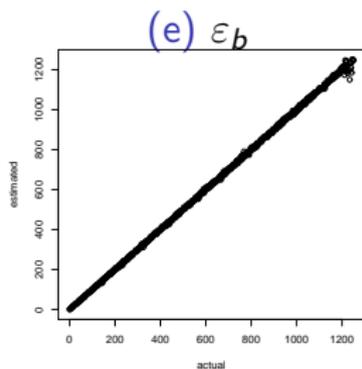
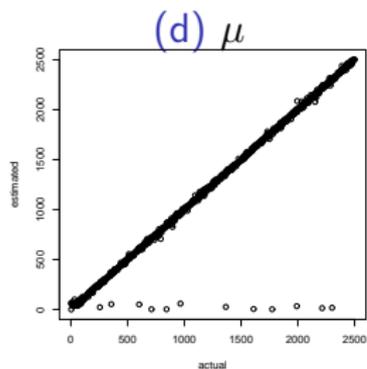
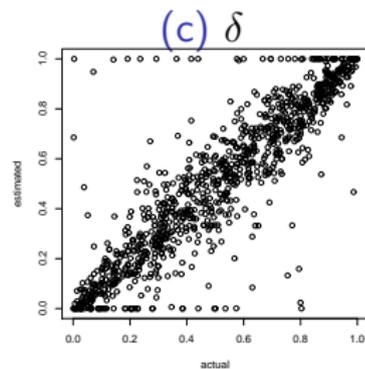
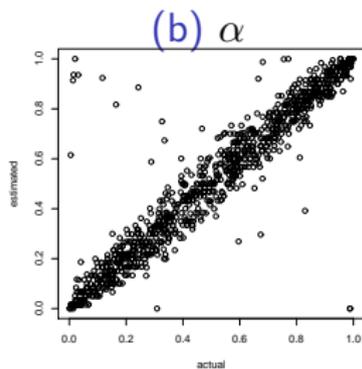
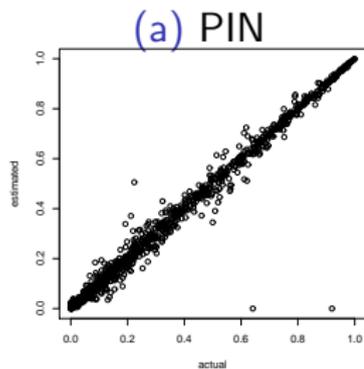
参考文献

- 1 Brennan, Michael J. and Huh, Sahn-Wook and Subrahmanyam, Avanidhar(2014) Asymmetric Effects of Informed Trading on the Cost of Equity Capital Available at SSRN:
- 2 David Jackson (2013) Estimating PIN for firms with high levels of trading Journal of Empirical Finance Volume 24, 116-120
- 3 Duarte, J., and L. Young (2009) Why is PIN priced? Journal of Financial Economics 91, 11 -138
- 4 Easley, D. N.M. Kiefer, M. O'Hara, and J.B. Paperman (1996) Liquidity, information, and infrequently traded stocks Journal of Finance, 51(4):1405-1436
- 5 Easley, D., Hvidkjaer, S. and M. O'Hara (2002) Is Information Risk a Determinant of Asset Returns?, Journal of Finance 10, 2185-2221
- 6 Easley, D., Hvidkjaer, S. and M. O'Hara, (2010) Factoring information into returns, Journal of Financial and Quantitative Analysis 45 (2), 293 - 309
- 7 Kubota, K. and H. Takehara (2009) Information based trade, the PIN variable, and portfolio style differences: evidence from Tokyo stock exchange firms, Pacific-Basin Finance Journal, 17, 319-337
- 8 Lin, W.W. and W.C. Ke (2011) A computing bias in estimating the probability of informed trading, Journal of Financial Markets 14, 625-640
- 9 Quan Gan David Johnstone Wang Chun Wei (2013) Cluster PIN: A new estimation method for the probability of informed trading, Discussion paper, The university of Sydney
- 10 Yan, Y. and S. Zhang (2012) An improved estimation method and empirical properties of the probability of informed trading, Journal of Banking and Finance 36, 454-467
- 11 音川和久 (2009) 投資家動向の実証分析-マーケットストラクチャーに基づく会計学研究, 中央経済社

