

非同期時系列の Lead-Lag 効果推定のための新しい推定量 Mining Lead-Lag Relationship from Non-synchronous Data

伊藤 克哉^{1*} 中川 慧²
Katsuya Ito¹ Kei Nakagawa²

¹ Preferred Networks 株式会社

¹ Preferred Networks, Inc.

² 野村アセットマネジメント株式会社

² Nomura Asset Management Co, Ltd.

Abstract: リードラグ効果は、金融市場のいたるところで観察され、特に高頻度データを用いた投資戦略を策定する上で重要な要因である。一方で、Lead-Lag 効果を推定するために、次の3つの課題が存在する。(1) 高頻度データにおいては、2つの時系列が常に同時に観測できるとは限らない(非同期)。(2) 高頻度データのサイズは大きく、応用のためには短期間で推定を完了する必要がある。(3) Lead-Lag 効果は時变的であり、かつ短期間しか持続せず、しばしば外部要因の影響を受ける。本研究では、これらの課題をすべて解決する、新しい Lead-Lag 効果の推定量 (NAPLES: Negative And Positive lead-lag EStimator) を提案する。人工データセットと実際の金融市場のデータセットを用いた実験の結果、NAPLES は、重要なマクロ経済のアナウンス (外部要因) によって引き起こされたものを含め、Lead-Lag 効果と強い相関関係が確認できた。

1 はじめに

裁定機会 (arbitrage opportunity) の解析は数理ファイナンスにおいて重要な課題の一つである。裁定機会の代表的なものに一つの銘柄の値動きが他の銘柄の値動きよりも先行または遅行する Lead-Lag 効果がある [19, 21, 12, 9]。先行している銘柄の値動きに合わせて遅行している銘柄を売買することによりリスクを取ることなくリターンを発生することができるため、Lead-Lag 効果については古くから計量経済学や金融工学の分野で研究がされてきた [21, 12]。

投資判断・市場構造分析において、より高速・正確に Lead-Lag 効果を推定することの重要性が近年増大している。その理由の1つは、近年、高速な情報通信と計算機を用いた電子取引が可能となり、市場はより効率的になり Lead-Lag 効果の探索は難しくなったからである [8, 6, 24]。2つ目の理由として、直近ではミリ秒・ナノ秒単位で電子取引を行う高頻度取引が登場したため、より短い時間の Lead-Lag 効果を利用して収益を得ることが可能となった [3, 25, 22]。

はじめに Lead-Lag 効果の推定の方法論を数学的に定式化する。この推定は、2つ時系列を観測し、一方のタイムスタンプをずらしながら相関係数を最大化す

る方法が一般的である。資産価格 X_t を $0 = s_0 < s_1 < s_2 < \dots < s_n$ という時刻で観測し、資産価格 Y_t を $0 = t_0 < t_1 < t_2 < \dots < t_m$ という時刻で観測する。 X_t が Y_t を **Lead-Lag** パラメタ θ で先行しているとは、任意の $\theta' \neq \theta$ について、 X_t と $Y_{t+\theta}$ の相関が X_t と $Y_{t+\theta'}$ の相関よりも強いことをいう。この Lead-Lag パラメタを推定するには、 $\hat{C}orr$ という相関係数の推定量を定義し、以下の最大化問題を解く。

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \hat{C}orr(\{X_{s_i}\}_{i=1}^n, \{Y_{t_j+\theta}\}_{j=1}^m) \quad (1)$$

高頻度取引における Lead-Lag 効果の推定における重要かつ難解な課題は大きく3つあり、それらは、(1) タイムスタンプの非同期性、(2) 計算量の膨大性、(3) Lead-Lag 効果の時変性である。

(1) タイムスタンプの非同期性とは、tick データに対して Lead-Lag 効果を計算するときの課題である。tick データは市場で起こった全てのイベントを保存するので、タイムスタンプが揃っていない。この課題に対しては Hayashi & Yoshida(2005) が非同期に観測される時系列データに対して共分散推定量を定義し [16]、Hoffman, Rosenbaum, & Yoshida(2013) がその推定量を最小化することによって得られる Lead-Lag 効果の推定量を定義したことにより、解決方法が提案された [17]。しかし、その手法を実世界に適用するために様々な理論的解析 [10, 14, 15] および実証的解析 [1, 5, 7, 18] が現時点でも行われている。

*連絡先: Preferred Networks 株式会社
〒100-0004 東京都千代田区大手町 1-6-1 大手町ビル 2F
E-mail: katsuyalito@gmail.com

