

ビッグデータ解析と再現可能研究 (JRM-BDA-RR-2024)

同時推測と高次元/高頻度データの可視化

小池祐太

2025 年 1 月 25 日

東京大学 大学院数理科学研究科/MI センター
CREST JST

はじめに

- 本報告では金融分野における高次元/高頻度データを用いた以下の2つの実証結果について紹介する

(A) **S&P 500 指数構成銘柄間のネットワーク構造の推定**¹

高次元データを用いた推定結果の可視化において同時推測/多重検定の問題を考慮することが重要となるような例として紹介

(B) **NASDAQ-BATS 取引所間のリード・ラグ解析**²

前年度までの報告で説明した「マイクロ秒単位で記録されたティックデータ」を用いることが“不可避”であるような実証例として紹介

¹詳細は Koike (2020) 参照

²慶應義塾大学の林高樹教授との共同研究. 詳細は Hayashi & Koike (2020) 参照

金融高次元データにおけるネットワーク構造の推定

金融高次元データにおけるネットワーク構造の推定

- 金融データにおいて、ネットワーク構造の把握は重要
 - ショックがどのように波及していくか
 - 「ハブ」となる個体があるか

偏相関ネットワーク

- データの共分散行列 Σ から計算可能
 - 「 i と j は辺で繋がっている ($i \neq j$)」 $\Leftrightarrow$ 「 Σ^{-1} の (i, j) 成分が非ゼロ」
 - データが正規分布に従う場合, $i \neq j$ に対して Σ^{-1} の (i, j) 成分がゼロであることは, i と j がそれ以外の頂点で条件づけると独立となることに由来
- 偏相関ネットワークの推定には, $\Theta = \Sigma^{-1}$ の非ゼロ成分の推定が必要

金融高次元データにおけるネットワーク構造の推定

- d 個の証券の n 期分の収益率のデータが手元にあるとする
- $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{d,t})^\top$: 時点 t での収益率データを並べたベクトル
- 標本共分散行列

$$\hat{\Sigma}_n := \frac{1}{n} \sum_{t=1}^n (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})^\top, \quad \bar{\mathbf{Y}} := \frac{1}{n} \sum_{t=1}^n \mathbf{Y}_t$$

- $\Theta = \Sigma^{-1}$ の推定には $\hat{\Sigma}_n$ の逆行列を用いるのが自然
 - 非ゼロ成分を決定するには, 例えば $\hat{\Sigma}_n$ の各成分がゼロか否か検定すればよさそう

問題点

$d \geq n$ なる高次元の設定では $\hat{\Sigma}_n$ は可逆でない

- 株価指数の構成銘柄間のネットワーク構造を調べる場合は $d \geq 500$
- 一方で、日次データを1年分使っても $n = 250$ 程度
- 高頻度データを使う場合でも、データの非同期性や観測誤差の問題から、1日分のデータは多くても $n = 390$ 程度 (実際には $n < 100$ 程度が多い)
⇒ 日毎のネットワーク構造を推定したければ $d \geq n$ という状況は不可避

- $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ が i.i.d. 列で多次元正規分布に従うならば, 対数尤度関数は以下で与えられる:

$$\ell_n(\Theta; \hat{\Sigma}_n) = \frac{n}{2} \left\{ \log \det \Theta - \text{tr}(\hat{\Sigma}_n \Theta) \right\}$$

(平均は最尤推定量 $\overline{\mathbf{Y}}$ で置き換えて, Θ と無関係な項は除いた)

- Θ の非ゼロ成分の多さに応じてマイナスの値を与える罰則項を $\ell_n(\Theta; \hat{\Sigma}_n)$ に追加した目的関数を最大化することでスパースな Θ の推定量を得ることを試みる