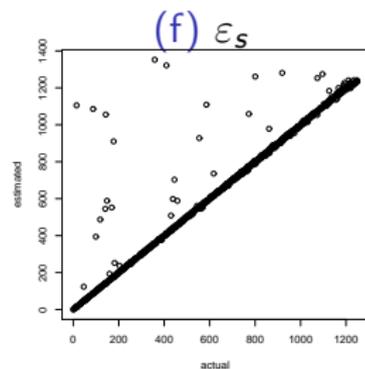
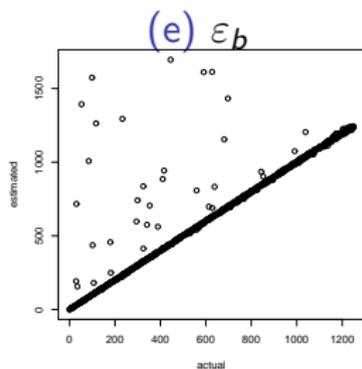
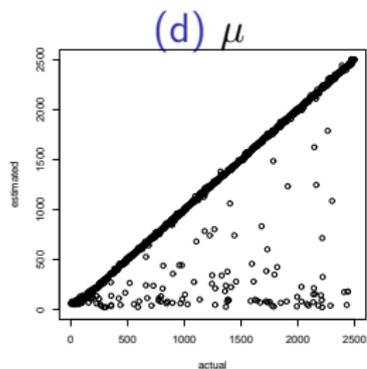
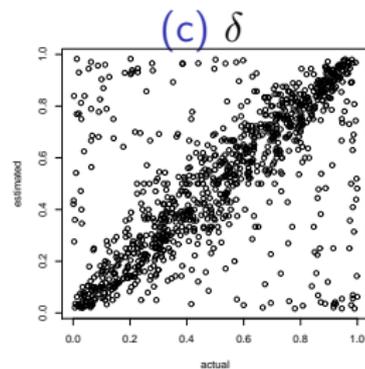
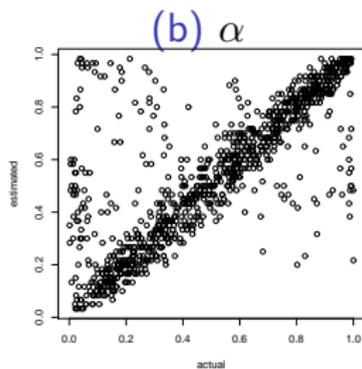
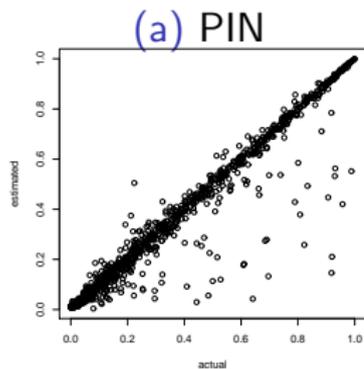
Appendix: EHO-YZ-PIN



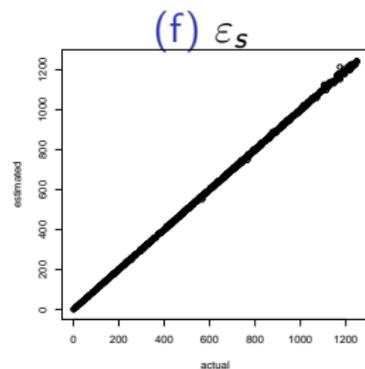
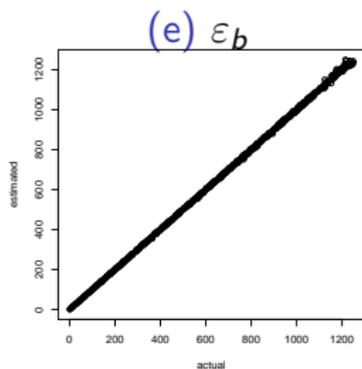
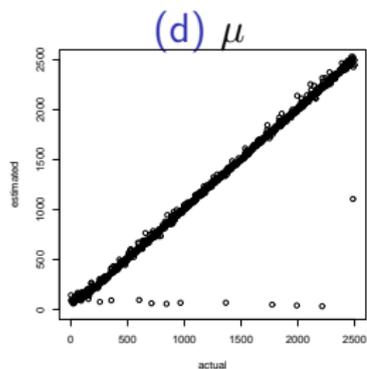
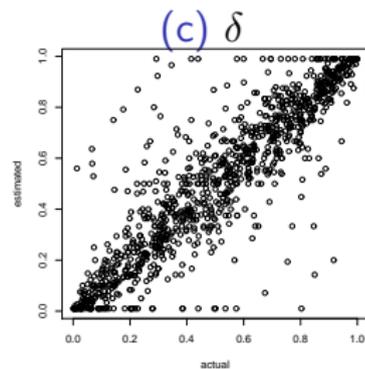
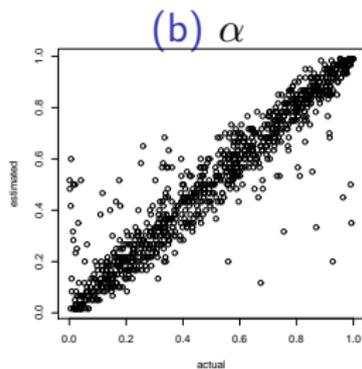
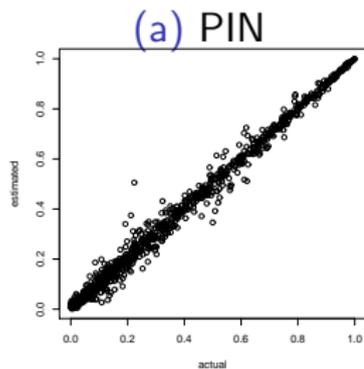
LK-YZ-PIN



