

JPX ワーキング・ペーパー特別レポート

## 低流動性銘柄の市場クオリティ計測に関する研究

On the evaluation of intraday market quality in the limit-order book markets: A collaborative filtering approach

---

林 高樹 (慶應義塾大学大学院経営管理研究科)

高橋 慎 (法政大学経営学部)

2021 年 6 月

本研究の詳細については, *Japanese Journal of Statistics and Data Science* 第 4 巻第 1 号に掲載.

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・関連会社の役職員及び外部研究者による調査・研究の成果をと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である個人に属し、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。

- ・ JSPS 科学研究費助成金 (課題番号 JP16K03601)
- ・ 文部科学省科学研究費補助金 (課題番号 17H01100)
- ・ 法政大学イノベーション・マネジメント研究センター研究プロジェクト「金融市場における情報伝播とその周辺に関する統計分析」
- ・ 日本取引所グループ東京証券取引所

1. 研究の概要
2. レコメンダ・システム
3. 方法論
  - 市場クオリティ指標の推定問題
  - モデル
4. 実証分析
  - データ
  - 推定と予測の手順
  - 結果
5. まとめ

## 研究の概要

---

# 本研究の目的と方法論

目的: 個別銘柄の高頻度 (短時間) の市場クオリティを評価

- ・ (真の) 流動性などの市場クオリティの評価は容易ではない
- ・ 高頻度注文板データにより評価
  - 流動性の度合は銘柄毎に異なる
  - 銘柄によっては欠損値が頻繁に発生
  - 市況や時間帯により欠損値の発生度合いも変化

方法論: ハイブリッド型 Collaborative Filtering (CF) の採用

- ・ Recommender System で用いられる手法のひとつ
- ・ 欠損値を埋めるために, 補助的な情報を活用
  - 個別銘柄に関する種々の属性情報
  - 株式市場全体のマクロ情報

## FLEX Full データを用いた実証分析

- ・ TOPIX100 銘柄, Mid400 銘柄のうちの 188 銘柄
- ・ データ期間: 2019 年 1 月 4 日～3 月 29 日 (58 営業日), 5 分間隔
- ・ 市場クオリティを特徴づける要因の特定
- ・ 提案手法は欠損率によらず安定した予測パフォーマンス

## 今後の課題

- ・ データ期間やカバレッジを広げての分析
- ・ 予測精度の向上

## レコメンダ・システム

---



## レコメンダ・システム (Recommender System)

“特定のユーザーが最も関心を持っている可能性が高いアイテムの提案を提供するソフトウェアおよびツール. ... それらの提案は, どのアイテムを購入するか, どの音楽を聴くか, どのオンラインニュースを読むかなど, さまざまな意思決定プロセスに関連している”

(Ricci et al., 2015)

- ・ “主として, 例えば, ウェブサイトが提供する可能性のある圧倒的な数のアイテムの選択肢を評価するのに, 十分な個人的経験や能力を欠く個人を対象とする” (同)
- ・ 例: Amazon.com, YouTube, Netflix, Spotify, LinkedIn, Facebook
- ・ システムのアウトプット
  - ユーザーの評価予測値 - ユーザー × アイテム 「行列補完問題」
  - 推薦リスト - 「トップ- $N$  推薦問題」

## 学術研究

- ・ 90年代半ば頃から始まり, 今世紀に入り実用の広がりと共に大きく進展
- ・ ACM Digital Library (<http://dl.acm.org/>) 内で  
“Recommender System” がタイトルに含まれる論文の件数:
  - 1999-2003: 26
  - 2004-2008: 221
  - 2008-2013: 943
  - 2013-2019: 1381

参考: Ricci, Rokach, Shapira (2015), *Recommender Systems Handbook*, Ch.1.

## 協調フィルタリング (Collaborative Filtering, CF)

アイテムやユーザーに関する外生的情報を使わずに、評価や利用履歴のパターンに基づいて、個別ユーザーに対してアイテムの推薦を行う方法

- ・ ユーザー同士が (気付くことなく) 互いに興味のあるアイテムを見つけ出すのを協力し合う
- ・ ネットフリックス賞 (2006.10 - 2009.9) が手法の発展を加速
  - ・ 1億件の映画評価データ (ユーザー 480,189 人, 映画 17,770 本)
  - ・ 186 カ国 41,305 チーム 51,051 人が参加
  - ・ 5,169 チーム 44,014 人がサブミット
- ・ CF システムは, ユーザー同士, アイテム同士, あるいはユーザー・アイテム間の近さ, 類似性, 相関性の情報を活用
  - 無相関なデータセットでは, 役に立たない
  - 主要な 2 手法: 近傍ベスト法と潜在因子モデル

### データのスパース性 (Sparsity)

- ・ 典型的に, 実用上の評価行列は極めてスパース  
⇒ 相対的にごく少数の評価から推薦を行う必要

### コールドスタート問題 (Cold-start problem)

- ・ 評価を付けていない新ユーザーに対してどのように推薦を行うか
- ・ 評価の付けられていない or 購入されたことのないアイテムをどのように扱うか
  - ハイブリッド法の採用 (追加的, 外生的情報の利用) or
  - ユーザーにサービス利用開始前に最小限の評価をしてもらう

# 本研究のアプローチ

回帰ベース潜在因子モデル (Agarwal & Chen, 2009)

Regression-based Latent Factor Model (RLFM)

- ・ 単一のモデリングの枠組の中で, ユーザーやアイテムの特徴量やユーザーのアイテムに対する過去の反応の情報を活用
- ・ ハイブリッド型: CF とコンテンツ・ベースの長所を取り込む
  - ・ 過去の反応データがユーザーやアイテムの潜在因子を決定するのに不十分な時, 特徴量に基づいた回帰モデルによってそれらを推定
  - ・ さもなくば, 行列分解によって推定
- ・ コールド・スタートからウォーム・スタートへ切れ目なく対応
- ・ 過去の反応データが僅少なユーザーに対して, 回帰関数を介してユーザーの特徴量に依存する値に縮小 (shrink) させる
  - ・ 僅少なアイテムについても同様

参考: Agarwal & Chen (2016), Ch.8

## 方法論

---

# 市場クオリティ

## 市場クオリティ

- ・ 市場が“適正に”機能する程度 (cf. Yano, 2009)

## 市場クオリティを左右 (or 評価) する主要要素 (or 計測量)

- ・ 取引コスト (ビッド・アスク・スプレッド等)
- ・ ボラティリティ
- ・ 充填率 or 約定比率 (約定数量 ÷ 注文数量)
- ・ 執行速度
- ・ 流動性

## 高頻度トレーダーの市場クオリティへの影響評価は重要なテーマ

- ・ Menkveld (2013), Brogaard et al. (2014), Kirilenko et al. (2017), etc.