Weighted graphical Lasso

- **Weighted graphical Lasso** (Rothman *et al.*, 2008; Janková & van de Geer, 2018):

$$\hat{\Theta}_\lambda := \arg \max_{\Theta \in \mathcal{S}_d^{++}} \left\{ \frac{2}{n} \ell_n(\Theta; \hat{\Sigma}_n) - \lambda \sum_{i,j=1, i \neq j}^d \sqrt{\hat{\Sigma}_n^{ii} \hat{\Sigma}_n^{jj}} |\Theta^{ij}| \right\}, \quad (1)$$

- $\mathcal{S}_d^{++}$ は d 次正定値対称行列全体の集合
- Θ^{ij} は Θ の (i, j) 成分を表す
- $\lambda > 0$ は罰則の強さをコントロールするチューニングパラメーター
- $\hat{\Sigma}_n$ の対角成分がすべて正ならば, 上の最適化問題の解は一意的に存在する (Duchi *et al.*, 2008, Lemma 1)
- 相関行列ベースの graphical Lasso と同等であり, 共分散行列ベースの重みなし graphical Lasso より理論的性質がよい

ファクター構造

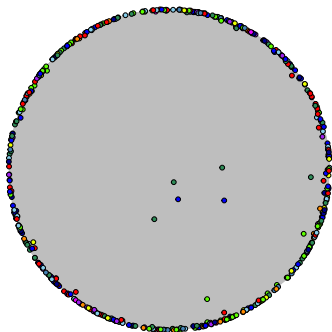
- 金融データにはしばしば既知のファクター構造 (共通因子) が存在
 - 例: 株式市場全体の変動 (株式指数の値動きで代表される)
- そのような場合, すべてのデータがファクターを通じて繋がってしまうので, ファクター由来の相関を取り除いた残差のネットワーク構造の方が興味ある
- $\mathbf{Y}_t$ が以下のようなファクター構造を持つ場合を考える:

$$\mathbf{Y}_t = \beta \mathbf{X}_t + \mathbf{Z}_t$$

- $\mathbf{X}_t = (X_{1,t}, \dots, X_{r,t})^\top$: r 個のファクターの時点 t での収益率を並べたベクトル (観測可能)
- β : $d \times r$ 行列 (ファクターローディング)
- $\mathbf{Z}_t = (Z_{1,t}, \dots, Z_{d,t})^\top$: 個別変動を表す d 次元確率変数 (観測不能)
- $\mathbf{Z}_t$ の共分散行列 Σ_Z の逆行列 $\Theta_Z = \Sigma_Z^{-1}$ のスパース推定に興味がある

- ファクター数 r は n に比べて十分小さく, $\mathbf{X}$ の標本共分散行列 $\hat{\Sigma}_{\mathbf{X},n}$ は可逆であると仮定
 $\Rightarrow \beta$ は OLS 型推定量 $\hat{\beta}_n = \hat{\Sigma}_{\mathbf{YX},n} \hat{\Sigma}_{\mathbf{X},n}^{-1}$ で推定できる ($\hat{\Sigma}_{\mathbf{YX},n}$ は $\mathbf{Y}$ と $\mathbf{X}$ の標本共分散行列)
- Σ_Z は, $\hat{\mathbf{Z}}_t := \mathbf{Y}_t - \hat{\beta}_n \mathbf{X}_t$ の標本共分散行列 $\hat{\Sigma}_{Z,n}$ で推定するのが自然
 $\Rightarrow$ (1) において $\hat{\Sigma}_n$ を $\hat{\Sigma}_{Z,n}$ に置き換えた weighted graphical Lasso 推定量 $\hat{\Theta}_{Z,\lambda}$ で Θ_Z をスパース推定できることが期待される
 - 実際, 適当な正則条件下で一致性を持つことが示せる (Koike (2020) の Theorem 1 参照)

図 1: S&P500 指数構成銘柄の残差系列の偏相関ネットワーク



S&P500 指数をファクター過程とし, 2018 年 3 月 1 日の S&P500 指数構成銘柄の 5 分足データに基づいて, weighted graphical Lasso によって推定した残差系列の偏相関ネットワーク. λ は (形式的) BIC を最小化するように選択した. S&P500 指数の高頻度データは SPY で代用. R パッケージ **yuima** の関数 `cce.factor` で同じ計算が可能. グラフはパッケージ **igraph** で作成