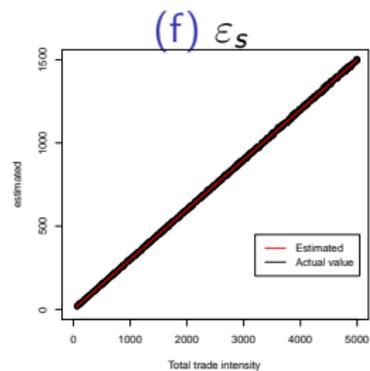
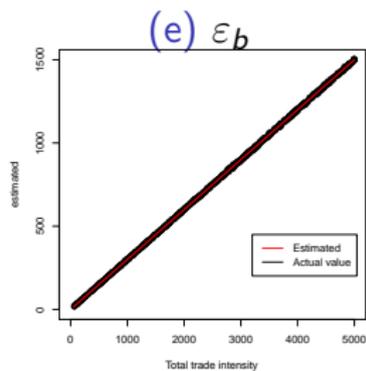
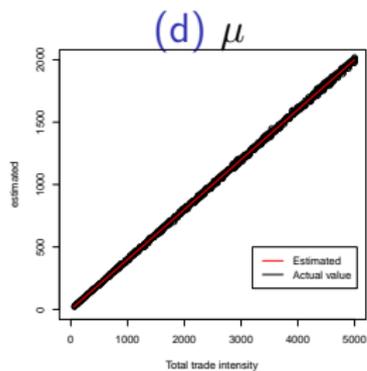
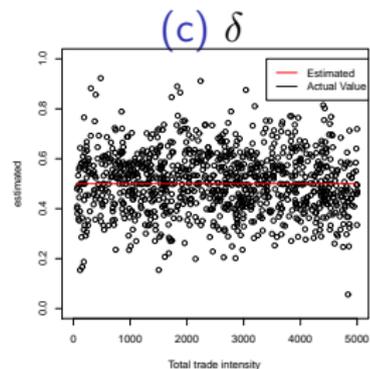
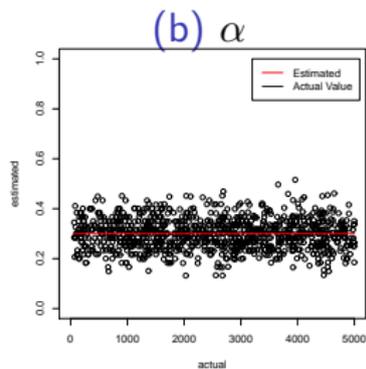
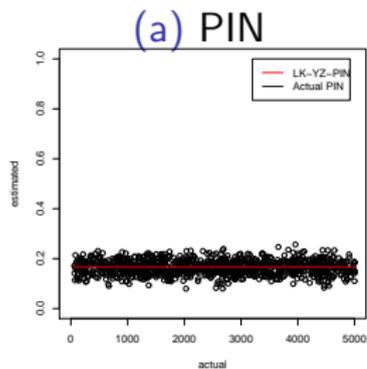
CPIN