## 流動性とは

- ・ 売買したい数量を, 時間をかけずに, 市場実勢価格に変化を与えず取引できるか否かの程度
- ・ マーケット・インパクト (自分の売買注文が売り値の下落や買い値の上昇をもたらすこと) の大小

## 主な流動性指標

- ・ 板の深さ or デプス (注文板状に存在する指値注文数量)
- ・ 出来高 (約定枚数・金額)
- ・ 売買回転率 (約定枚数 ÷ 発行済株式数)
- ・ Kyle (1985) の  $\lambda$  (価格インパクト指標)
- ・ Amihud (2002) の ILLIQ (非流動性指標)



# Inverse Limit Order Book Slope (ILOBS)

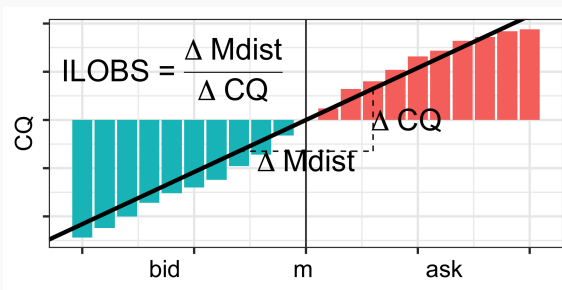
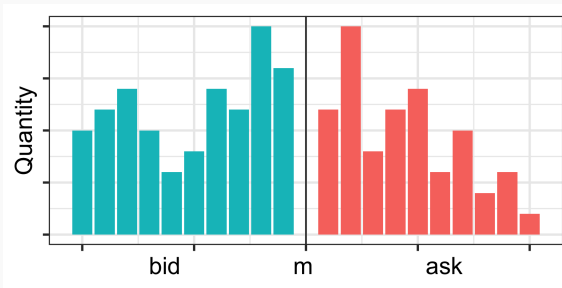
Deuskar & Johnson (2011) が提案した非流動性指標

- ・ 買い気配と売り気配の累積枚数の, 中値を中心とする気配価格の変化に対する傾き, の逆数

$$ILOBS = \frac{\sum_{k=1}^K (P_k^b - m)^2 + \sum_{k=1}^K (P_k^a - m)^2}{\sum_{k=1}^K (P_k^b - m) (-CQ_k^b) + \sum_{k=1}^K (P_k^a - m) CQ_k^a}$$

- $P_k^{a,b}$ : 最も内側 (最良気配) から  $k$  番目の板の気配価格
- $a$ : ask,  $b$ : bid
- $m$ : 中値 (最良買い気配と最良売り気配の平均)
- $CQ_k^{a,b}$ : 最も内側 (最良気配) から  $k$  番目の板までの累積枚数

# ILOBS の概念図



Deuskar & Johnson (2011) は, ILOBS が価格インパクト (注文に伴う価格変化) とほぼ完全な対応関係を持つことを示した

- ・ ILOBS は, 価格インパクトを予測するのに有効
- ・ 市場の (非) 流動性の計測に有効な指標

## 価格インパクトに関する実証分析

- ・ Bouchaud et al. (2009)
- ・ Eisler et al. (2012)
- ・ Hautsch & Huang (2012)
- ・ Cont et al. (2014)
- ・ 高橋 (2018)

# ボラティリティ

ボラティリティとは

- ・ 資産価格の変動性を表す観測できない状態変数
- ・ 金融リスク管理や資産の価格付けで重要

高頻度データを用いたボラティリティの推定

- ・ 価格変化率 (の二乗) から計算される Realized Volatility (RV)
  - Andersen et al. (2001), Barndorff-Nielsen & Shephard (2002)
- ・ 価格のレンジ (最高値と最安値との差) から計算される推定量
  - Parkinson (1980), Garman & Klass (1980), Rogers & Satchell (1991)

本研究では, tick-by-tick の中値から計算した RV を採用

- ・ 他のボラティリティ推定量にも容易に応用できる

# Regression-based Latent Factor Model (RLFM)

First layer (observation equation):

$$y_{i,j} = x'_{i,j}b + \alpha_i + \beta_j + u'_i v_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

Second layer (state equation):

$$\alpha_i = g'_0 w_i + \epsilon_i^\alpha, \quad \epsilon_i^\alpha \sim N(0, \sigma_\alpha^2)$$

$$\beta_j = d'_0 z_j + \epsilon_j^\beta, \quad \epsilon_j^\beta \sim N(0, \sigma_\beta^2)$$

$$u_i = G w_i + \epsilon_i^u, \quad \epsilon_i^u \sim N(0, \sigma_u^2)$$

$$v_j = D z_j + \epsilon_j^v, \quad \epsilon_j^v \sim N(0, \sigma_v^2)$$

- $\alpha_i$ : 個別銘柄の潜在因子
- $\beta_j$ : 時間帯の潜在因子
- $u'_i v_j$ : 銘柄-時間帯の潜在因子 ( $u_i, v_j$  は  $r$  次元)

レコメンダ・システムが想定しているのはユーザー × アイテム行列

- ・ 1時点または1期間のクロス・セクション・データ
- ・ トレンドや周期性を考慮する必要はない

本研究で扱うのは高頻度の時系列データ

- ・ 変数 (取引枚数やボラティリティなど) には日内周期性がある
  - Wood et al. (1985), McNish & Wood (1992), Lee et al. (1993)
  - Andersen & Bollerslev (1997), Dacorogna et al. (2001)
- ・ 予測には問題ないが、パラメータ推定では多重共線性の問題
- ・ 慎重な変数選択やデータの前処理で多重共線性を回避

# 実証分析

---

## FLEX Full

- ・ データ期間: 2019 年 1 月 4 日～2019 年 3 月 29 日 (58 営業日)
  - 推定期間: 1 月 4 日～2 月 28 日 (38 日)
  - 予測期間: 3 月 1 日～3 月 29 日 (20 日)
- ・ 対象銘柄: 東証一部 188 銘柄
  - TOPIX100 内の 98 銘柄, Mid400 内の 90 銘柄
- ・ 各営業日を 5 分間隔に分割, 計 60 時間帯
  - 引け (11:30, 15:00) のレコードは除去

## 日経 NEEDS-FQ

- ・ 属性情報: 日経 225 採用銘柄, 時価総額



## 次元

- ・ 銘柄:  $i = 1, \dots, M = 188$
- ・ 時間帯:  $j = 1, \dots, N = 60$

## 目的変数 $y_{i,j}$

- ・ 銘柄-時間帯 ( $M \times N$ ) 行列を構成
  - (a) 両サイド 10 本値から計算した ILOBS (非流動性指標) の対数値
  - (b) tick-by-tick の中値から計算した RV の対数値

## 銘柄-時間帯についての説明変数 $x_{i,j}$ (定数項を入れて $s = 7$ 次元)

- ・ 対数収益率  $\ln \text{Ret.D}$
- ・ 約定金額  $\log \text{Damt}$
- ・ スプレッド中値比  $\text{Spr}$
- ・ 約定件数の対数値  $\log nD$
- ・ 約定不均衡  $\text{ImbD}$
- ・ 気配更新件数の対数値  $\log nQ$

個別銘柄についての説明変数  $w_i$  ( $p = 18$  次元)

- ・ 日経 225 採用ダミー
- ・ TOPIX17 業種ダミー (17 番目省略)
  1. 食品, 2. エネルギー, 3. 建設・資材, 4. 素材・化学, 5. 医薬品,
  6. 自動車・輸送機, 7. 鋼鉄・非鉄, 8. 機械, 9. 電機・精密,
  10. 情報通信・サービスその他, 11. 電気・ガス, 12. 運輸・物流,
  13. 商社・卸売, 14. 小売, 15. 銀行, 16. 金融 (銀行を除く),
  17. 不動産
- ・ 前日終値に基づく時価総額 (兆円) の対数值 (MCAP)

時間帯についての説明変数  $z_j$  ( $q = 6$  次元)