(2) 計算量の膨大性とは、高頻度取引が一般的になるにつれて、計算量が膨大となる課題である。データ数 T 、銘柄数 N 、ラグ候補数 L としたときに現在の計算量は、 $O(T^2 N^2 L)$ のオーダーでの計算量が多く、これ以上の計算オーダーの計算アルゴリズムは開発されていない。

(3) Lead-Lag 効果の時変性とは、Lead-Lag パラメタが時間に応じて変化することで推定が困難になるという課題である。基本的に Lead-Lag 効果は、情報の非対称性 [24] やトレーディング・ボリュームの差 [11] によって価格変更が遅れていること等に起因して発生し、常に発生しているわけではない [2, 4, 23]。そのために時間変化するラグを扱うことのできる方法が必要とされている [20]。本論文では、これらの3つの問題を解決するための Lead-Lag 効果の新しい推定量を提案する。提案方法は、非同期に観測される時系列のセットを使用して実行できるだけでなく、観測数に対して線形な計算複雑性で実行が可能である。本論文の主な貢献は下記の通りである。

- 3つの課題を解決する Lead-Lag 効果の新しい推定法を提案した。
- 推定法の数学的な性質を与えるとともにその解釈を行った。
- 人工データ・実データを用いた実験により、提案手法の有効性を示すとともに、マクロ指標の発表前にリードラグの推定を行い、その考察を行った。

2 問題設定と先行研究

ここでは、リードラグ効果の推定の定式化と先行研究において提案された手法を概観する。

2.1 Notation

2つの実数 $a, b \in \mathbb{R}$ の区間を次のように書く。 $[a, b] = \{x | a \leq x \leq b\}$, $]a, b[= \{x | a < x < b\}$, $[a, b[= \{x | a \leq x < b\}$, $]a, b] = \{x | a < x \leq b\}$ また、 $a \vee b$ と $a \wedge b$ はそれぞれ $\max\{a, b\}$ と $\min\{a, b\}$ を表す。 1_A は条件 A についての指示関数である。

2.2 問題設定

X_t, Y_t を2つの確率過程 (e.g., 資産価格) とする。 X_t は時間間隔 $0 = s_1 < s_2 < \dots < s_n = T$ で観察され、 Y_t は時間間隔 $0 = t_1 < t_2 < \dots < t_m = T$ でそれぞれ観察されるとする。はじめに Lead-Lag 効果と非同期観測の定義を行う。

Definition 1 X_t が Y_t に対して、Lead-Lag パラメタ θ の Lead-Lag 効果を持つとは、任意の α に対して、 $X_t, Y_{t+\alpha}$ の相関係数の絶対値が $X_t, Y_{t+\theta}$ よりも大きいことをいう

Definition 2 X_t と Y_t が、すべての i に対して、 $s_i = t_i$ が成立する場合、同期観測であるといい、同期観測でない場合、非同期観測という。

本研究における問題設定は、非同期観測されたデータから Lead-Lag 効果を推測することである。つまり、

Problem 1 X_t が Lead-Lag パラメタ θ を持ち、 Y_t を先行すると仮定する。このとき、非同期観測データ $\{X_{s_i}\}_{i=0}^n$ および $\{Y_{t_j}\}_{j=0}^m$ から、Lead-Lag パラメタ θ を推定する。

2.3 先行研究

高頻度取引が盛んになる前は、同期観測のデータに対する Lead-Lag の推定量が Lo & MacKinlay(1990)[21] や de Jong & Nijman [12] によって考案されていた。

高頻度取引が盛んになり非同期観測場合における Lead-Lag の推定のために Hayashi & Yoshida(2005) は次のような共分散推定量を提案した、

Definition 3 資産価格 X_t を $s_1 < s_2 < \dots < s_n$ という時刻で観測し、資産価格 Y_t を $t_1 < t_2 < \dots < t_m$ という時刻で非同期的に観測するとき、Hayashi-Yoshida の共分散推定量は $\text{HY}(X_s, Y_t, \{s_i\}_{i=1}^n, \{t_j\}_{j=1}^m)$

$$= \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} (X_{s_{i+1}} - X_{s_i})(Y_{t_{j+1}} - Y_{t_j}) 1_{\{[t_{i-1}, t_i] \cap [s_{j-1}, s_j] \neq \emptyset\}}$$