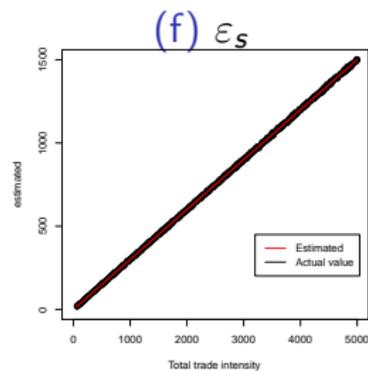
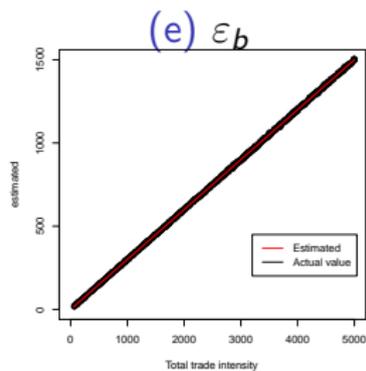
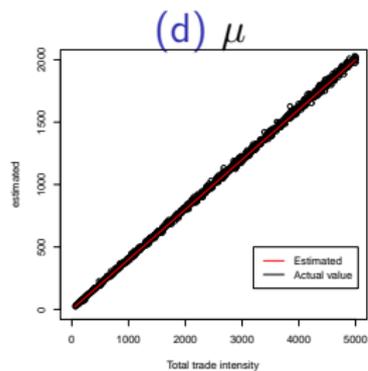
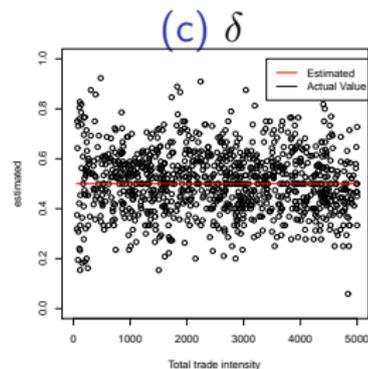
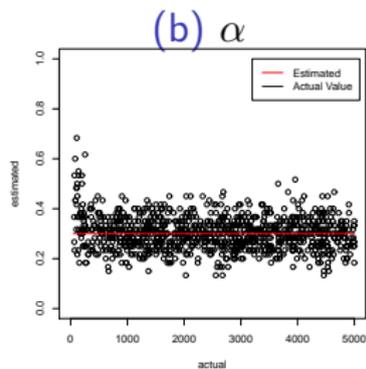
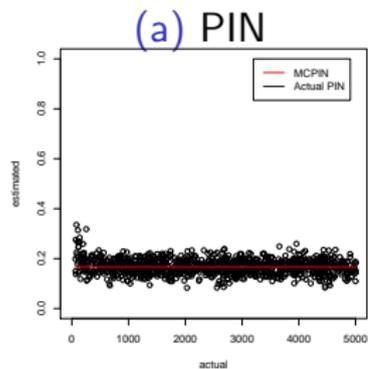
MCPIN