- ・ 基準化した時間帯  $j/N$  についての 6 次までの多項式

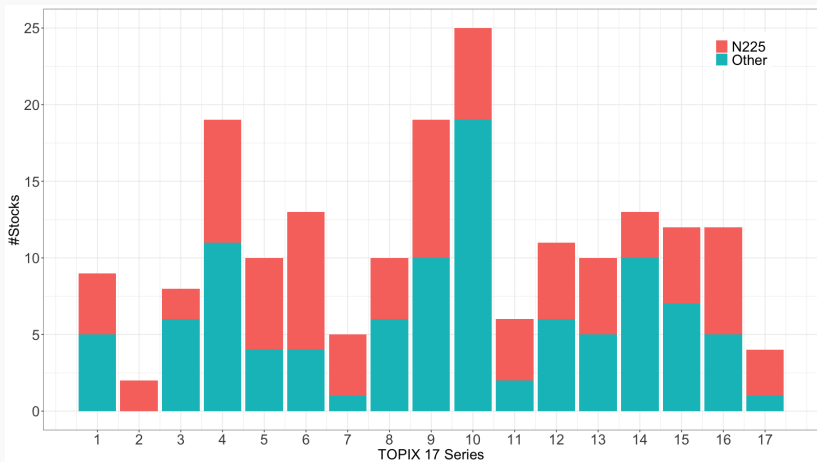
$$\left(\frac{j}{N}\right)^k, \quad k = 1, 2, \dots, 6$$

# 要約統計量

		平均	標準偏差	最小値	中央値	最大値
$y_{ij}$	logILOBS	1.4713	1.7946	-6.3261	1.6740	5.6373
	logRV	-2.0052	0.3565	-2.7264	-2.0717	-0.7483
$x_{ij}$	logRet.D	0.0007	0.0033	-0.0087	0.0006	0.0139
	lognD	3.6260	2.6912	0.7955	2.6905	16.7226
	logDamt	7.4288	0.6147	6.1088	7.5542	8.8933
	lmbD	0.2610	2.8760	-7.5759	0.0749	13.5832
	Spr	8.2578	5.7738	1.7525	6.0097	33.4802
	lognQ	2.7772	0.4188	1.7550	2.8651	3.6946
$w_i$	logMCAP	-0.0602	1.2181	-2.3862	0.1018	3.0944

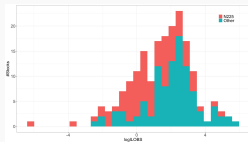
各銘柄の全期間の平均値 188 個から計算した統計量

# TOPIX17 業種別の日経 225 採用銘柄の分布

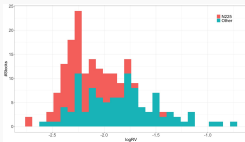


# 各銘柄の全期間の平均値の分布

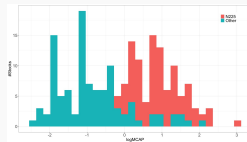
## logLOBS



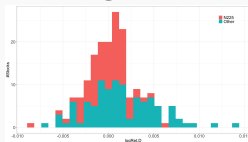
## logRV



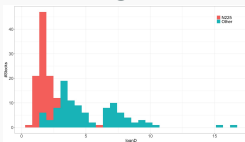
## logMCAP



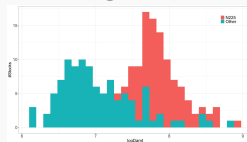
## logRet.D



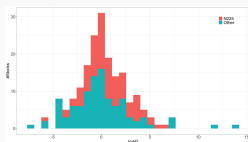
## lognD



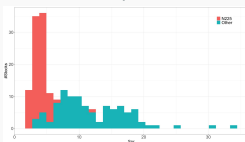
## logDamt



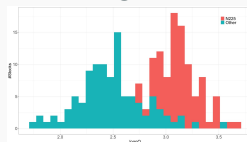
## lmbD



## Spr

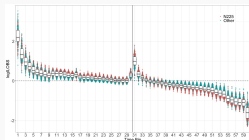


## lognQ

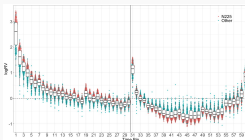


# 日内周期性

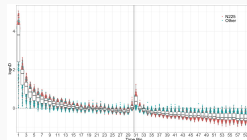
logLOBS



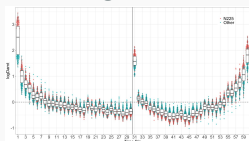
logRV



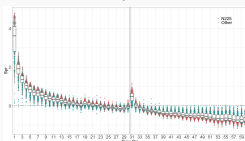
lognD



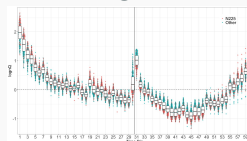
logDamt



Spr



lognQ

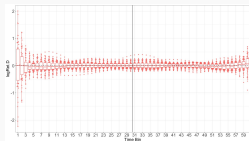


各日の推定の前に以下の手順を実行

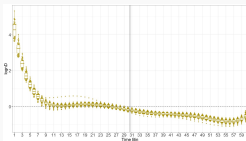
1.  $x_{i,j}$  を各  $i$  に関して ( $j$  方向の平均と標準偏差を使用し) 標準化
2.  $y_{i,j}$  の欠損値 ( $-\infty$  を含む) に各  $i$  に関する平均を代入
3.  $x_{i,j}$  の欠損値に 0 を代入
4. 各  $x_{i,j}$  を  $z_j$  に回帰し日内周期性を除去

# 除去された日内周期性

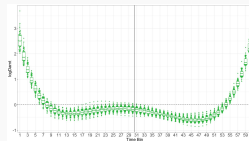
logRet.D



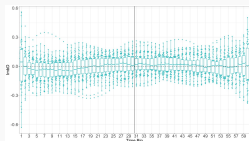
lognD



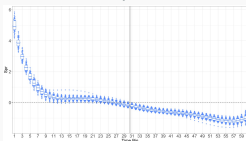
logDamt



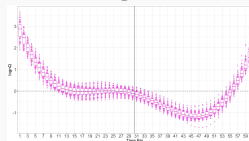
lmbD



Spr



lognQ





## 推定対象

- ・ モデルパラメータ:  $\Theta = (b, g_0, d_0, G, D, \sigma^2, \sigma_\alpha^2, \sigma_\beta^2, \sigma_u^2, \sigma_v^2)$
- ・ 潜在因子:  $\Delta = \{\alpha_i, \beta_j, u_i, v_j\}$  (の平均値)

## MCEM アルゴリズム (Agarwal & Chen, 2016, 8.2)

- ・ E ステップ:
  - Gibbs Sampler によって潜在因子を生成し, 事後平均を計算
  - 各回 “burn-in”10 回後, Gibbs サンプル 100 個抽出 (同, p.155)
- ・ M ステップ:
  - 得られた潜在因子  $\Delta$  の事後平均を使い, 回帰を実行
- ・ 上記の EM 計算を 20 回反復 (同, p.155)

## 線形モデル

- $y_{i,j}$  を  $x_{i,j}, w_i, z_j$  に回帰 (LMO)
- LMO + 個別銘柄および時間帯の固定効果 (LMD)
- LMO + 個別銘柄および時間帯の変量効果 (LMM)

## 従来型 CF

- User-based CF (UBCF)
- Item-based CF (IBCF)
- Singular Value Decomposition (SVD)

データ期間内の 1 日毎に前処理後のデータを使い推定

- ・ RLFM: R で Monte Carlo EM (MCEM) アルゴリズムを実装
- ・ LMM: R のパッケージ *lmer*
- ・ UBCF, IBCF, SVD: R のパッケージ *recommenderlab*