- 一般に, (weighted) graphical Lasso は

- (作用素ノルムに関する) 一致性
- スパース一致性 (ゼロ成分を正しくゼロと推定する)

を両立できない (Lam & Fan (2009) や Bühlmann & van de Geer (2011) の Chapter 7 参照)

- 解決策

- (1) 「有意な」成分のみを残す
- (2) 非凸な罰則項を用いる

⇒ ここでは第1のアプローチを考える: 各 $i < j$ に対して, 仮説検定

$$H_0 : \Theta_Z^{ij} = 0 \quad \text{vs} \quad H_1 : \Theta_Z^{ij} \neq 0 \quad (2)$$

を考え, 帰無仮説が棄却されたペア (i, j) に対応する辺のみ残す

- このためには, $\hat{\Theta}_{Z,\lambda}$ の推定誤差の分布を近似する必要がある

- $\hat{\Theta}_{Z,\lambda}$ は 1 次のオーダーのバイアスを持つため補正する必要がある
- これは「ワンステップ推定」で実行できる: 適当な正則条件の下で,

$$\hat{\Theta}_{Z,\lambda_n} - \Theta_Z = \Gamma_{Z,n} - \Theta_Z(\check{\Sigma}_{Z,n} - \Sigma_Z)\Theta_Z + O_p(\lambda^2)$$

ここに,

- $\Gamma_{Z,n} := -\hat{\Theta}_{Z,\lambda}(\hat{\Theta}_{Z,\lambda}^{-1} - \hat{\Sigma}_{Z,n})\hat{\Theta}_{Z,\lambda}$
- $\check{\Sigma}_{Z,n} := \sum_{t=1}^n (\mathbf{Z}_t - \mathbb{E}[\mathbf{Z}_t])(\mathbf{Z}_t - \mathbb{E}[\mathbf{Z}_t])^\top$
- λ はおよそ $O(\sqrt{(\log d)/n})$ のオーダーにとれる
- $\sqrt{n}(\check{\Sigma}_{Z,n} - \Sigma_Z)$ の分布は CLT によって正規近似できる
⇒ バイアス項を取り除いた

$$\sqrt{n}(\hat{\Theta}_{Z,\lambda} - \Theta_Z - \Gamma_{Z,n}) \tag{3}$$

の分布は正規近似できる

図 2: S&P500 指数構成銘柄の残差系列の偏相関ネットワーク (有意な辺のみ)

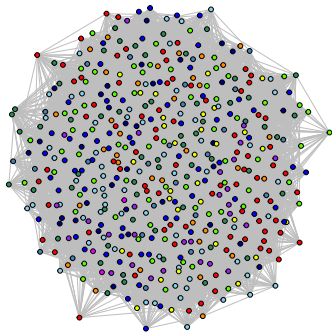
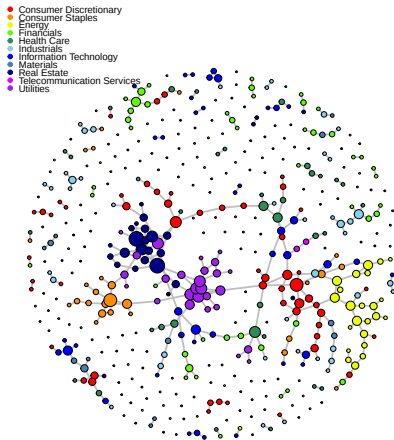


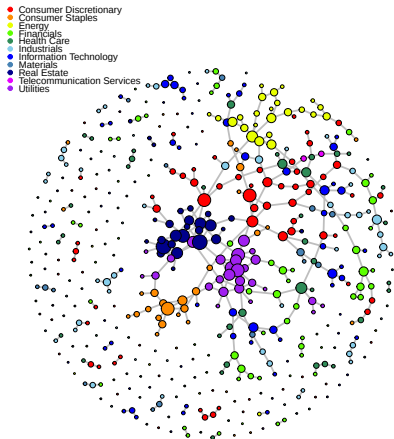
図 1 と同じ設定で, 統計量 (3) の各成分の正規近似に基づく z -検定によって仮説検定 (2) を有意水準 5% で実行し, 帰無仮説が棄却された辺だけ残したグラフ. まだ辺が残りすぎていて解釈が難しい. 多重検定の問題だと思われる.