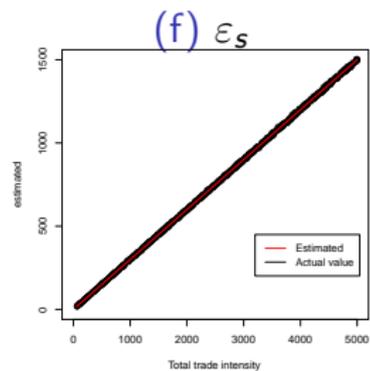
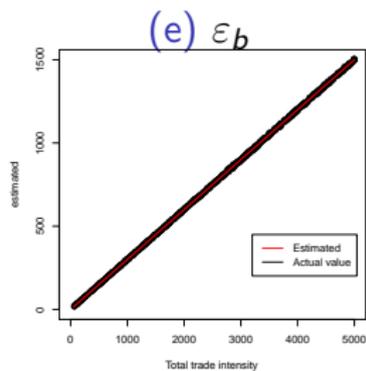
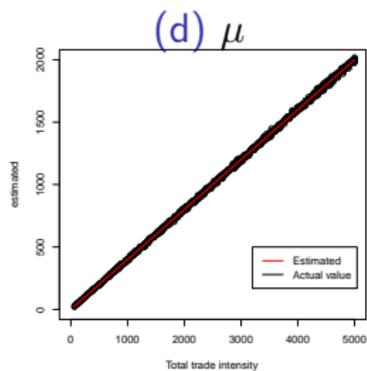
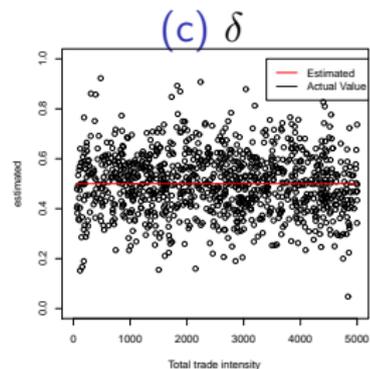
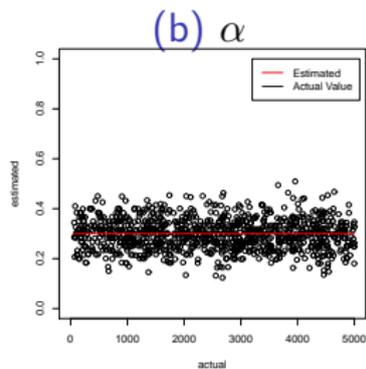
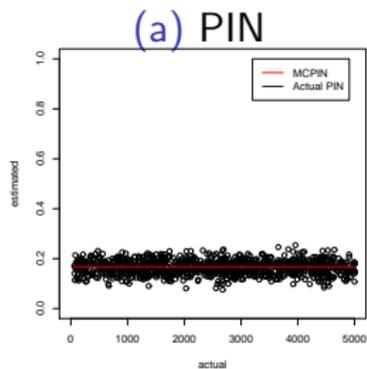
LK-YZ-PIN



MCPIN



MCPIN



Appendix: LK(2011) による対数尤度関数の分解

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\eta}|\boldsymbol{\Omega}) &= \sum_{t=1}^T \ln \left[(1-\alpha) e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} \right. \\
&\quad \left. + \alpha(1-\delta) e^{-(\mu+\varepsilon_b)} \frac{(\mu+\varepsilon_b)^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} + \alpha\delta e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-(\mu+\varepsilon_s)} \frac{(\mu+\varepsilon_s)^{S_t}}{S_t!} \right] \\
&= \sum_{t=1}^T \ln \left[(1-\alpha) \frac{\varepsilon_b^{B_t}}{(\varepsilon_b+\mu)^{B_t}} \frac{\varepsilon_s^{S_t}}{(\varepsilon_s+\mu)^{S_t}} + \alpha(1-\delta) e^{-\mu} \frac{\varepsilon_s^{S_t}}{(\varepsilon_s+\mu)^{S_t}} + \alpha\delta e^{-\mu} \frac{\varepsilon_b^{B_t}}{(\varepsilon_b+\mu)^{B_t}} \right] \\
&\quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\varepsilon_b + \varepsilon_s) - \ln(B_t! S_t!)] \\
&= \sum_{t=1}^T \ln \left[(1-\alpha) e^{-B_t \ln[(\varepsilon_b+\mu)/\varepsilon_b] - S_t \ln[(\varepsilon_s+\mu)/\varepsilon_s]} + \alpha(1-\delta) e^{-\mu - S_t \ln[(\varepsilon_s+\mu)/\varepsilon_s]} + \alpha\delta e^{-\mu - \ln[(\varepsilon_b+\mu)/\varepsilon_b]} \right] \\
&= \sum_{t=1}^T \ln \left[(1-\alpha) e^{k_{3,t}} + \alpha(1-\delta) e^{k_{2,t}} + \alpha\delta e^{k_{1,t}} \right] + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\varepsilon_b + \varepsilon_s) - \ln(B_t! S_t!)] \\
&= \sum_{t=1}^T \ln [\alpha\delta e^{(k_{1,t} - k_{max,t})} + \alpha(1-\delta) e^{(k_{2,t} - k_{max,t})} + (1-\alpha) e^{(k_{3,t} - k_{max,t})}] \\
&\quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\varepsilon_b + \varepsilon_s) + k_{max,t} - \ln(B_t! S_t!)]
\end{aligned}$$