推定結果の評価

- ・ 推定期間内の各日の推定値を 1 データ点とした箱ひげ図
- ・ 各モデルの適合度を RMSE と AIC で比較

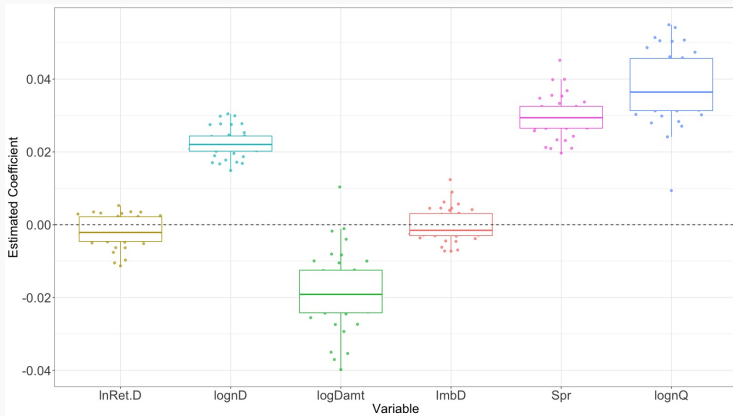
## 手順

1. 銘柄-時間帯行列から, ランダムに予測対象ペア  $(i, j)$  を  $p\%$  抽出  
- 今回の分析では  $p = 20, 50, 80\%$
2. 抽出された予測対象ペアに対応する  $y_h, x_h$  ( $h = 1, \dots, H$ ) に対して, データの前処理を実行
3. 予測期間内の 1 日毎に欠損処理後のデータを使いモデル推定
4. 推定されたモデルに対して, 抽出された予測対象ペア毎に, 真の説明変数  $x_h$  を用いて,  $y_h$  の予測値  $\hat{y}_h$  を計算
5. RMSE と MAE を計算し予測力を評価

$$RMSE = \sqrt{\frac{1}{H} \sum_{h=1}^H (y_h - \hat{y}_h)^2} \quad MAE = \frac{1}{H} \sum_{h=1}^H |y_h - \hat{y}_h|$$

# 目的変数を logILOBS とした推定結果

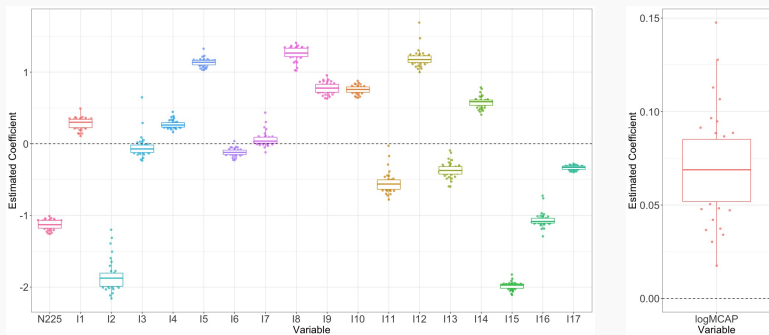
銘柄-時間帯についての説明変数  $x_{ij}$  の係数  $b$



- lognD, Spr, lognQ が大きいほど logILOBS 増加 (流動性低下)
- logDamt が大きいほど logILOBS 減少 (流動性向上)

# 目的変数を logILOBS とした推定結果

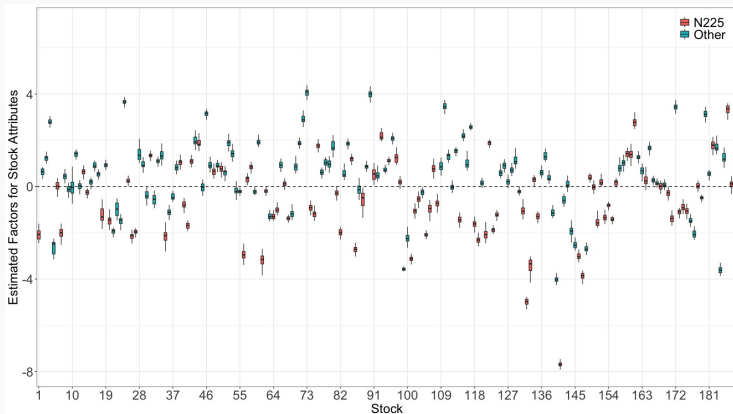
銘柄についての説明変数  $w_i$  の係数  $g_0$



- ・ 業種 5, 8, 12 (医薬品, 機械, 運輸・物流) の銘柄は流動性が低い
- ・ 日経 225, 業種 2, 15, 16 (エネルギー, 銀行, 金融) の銘柄は高い
- ・ 時価総額が高いほど流動性低下

# 目的変数を logILOBS とした推定結果

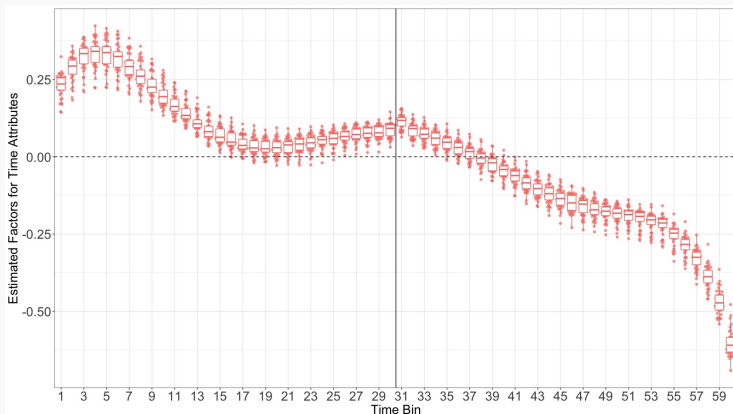
銘柄についての潜在因子  $\alpha_i$



- ・ 日経 225 採用銘柄の方が流動性が高い傾向

# 目的変数を logILOBS とした推定結果

時間帯についての潜在因子  $\beta_j$

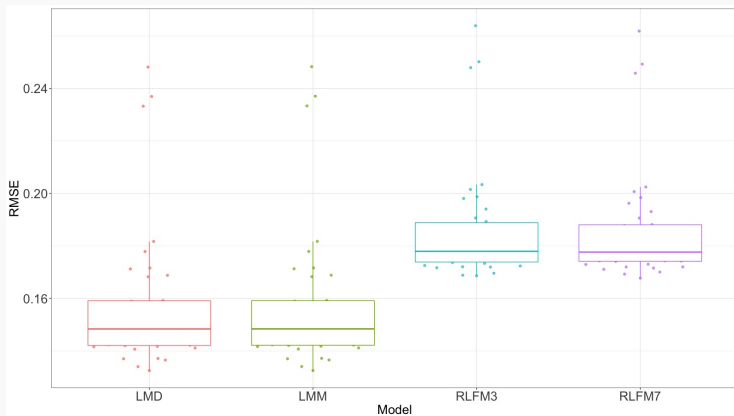


- ・ 寄り付き後は流動性が低く、大引けにかけて大幅に高くなる傾向



# 目的変数を logILOBS とした推定結果

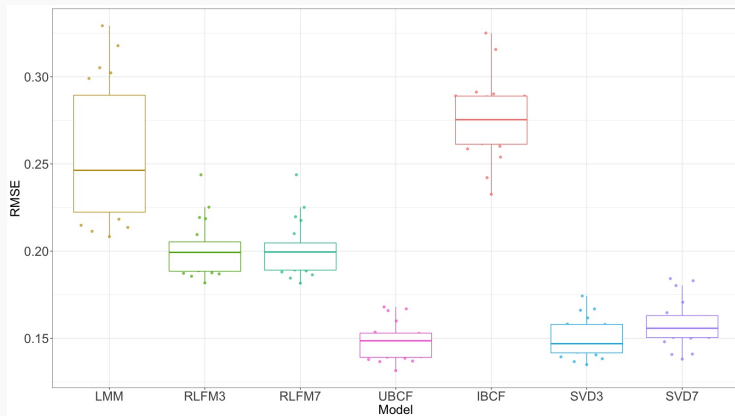
モデルの適合度: RMSE (AIC についても同様の結果)