多重検定の問題

- すべてのペア (i, j) について検定 (2) を実行すると, 計 $\binom{d}{2}$ 個の検定を同時に実行していることになる
 - いまの場合 $d = 504$ なので $\binom{d}{2} = 126756$
- このようなケースでは, あるペアについて帰無仮説が誤って棄却される確率は個々の検定で設定した有意水準よりもはるかに高くなってしまう危険性がある
 - 赤字の確率は Family-wise error rate (FWER) と呼ばれる
- FWER を指定した有意水準以下に抑えるためには個々の検定の p 値を適切に補正する必要がある
- ここでは p 値の間の従属関係に仮定を置かずに利用可能な Holm の方法および Romano–Wolf の方法を適用した結果を紹介する
 - これらの方法の詳細は (Belloni et al., 2018, Section 2.4) 参照



仮説検定 (2) の多重性を考慮し, Holm の方法で p 値を補正した上で, 帰無仮説が棄却された辺だけ残したグラフ. 同じセクターに属する銘柄どうしが繋がる傾向が強いことがわかる. 特に, エネルギー (Energy), 公共事業 (Utilities) および不動産 (Real Estate) でこの傾向が強いように見える.



仮説検定 (2) の多重性を考慮し, Romano-Wolf の方法で p 値を補正した上で, 帰無仮説が棄却された辺だけ残したグラフ. 基本的に Holm の方法と近い結果が得られているが, より検出力が高いため辺の数が増えている.

ティックデータにおける先行遅行関係 の解析

ティックデータにおける先行遅行関係の解析

- 先行遅行関係 (リード・ラグ関係)

- 2つの時系列データの間に時間差を伴って生じる相関関係

- 計量ファイナンスでは古くから多くの実証研究がある

- 株価指数 vs その先物 [先物先行]

- Kawaller *et al.* (1987), Stoll & Whaley (1990), Chan (1992), de Jong & Nijman (1997), ...

- 株価 (指数) vs オプション

- Finucane (1991), Chan *et al.* (1993), de Jong & Donders (1998), ...

- 同一銘柄の複数取引所間

- Garbade & Silber (1979), Hasbrouck (1995), ...

- 外為市場

- de Jong, Mahieu & Schotman (1998), Covrig & Melvin (2002), ...

- 大型株 vs 小型株 [大型株先行]

- Lo & MacKinlay (1990), Hou (2007), ...

...

ティックデータにおける先行遅行関係の解析

- ティックデータ (金融高頻度データ)

- 1 日内にあった約定・注文を全て記録したデータ
- 約定・注文が生じた時刻ごとにデータが記録されている (観測時刻が非等間隔)

- ティックデータではタイムスタンプが非常に重要

⇒ 明示的にモデルに組み込む必要

- 連続時間確率過程の離散観測データとしてモデル化するのが便利

- 価格系列に基づく解析手法

- タイムラグ付き連続セミマルチンゲールによるモデル化 (Hoffmann, Rosenbaum & Yoshida, 2013, HRY モデル)
- 多次元 Hawkes 過程によるモデル化 (Bacry *et al.*, 2013; Da Fonseca & Zaatour, 2015)

- タイムスタンプに基づく解析手法

- 同時に起きた「イベント」の回数を計測 (Dobrev & Schaumburg, 2016, DS 推定量)
- Neyman–Scott 過程によるモデル化 (Shiotani & Yoshida, 2024)

DS 推定量 (Dobrev & Schaumburg, 2016)