Appendix: DY(2009) モデルの対数尤度関数分解

$$\begin{aligned}
\mathcal{L}(\eta|\Omega) &= \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta) e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} + \alpha(1-\delta) e^{-(\mu+\varepsilon_b)} \frac{(\mu+\varepsilon_b)^{B_t}}{B_t!} e^{-\varepsilon_s} \frac{\varepsilon_s^{S_t}}{S_t!} \right. \\
&\quad \left. + \alpha\delta e^{-\varepsilon_b} \frac{\varepsilon_b^{B_t}}{B_t!} e^{-(\mu+\varepsilon_s)} \frac{(\mu+\varepsilon_s)^{S_t}}{S_t!} + (1-\alpha)\theta e^{-(\varepsilon_b+\Delta_b)} \frac{(\varepsilon_b+\Delta_b)^{B_t}}{B_t!} e^{-(\varepsilon_s+\Delta_s)} \frac{(\varepsilon_s+\Delta_s)^{S_t}}{S_t!} \right] \\
&= \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta) \frac{\varepsilon_b^{B_t}}{(\varepsilon_b+\mu)^{B_t}} \frac{\varepsilon_s^{S_t}}{(\varepsilon_s+\mu)^{S_t}} + \alpha(1-\delta) e^{-\mu} \frac{\varepsilon_s^{S_t}}{(\varepsilon_s+\mu)^{S_t}} + \alpha\delta e^{-\mu} \frac{\varepsilon_b^{B_t}}{(\varepsilon_b+\mu)^{B_t}} \right. \\
&\quad \left. + (1-\alpha)\theta e^{-\Delta_b} \frac{(\varepsilon_b+\Delta_b)^{B_t}}{(\varepsilon_b+\mu)^{B_t}} e^{-\Delta_s} \frac{(\varepsilon_s+\Delta_s)^{S_t}}{(\varepsilon_s+\mu)^{S_t}} \right] + \sum_{t=1}^T [B_t \ln(\varepsilon_b+\mu) + S_t \ln(\varepsilon_s+\mu) - (\varepsilon_b+\varepsilon_s) - \ln(B_t!S_t!)] \\
&= \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta) e^{-B_t \ln[(\varepsilon_b+\mu)/\varepsilon_b] - S_t \ln[(\varepsilon_s+\mu)/\varepsilon_s]} + \alpha(1-\delta) e^{-\mu - S_t \ln[(\varepsilon_s+\mu)/\varepsilon_s]} \right. \\
&\quad \left. + \alpha\delta e^{-\mu - B_t \ln[(\varepsilon_b+\mu)/\varepsilon_b]} + (1-\alpha)\theta e^{-\Delta_b + B_t \ln(\frac{\varepsilon_b+\Delta_b}{\varepsilon_b+\mu}) - \Delta_s + S_t \ln(\frac{\varepsilon_s+\Delta_s}{\varepsilon_s+\mu})} \right] \\
&\quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b+\mu) + S_t \ln(\varepsilon_s+\mu) - (\varepsilon_b+\varepsilon_s) - \ln(B_t!S_t!)]
\end{aligned}$$