- RLFM の適合度は LMD, LMM にやや劣る

# 目的変数を logILOBS とした予測結果

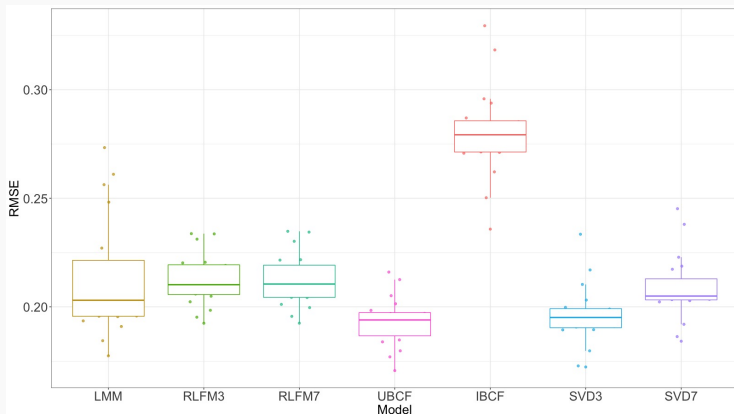
欠損率  $p = 20\%$  での RMSE (MAE も同様の結果)



- 欠損率が低いときは従来型 CF の UBCF, SVD のパフォーマンスが良い

# 目的変数を logILOBs とした予測結果

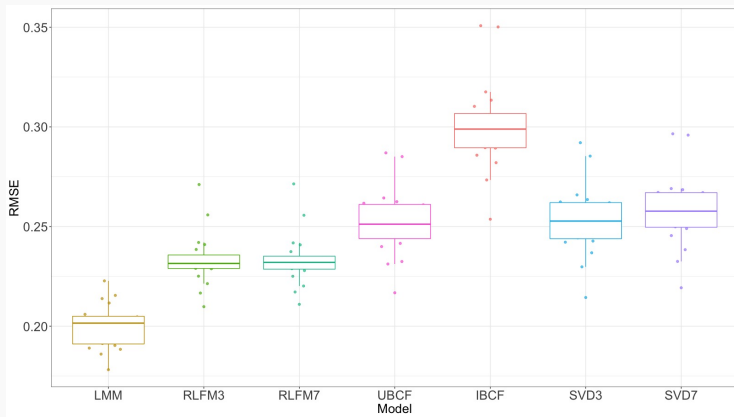
欠損率  $p = 50\%$  での RMSE (MAE も同様の結果)



- 欠損率が高くなると回帰型 LMM のパフォーマンスが良くなる

# 目的変数を logILOBS とした予測結果

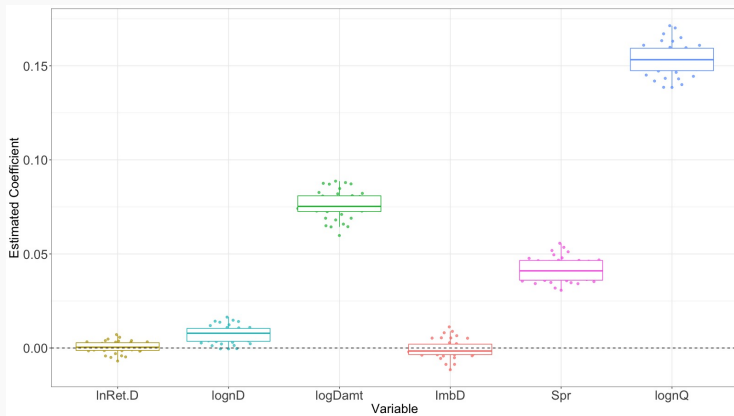
欠損率  $p = 80\%$  での RMSE (MAE も同様の結果)



- ・ 欠損率が高いときは回帰型 LMM のパフォーマンスが良い
- ・ RLFM は欠損率によらず安定したパフォーマンス

# 目的変数を logRV とした推定結果

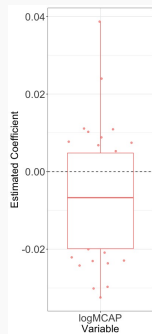
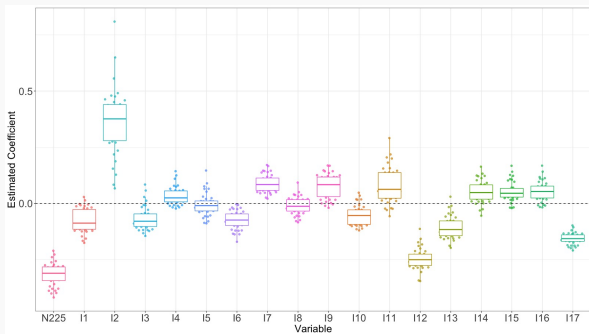
銘柄-時間帯についての説明変数  $x_{ij}$  の係数  $b$