- 2 種類の金融資産 A と B の観測データについて, **タイムラグ $\theta \in \mathbb{R}$ をもって同時に生じた観測の個数**を以下の統計量で計測:

$$U^{DS}(\theta) = \frac{1}{\min\{n_1, n_2\}} \sum_{k=1}^{\lfloor T/\delta_N \rfloor} I_{k\delta_N}^A I_{k\delta_N+\theta}^B$$

ここに,

- $X = A \text{ or } B$ と $t \geq 0$ に対して

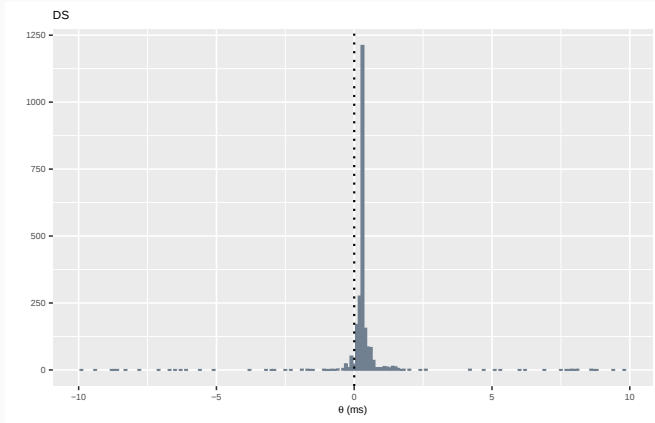
$$I_t^X = \begin{cases} 1 & \text{資産 } X \text{ が時刻 } t \text{ で観測される場合,} \\ 0 & \text{それ以外} \end{cases}$$

- δ_N はタイムスタンプの最小解像度 (本分析では $1\mu s$)
- 本分析では「観測＝注文の発生」
- DS 推定量 $\hat{\theta}^{DS}$ は以下で定義される:

$$\hat{\theta}^{DS} = \arg \max_{\theta \in \mathcal{G}} U^{DS}(\theta).$$

NASDAQ-BATS 取引所間のリード・ラグ解析

- 米国株式市場には 13 の主要な取引所があり、それぞれの取引所に売買注文をかけることができる
 - 同一の銘柄であっても、取引所ごとに異なる価格がつく
⇒ 取引所ごとの価格系列どうしの先行遅行関係を調べられる
 - これは、取引所ごとの市場参加者の違いや、どのような要素が時間差の大小に影響を与えるかということに関する知見を得る上で有用
- NASDAQ 取引所と BATS 取引所に着目して、下記の設定で取引所ごとの価格系列どうしの先行遅行関係を調べる
 - 期間: 2015 年 8 月の各取引日 (計 21 日)
 - 市場の開閉前後の非正常な振る舞いの影響を排除するため、開閉直後の 15 分のデータは取り除く (9:45 から 15:45 のデータを使用)
 - 銘柄: 2015 年 8 月時点での NASDAQ100 指数構成銘柄 (計 108 銘柄)
 - データソース: Daily TAQ データベースの最良気配データを使用
 - タイムスタンプの精度はマイクロ秒



- $\mathcal{G}$ を -10ms から 10ms まで 0.1ms 刻みのグリッドにとって DS 推定量を計算し, $21 \times 108 = 2268$ の推定値を得てヒストグラムを作成
- $\theta > 0$ は NASDAQ が BATS を先行することを意味する
- DS 推定量の推定値は $\theta = +0.3\text{ms}$ に集中している
⇒ 「NASDAQ が BATS に対して一貫して 0.3ms 先行する」というリード・ラグ関係の存在が示唆される

NASDAQ-BATS 取引所間のリード・ラグ解析

- Dobrev & Schaumburg (2016) は,
 - 米国債 10 年物とその先物
 - S&P500 指数の ETF とその先物

のそれぞれに対して DS 推定量によるリード・ラグ解析を実施し、いずれの場合も先物が **5ms** 先行することを報告

- Dobrev & Schaumburg (2016) は、この **5ms** というタイムラグが、地理的要因から生じていると推測
 - 米国債の取引: Brokertec, サーバ@シコーカス (Secaucus), NJ
 - S&P500 指数の ETF (SPDR S&P 500 ETF) の取引: 上記 13 の取引所, サーバ@シコーカス周辺
 - 米国債先物・S&P500 指数先物 (E-Mini S&P 500) の取引: シカゴ・マーカンタイル取引所 (CME), サーバ@オーロラ (Aurora), IL
 - シコーカス-オーロラ間を光速度で移動すると約 4.7ms かかる