$$\begin{aligned}
& \dots = \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta)e^{\frac{\Delta_b + \Delta_s}{2} - B_t \ln[(\varepsilon_b + \mu)/\varepsilon_b] - S_t \ln[(\varepsilon_s + \mu)/\varepsilon_s]} + \alpha(1-\delta)e^{\frac{\Delta_b + \Delta_s}{2} - \mu - S_t \ln[(\varepsilon_s + \mu)/\varepsilon_s]} \right. \\
& \quad \left. + \alpha\delta e^{\frac{\Delta_b + \Delta_s}{2} - \mu - B_t \ln[(\varepsilon_b + \mu)/\varepsilon_b]} + (1-\alpha)\theta e^{-\frac{\Delta_b + \Delta_s}{2} + B_t \ln(\frac{\varepsilon_b + \Delta_b}{\varepsilon_b + \mu}) + S_t \ln(\frac{\varepsilon_s + \Delta_s}{\varepsilon_s + \mu})} \right] \\
& \quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\frac{\Delta_b + \Delta_s}{2} + \varepsilon_b + \varepsilon_s) - \ln(B_t! S_t!)] \\
& = \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta)e^{z_3, t} + \alpha(1-\delta)e^{z_2, t} + \alpha\delta e^{z_1, t} + (1-\alpha)\theta e^{z_4, t} \right] \\
& \quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\frac{\Delta_b + \Delta_s}{2} + \varepsilon_b + \varepsilon_s) - \ln(B_t! S_t!)] \\
& = \sum_{t=1}^T \ln \left[(1-\alpha)(1-\theta)e^{z_3, t - z_{max, t}} + \alpha(1-\delta)e^{z_2, t - z_{max, t}} + \alpha\delta e^{z_1, t - z_{max, t}} + (1-\alpha)\theta e^{z_4, t - z_{max, t}} \right] \\
& \quad + \sum_{t=1}^T [B_t \ln(\varepsilon_b + \mu) + S_t \ln(\varepsilon_s + \mu) - (\frac{\Delta_b + \Delta_s}{2} + \varepsilon_b + \varepsilon_s) + z_{max, t} - \ln(B_t! S_t!)] \\
& (z_{1, t} = \frac{\Delta_b + \Delta_s}{2} - \mu - B_t \ln(1 + \frac{\mu}{\varepsilon_b}), z_{2, t} = \frac{\Delta_b + \Delta_s}{2} - \mu - S_t \ln(1 + \frac{\mu}{\varepsilon_s}), \\
& z_{3, t} = \frac{\Delta_b + \Delta_s}{2} - B_t \ln(1 + \frac{\mu}{\varepsilon_b}) - S_t \ln(1 + \frac{\mu}{\varepsilon_s}), z_{4, t} = -\frac{\Delta_b + \Delta_s}{2} + B_t \ln(\frac{\varepsilon_b + \Delta_b}{\varepsilon_b + \mu}) + S_t \ln(\frac{\varepsilon_s + \Delta_s}{\varepsilon_s + \mu}) \\
& z_{max, t} = \max(z_{1, t}, z_{2, t}, z_{3, t}, z_{4, t}))
\end{aligned}$$

また、本稿の APIN の推定シミュレーションでは、DY(2009)に載っている中でもパラメータに制約を設けたモデルを用いているが、他のモデルでも同様な結果が示せた。

Appendix : 買い主導約定・売り主導約定

ここでは音川 (2009) を参考に「B」と「S」について簡単に解説する。

- ・東京証券取引所で株式注文方法には売買値段と数量を指定する「指値注文」と数量のみを指定する「成行注文」の2種類がある。

- ・また、株式の約定 (売買成立) は2つの種類に分類することが可能であり、それが買い主導約定 (B) と売り主導約定 (S) である。

- ・前者は、約定により株価が上昇した場合を指す。具体的には、指値売り注文に対し成行買い注文を出すことで約定した場合を言うことが大半である。買い手が即時の取引執行を望んだという意味合いが「買い手主導」という言葉に表れている。同様に、後者は約定により株価が下落した場合を指し、具体的には、指値買い注文に対して成行売り注文が出されることで約定した場合を言うことが多い。

Appendix : 実証分析への応用

表は音川 (2009) 掲載の、トヨタ自動車における記録の抜粋である。実証分析では、このような膨大な tick データから B,S の日次データを計算して整備する必要がある (Kubota and Takehara(2009) では B と S が正となる日が 120 日以上となるような銘柄のみをサンプルとしている)。私の場合はデータベース mysql でこれらのカウントを行っている。

データ日付	取引 所コ ード	証券 種別	場区 分	銘柄 コ ード	時刻	レコ ー ド種 別	管理番 号	価格	約定種 別	歩み値 フラグ
20000331	11	1	1	7203	1003	0	1	5370	16	0
20000331	11	1	1	7203	1003	33	2	5380	0	0
20000331	11	1	1	7203	1003	33	3	5370	128	0
20000331	11	1	1	7203	1003	0	4	5370	16	0
20000331	11	1	1	7203	1003	33	5	5380	0	0
20000331	11	1	1	7203	1003	33	6	5360	128	0
20000331	11	1	1	7203	1003	33	7	5380	0	0
20000331	11	1	1	7203	1003	33	8	5360	128	0
20000331	11	1	1	7203	1003	33	9	5380	0	0
20000331	11	1	1	7203	1003	33	10	5360	128	0
20000331	11	1	1	7203	1004	0	1	5360	16	0
20000331	11	1	1	7203	1004	33	2	5380	0	0
20000331	11	1	1	7203	1004	33	3	5360	128	0