- lognD, logDamt, Spr, lognQ が大きいほどボラティリティ増加

# 目的変数を logRV とした推定結果

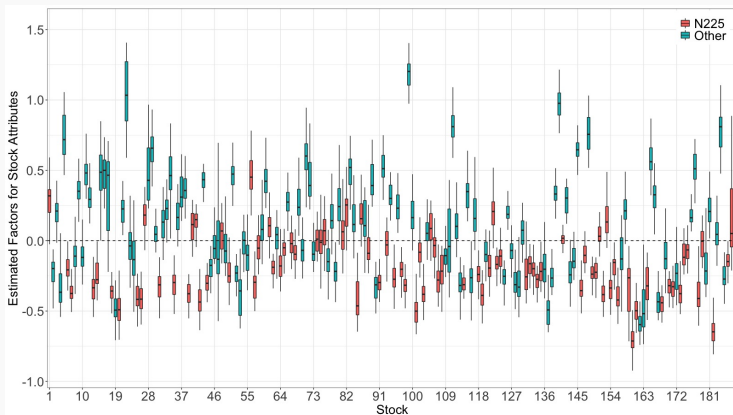
銘柄についての説明変数  $w_i$  の係数  $g_0$



- ・ 日経 225 採用銘柄のボラティリティは低い傾向
- ・ 業種別の違いは logILOBs と比べると小さい
- ・ 時価総額の影響は不定

# 目的変数を logRV とした推定結果

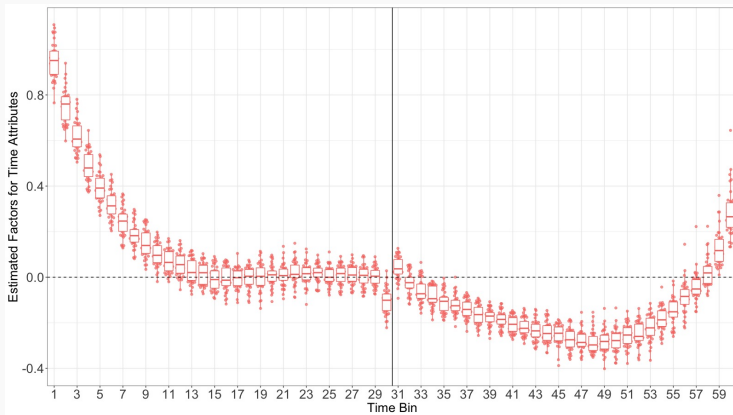
## 銘柄についての潜在因子 $\alpha_i$



- ・ 日経 225 採用銘柄の方がボラティリティが低い傾向

# 目的変数を logRV とした推定結果

時間帯についての潜在因子  $\beta_j$

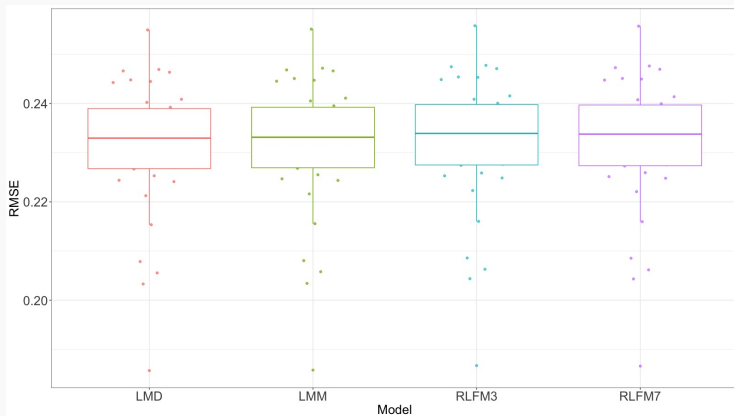


- ・ 寄り付き後はボラティリティが高く、大引けにかけて高くなる傾向



# 目的変数を logRV とした推定結果

モデルの適合度: RMSE (AIC についても同様の結果)



- RLFM の適合度は LMD, LMM と同等

# 目的変数を logRV とした予測結果

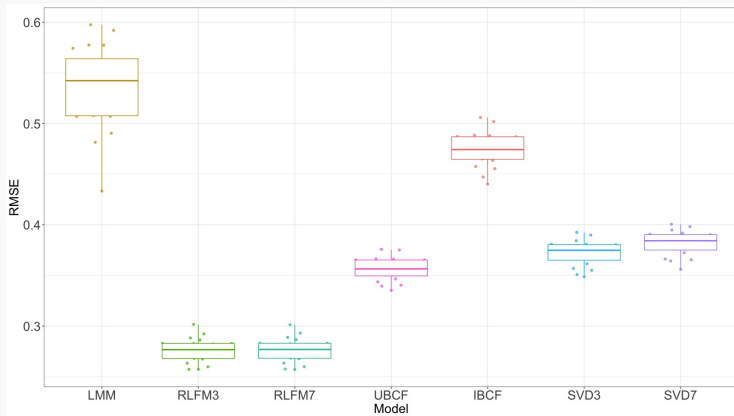
欠損率  $p = 20\%$  での RMSE (MAE も同様の結果)



- ・ 欠損率が低いときは RLFM のパフォーマンスが良い

# 目的変数を logRV とした予測結果

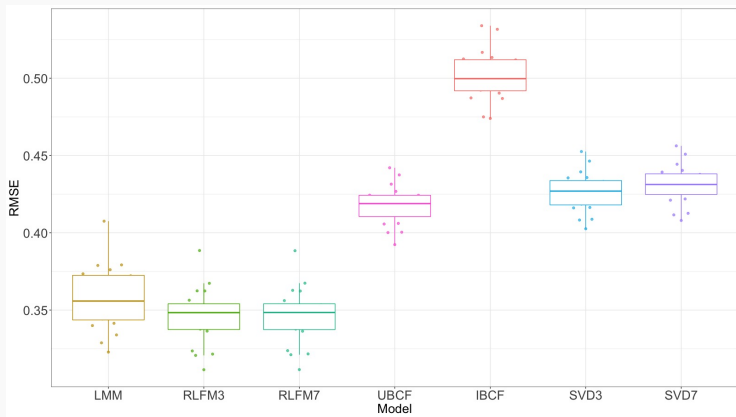
欠損率  $p = 50\%$  での RMSE (MAE も同様の結果)



- ・ 欠損率が高くなると従来型 CF のパフォーマンスが悪くなる
- ・ RLFM のパフォーマンスが良い

# 目的変数を logRV とした予測結果

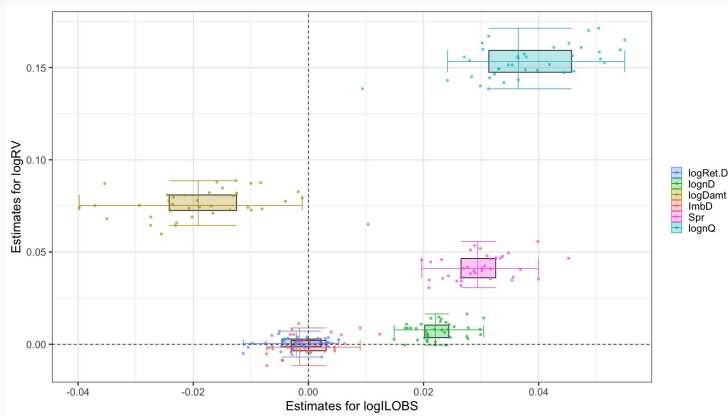
欠損率  $p = 80\%$  での RMSE (MAE も同様の結果)



- ・ 欠損率が高くなると回帰型 LMM のパフォーマンスが良くなる
- ・ RLFM は欠損率によらず安定して良いパフォーマンス

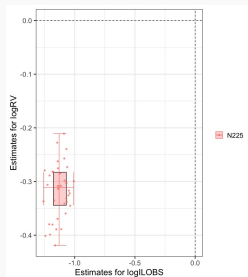
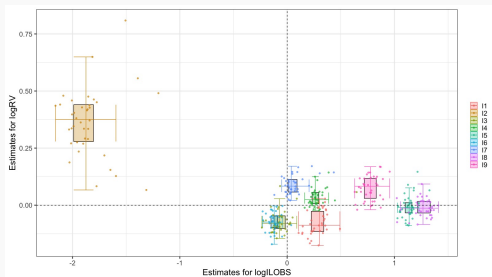
# logILOBs vs logRV

銘柄-時間帯についての説明変数  $x_{ij}$  の係数  $b$



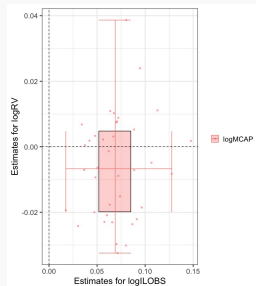
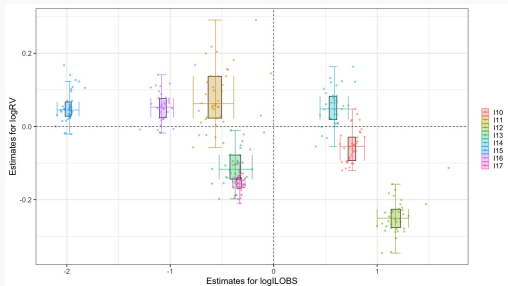
- lognD, Spr, lognQ が大きい → 流動性 ↓, ボラティリティ ↑
- logDamt が大きい → 流動性 ↑, ボラティリティ ↑