NASDAQ-BATS 取引所間のリード・ラグ解析

- 同様の考察を我々の設定で行ってみる
 - NASDAQ のサーバ@カーテレット (Carteret), NJ
 - BATS のサーバ@シコーカス, NJ
 - カーテレット-シコーカス間を光速度で移動すると約 **0.09ms** かかる (Tivnan et al., 2020, Table 2)
- DS 推定値 = 0.3ms $\Rightarrow$ **余計な 0.2ms はどこから来たのか？**
 - 米国株式市場では、各取引所での約定・気配情報が Securities Information Processor (SIP) に集約されたのち、市場全体での最良売買気配 (NBBO) が (一般の) 投資家に提供される
 - Daily TAQ データの「Time」の項目に記載されているタイムスタンプは、各取引所で生じた約定・気配情報が SIP で処理された時刻に対応
 - 一方で、一部の投資家は各取引所から直接約定・気配情報を入手
 - $\Rightarrow$ **各取引所から SIP に情報が送られるまでのタイムラグを考慮すべき**

NASDAQ-BATS 取引所間のリード・ラグ解析

- 実は, Daily TAQ データには, 「Participant Timestamp」という項目があり, 取引所から SIP へと情報が送信された時刻が記録されている
- そこで, 「Participant Timestamp」をタイムスタンプに取り直して DS 推定量を計測し直してみる

表 1: DS 推定値の要約統計量の比較 (単位は ms)

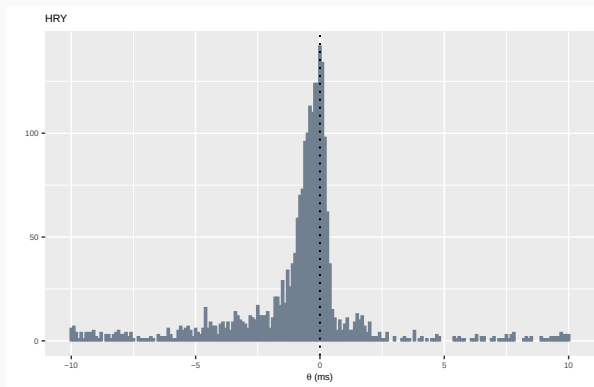
	1st Qu.	Median	3rd Qu.	Mode
SIP Timestamp	0.20	0.30	0.30	0.30
Participant Timestamp	-0.10	0.10	0.10	0.10

- Participant Timestamp で測ると, DS 推定値は 0.1ms に集中
- DS 推定量は「超高速な」市場参加者間の先行遅行関係, もしくは「通信上の限界値」をとらえていそう

補足 1: Participant Timestamp を分析に使う際の問題点

- Participant Timestamp は各取引所の時計に基づいて記録されているため、取引所間の時計のズレに影響される (clock synchronization の問題)
- 各取引所は協定世界時 (UTC) からの時計のズレを 0.1ms 以内に抑えることを要求されているが、微小なズレは存在しうる
- 実際のズレは平均 36 マイクロ秒程度らしい (Hasbrouck, 2019, p.12)

補足 2: HRY モデルによる推定値



- HRY モデルに基づくタイムラグの推定値 (価格系列に基づく方法) のヒストグラムは負の方向に歪んでおり, 「BATS が NASDAQ に対して先行する」というリード・ラグ関係の存在を示唆 $\Rightarrow$ DS 推定値による結果と矛盾
- この結果は「市場参加者の不均一性から生じる異なるタイムスケールにおける価格形成の違い」(不均一市場仮説) の現れとして説明できそう
- 詳細は Hayashi & Koike (2020) 参照