銘柄についての説明変数  $w_i$  の係数  $g_0$



- ・ 業種 2 (エネルギー) の銘柄は、流動性が高くボラティリティも高い傾向
- ・ 日経 225 採用銘柄は、流動性が高くボラティリティが低い傾向

銘柄についての説明変数  $w_i$  の係数  $g_0$



- ・ 業種 13, 17 (商社・卸売, 不動産) → 流動性 ↑, ボラティリティ ↓
- ・ 業種 12 (運輸・物流) → 流動性 ↓, ボラティリティ ↓
- ・ 時価総額が大きい → 流動性 ↓, ボラティリティ ↑

RLFM の予測パフォーマンス (RMSE) から以下を選択

- ・ 悪かった例: 2019 年 3 月 25 日の日本たばこ産業 (JT)
- ・ 良かった例: 2019 年 3 月 25 日のアルフレッサ

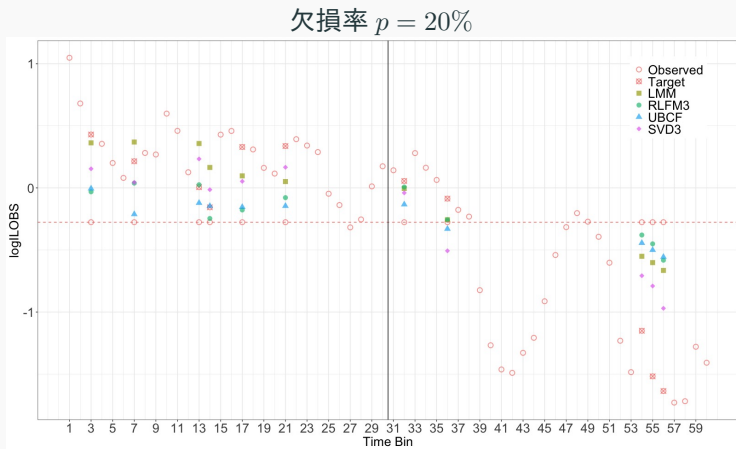
銘柄の特性

- ・ JT: 日経 225 採用, 業種 2 (食品)
  - 時価総額の平均 5.58 兆円 (188 銘柄中 11 位)
- ・ アルフレッサ: 日経 225 非採用, 業種 13 (商社・卸売)
  - 時価総額の平均 0.76 兆円 (188 銘柄中 107 位)



# 目的変数を logILOBS とした予測の実例

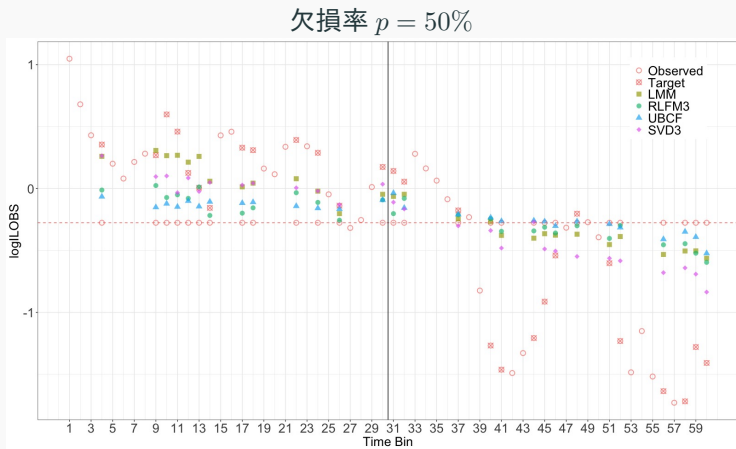
パフォーマンスが悪かった例: 2019年3月25日のJT



- ・ 1日を通して平均から大きく乖離 (日内周期性では説明できない)