References

- Bacry, E., Delattre, S., Hoffmann, M. & Muzy, J. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123**, 2475–2499.
- Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C. & Kato, K. (2018). High-dimensional econometrics and regularized GMM. Working paper. arXiv: 1806.01888.
- Bühlmann, P. & van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer.
- Chan, K. (1992). A further analysis of the lead–lag relationship between the cash market and stock index futures market. *Review of Financial Studies* **5**, 123–152.
- Chan, K., Chung, Y. P. & Johnson, H. (1993). Why option prices lag stock prices: A trading-based explanation. *Journal of Finance* **48**, 1957–1967.

- Covrig, V. & Melvin, M. (2002). Asymmetric information and price discovery in the FX market: does Tokyo know more about the yen? *Journal of Empirical Finance* **9**, 271–285.
- Da Fonseca, J. & Zaatour, R. (2015). Correlation and lead-lag relationships in a Hawkes microstructure model. *Journal of Futures Markets* **37**, 260–285.
- de Jong, F. & Donders, M. W. M. (1998). Intraday lead-lag relationships between the futures-, options and stock market. *European Finance Review* **1**, 337–359.
- de Jong, F., Mahieu, R. & Schotman, P. (1998). Price discovery in the foreign exchange market: an empirical analysis of the yen/dmark rate. *Journal of International Money and Finance* **17**, 5–27.
- de Jong, F. & Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance* **4**, 259–277.
- Dobrev, D. & Schaumburg, E. (2016). High-frequency cross-market trading: Model free measurement and applications. Working paper.

- Duchi, J., Gould, S. & Koller, D. (2008). Projected subgradient methods for learning sparse Gaussians. In *Proceedings of the twenty-fourth conference on uncertainty in artificial intelligence*. AUAI Press, pp. 153–160.
- Finucane, T. J. (1991). Put-call parity and expected returns. *Journal of Financial and Quantitative Analysis* **26**, 445–457.
- Garbade, K. D. & Silber, W. L. (1979). Dominant and satellite markets: A study of dually-traded securities. *The Review of Economics and Statistics* **61**, 455–460.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance* **50**, 1175–1199.
- Hasbrouck, J. (2019). Price discovery in high resolution. *Journal of Financial Econometrics (forthcoming)* .
- Hayashi, T. & Koike, Y. (2020). Multi-scale analysis of lead-lag relationships in high-frequency financial markets. Working paper. Available at arXiv: <https://arxiv.org/abs/1708.03992>.

- Hoffmann, M., Rosenbaum, M. & Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.
- Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies* **20**, 1113–1138.
- Janková, J. & van de Geer, S. (2018). Inference in high-dimensional graphical models. In *Handbook of graphical models*, chap. 14. CRC Press, pp. 325–351.
- Kawaller, I. G., Koch, P. D. & Koch, T. W. (1987). The temporal price relationship between S&P 500 futures and the S&P 500 index. *Journal of Finance* **42**, 1309–1329.
- Koike, Y. (2020). De-biased graphical Lasso for high-frequency data. *Entropy* **22**, 456.
- Lam, C. & Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist.* **37**, 4254–4278.
- Lo, A. W. & MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies* **3**, 175–205.

- Rothman, A. J., Bickel, P. J., Levina, E. & Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Stat.* **2**, 494–515.
- Shiotani, T. & Yoshida, N. (2024). Statistical inference for highly correlated stationary point processes and noisy bivariate Neyman–Scott processes. Working paper. arXiv: 2410.05732.
- Stoll, H. R. & Whaley, R. E. (1990). The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative Analysis* **25**, 441–468.
- Tivnan, B. F., Dewhurst, D. R., Van Oort, C. M., Ring, J. H., Gray, T. J., Tivnan, B. F., Koehler, M. T. K., McMahon, M. T., Slater, D. M., Veneman, J. G. & Danforth, C. M. (2020). Fragmentation and inefficiencies in US equity markets: Evidence from the Dow 30. *PLoS ONE* **15**, e0226968.