# 目的変数を logILOBS とした予測の実例

パフォーマンスが悪かった例: 2019年3月25日のJT

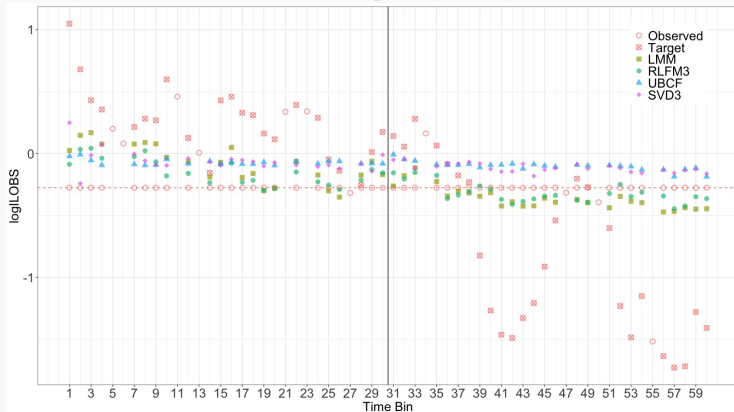


- ・ どのモデルも後場の大きな変動を捉えられていない

# 目的変数を logILOBS とした予測の実例

パフォーマンスが悪かった例: 2019年3月25日のJT

欠損率  $p = 80\%$

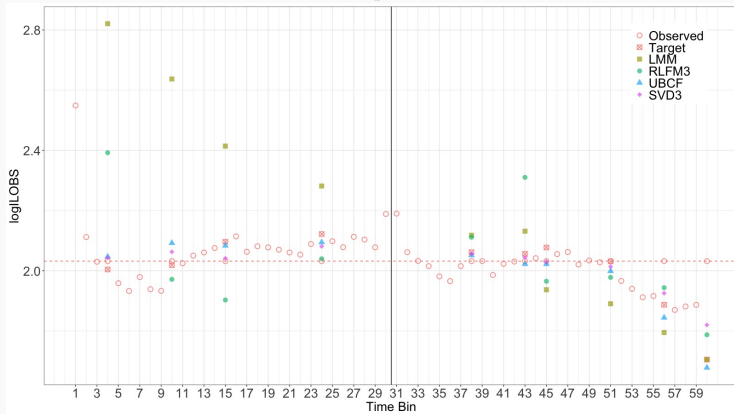


- ・ 欠損率が高くなると予測値が平均周辺に集中していく

# 目的変数を logILOBS とした予測の実例

パフォーマンスが良かった例: 2019年3月25日のアルフレッサ

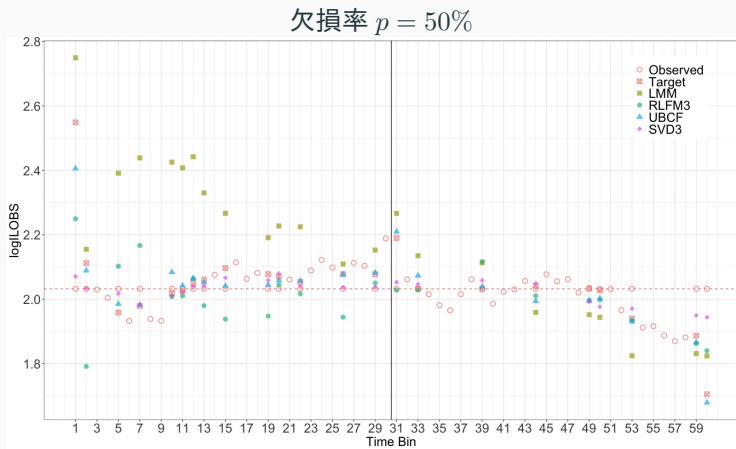
欠損率  $p = 20\%$



・ 1日を通して平均の周辺で変動

# 目的変数を logILOBS とした予測の実例

パフォーマンスが良かった例: 2019 年 3 月 25 日のアルフレッサ

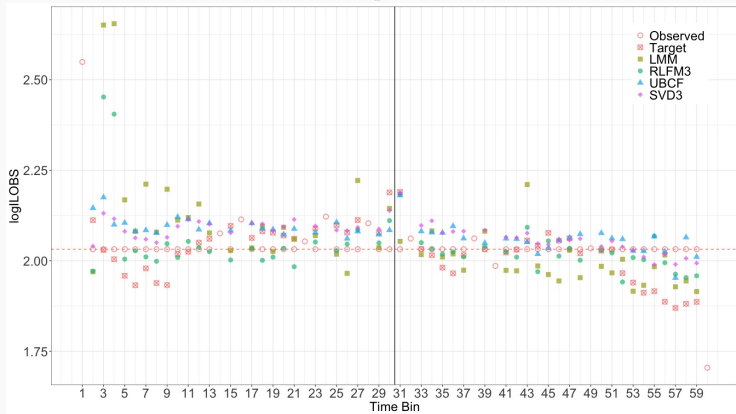


- ・ LMM は (全体の) 日内周期性により寄り付き後を過大に予測

# 目的変数を logILOBS とした予測の実例

パフォーマンスが良かった例: 2019年3月25日のアルフレッサ

欠損率  $p = 80\%$



- 欠損率が高くなると予測値が平均周辺に集中していく

まとめ

---

## 証券市場の市場クオリティの評価に有用な統計的機械学習手法の調査

- ・ 市場クオリティ指標: 非流動性指標 (ILOBS), ボラティリティ (RV)
- ・ レコメンダ・システムのアプローチによって, 流動性の高い低い個別銘柄の市場クオリティを評価する手法 (RLFM) を検討
  - ハイブリッド型 CF で “コールド・スタート問題” に対処
  - 個別銘柄や時間帯の特徴を表わす外生変数の使用

## FLEX Full (東証の高頻度注文板データ) を用いた実証分析

- ・ TOPIX100 銘柄, Mid400 銘柄のうちの 188 銘柄
- ・ データ期間: 2019 年 1 月 4 日 ~ 3 月 29 日 (58 営業日), 5 分間隔
- ・ 流動性を特徴づける要因の特定
- ・ RLFM は欠損率によらず安定した予測パフォーマンス



### データ期間やカバレッジを広げての分析

- ・ 低流動性銘柄の評価には, 財務諸表や業績データ等, 市場データ以外の情報の利用も考えられる

### 流動性予測に関する精度の向上

- ・ 説明変数の選択, 変数の加工法, 時系列構造の考慮
- ・ ILOBS とボラティリティの同時モデリング
- ・ 欠損値の取り扱い (欠損を含んだモデリング)

- Agarwal, Deepak K. and Bee-Chung Chen (2009) “Regression-based latent factor models,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’09)*.
- ——— (2016) *Statistical Methods for Recommender Systems*: Cambridge Univ. Press.
- Amihud, Yakov (2002) “Illiquidity and stock returns: cross-section and time-series effects,” *J. Financial Markets*, Vol. 5, No. 1, pp. 31 – 56.
- Andersen, Torben G. and Tim Bollerslev (1997) “Intraday periodicity and volatility persistence in financial markets,” *Journal of Empirical Finance*, Vol. 4, No. 2, pp. 115 – 158. High Frequency Data in Finance, Part 1.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and P. Labys (2001) “The Distribution of Realized Exchange Rate Volatility,” *Journal of American Statistical Association*, Vol. 96, pp. 42–55.
- Barndorff-Nielsen, Ole E. and Neil Shephard (2002) “Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models,” *Journal of Royal Statistical Society, Series B*, Vol. 64, pp. 253–280.

- Bouchaud, Jean-Philippe, J. Dooyne Farmer, and Fabrizio Lillo (2009) “How Markets Slowly Digest Changes in Supply and Demand,” in Thorsten Hens and Klaus Reiner Schenk-Hoppé eds. *Handbook of Financial Markets: Dynamics and Evolution*, San Diego: North-Holland, Chap. 2, pp. 57 – 160.
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014) “High-frequency trading and price discovery,” *Review of Financial Studies*, Vol. 27, No. 8, pp. 2265–2306.
- Cont, Rama, Arseniy Kukanov, and Sasha Stoikov (2014) “The price impact of order book events,” *J. Financial Econometrics*, Vol. 12, No. 1, pp. 47–88.
- Dacorogna, Michel M., Ramazan Gençay, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet (2001) *An Introduction to High-Frequency Finance*, San Diego: Academic Press.
- Deuskar, Prachi and Timothy C. Johnson (2011) “Market Liquidity and Flow-driven Risk,” *Review of Financial Studies*, Vol. 24, No. 3, pp. 721–753.
- Eisler, Zoltan, Jean-Philippe Bouchaud, and Julien Kockelkoren (2012) “The price impact of order book events: Market orders, limit orders and cancellations,” *Quant. Finance*, Vol. 12, No. 9, pp. 1395–1419.

- Garman, Mark B. and Michael J. Klass (1980) "On the Estimation of Security Price Volatility from Historical Data," *Journal of Business*, Vol. 53, No. 1, pp. 67–78.
- Hautsch, Nikolaus and Ruihong Huang (2012) "The market impact of a limit order," *J. Economic Dynamics & Control*, Vol. 36, No. 4, pp. 501–522.
- Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun (2017) "The Flash Crash: High-Frequency Trading in an Electronic Market," *J. Finance*, Vol. 72, No. 3, pp. 967–998.
- Lee, Charles M. C., Belinda Mucklow, and Mark J Ready (1993) "Spreads, depths, and the impact of earnings information: An intraday analysis," *Review of Financial Studies*, Vol. 6, No. 2, pp. 345–374.
- McNish, Thomas H. and Robert A Wood (1992) "An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks," *The Journal of Finance*, Vol. 47, No. 2, p. 753–764.
- Menkveld, Albert J. (2013) "High frequency trading and the new market makers," *J. Financial Markets*, Vol. 16, No. 4, pp. 712–740.
- ——— (2016) "The Economics of High-Frequency Trading: Taking Stock," *Annual Review of Financial Economics*, Vol. 8, No. 1, pp. 1–24.

- Parkinson, Michael (1980) “The extreme value method for estimating the variance of the rate of return,” *Journal of Business*, Vol. 53, No. 1, pp. 61–65.
- Ricci, Francesco., Lior Rokach, and Bracha Shapira eds. (2015) *Recommender Systems Handbook*: Springer, 2nd edition.
- Wood, Robert A, Thomas H. McInish, and J. Keith Ord (1985) “An Investigation of Transactions Data for NYSE Stocks,” *The Journal of Finance*, Vol. 40, No. 3, pp. 723–739.
- Yano, Makoto (2009) “The Foundation of Market Quality Economics,” *The Japanese Economic Review*, Vol. 60, No. 1, pp. 1–32.
- 高橋慎 (2018). J-GATE 稼働と日経 225 先物市場の日中流動性, 『先物・オプションレポート』, 2018 年 3 月号.