のように定義されるものである。この共分散推定量を用いて、

$$U^n(\tilde{\theta}) := 1_{\tilde{\theta} \geq 0} \text{HY}(X_{s-\tilde{\theta}}, Y_t, \{s_i - \tilde{\theta}\}_{i=1}^n, \{t_j\}_{j=1}^m)$$

$$+ 1_{\tilde{\theta} < 0} \text{HY}(X_{s+\tilde{\theta}}, Y_t, \{s_i + \tilde{\theta}\}_{i=1}^n, \{t_j\}_{j=1}^m)$$

のように時系列をシフトした共分散推定量を作ることによって Lead-Lag 効果の推定する。

この推定量の理論的拡張として、Chiba(2017) はこの推定量を非整数ブラウン運動の場合に拡張した。また、Hayashi & Koike(2017) [14] はフーリエ変換を用いてこの推定量の改良を行い、Hayashi & Koike(2018) [15] はウェーブレット変換を用いてこの推定量の改良を行った。

一方でこれらの研究と独立した形で、Dobrev & Schaumburg は次のような Lead-Lag 推定のための数値を考案している。

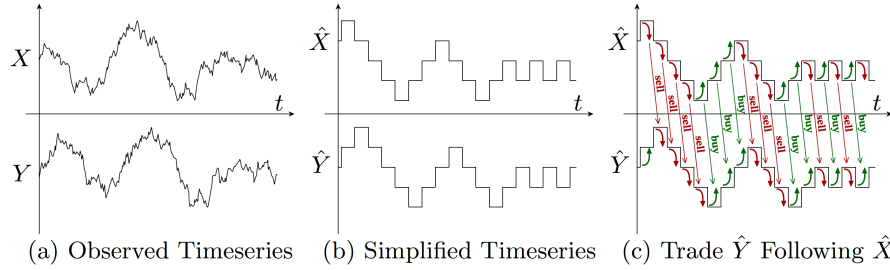


図 1: アルゴリズムの概念図

Definition 4 (式 (2) in [13]) 2 資産 X, Y が観測されている状況を考える。確率過程 \hat{Z}_t は、ある資産 $Z = X, Y$ が時刻 t において取引されているときに 1, それ以外の時に 0 を返すような確率過程である。このとき、Dobrevva-Schaumburg の指数 DS_t とは次のように計算されるものである。

$$DS_t := \frac{\sum_{i=|t|}^{N-|t|} \hat{X}_{i+t} \hat{Y}_i}{\sum_{i=|t|}^{N-|t|} \hat{X}_{i+t} \wedge \sum_{i=|t|}^{N-|t|} \hat{Y}_i}.$$

Dobrevva-Schaumburg の指数は Hoffman-Rosenbaum-Yoshida と同様に t に関して探索を行うことによって Lead-Lag 効果の推定を行うことができる。その統計的性質は明らかになっていないが、実証的研究によってその有効性が明らかにされている。

3 提案手法:NAPLES

提案手法は、(1) 時系列の符号による単純化と (2) 単純化した時系列をもちいた投資戦略の計算をする。(1) によって、時系列を単純化し、Lead-Lag に機敏に反応する推定量を作ることができる。また (2) によって、タイムスタンプの非同期性や Lead-Lag パラメタの時間変化を考慮した、高速な計算をすることができる。

ここで、 $\{Z_{u_k}\}_{k=1}^n \in \{\{X_{s_i}\}_i, \{Y_{t_j}\}_j\}$ を観測したとき、その対数リターンとリターンの符号をそれぞれ $r_{u_k}^{(Z)} = \log(Z_{u_k}/Z_{u_{k-1}})$, $b_{u_k}^{(Z)} = \text{sign}\{r_{u_k}^{(Z)}\}$ と書く。このとき、 \hat{Z}_t を $b_{u_k}^{(Z)}$ の累積和 $\hat{Z}_t = \sum_{k=1}^n 1_{u_k < t} b_{u_k}^{(Z)}$ として定義する。

Definition 5 (NAPLES; Negative And Positive lead-lag EStimator) X_t を時刻 $0 = s_1 < s_2 < \dots < s_n = T$ で観察し、 Y_t を $0 = t_1 < t_2 < \dots < t_m = T$ で観察する。このとき、指数 NAPLES、 $R(t; X_t, Y_t)$ は次のように定義される。

$$R(t; X_t, Y_t) := \sum_{i=1}^{n-1} \left(b_{s_i}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i} \right) 1_{s_{i+1} < t} - \sum_{j=1}^{m-1} \left(b_{t_j}^{(Y)} \hat{X}_{t_{j+1}} - b_{t_j}^{(Y)} \hat{X}_{t_j} \right) 1_{t_{j+1} < t}$$

この定義は一見すると複雑に見えるが、以下のような単純な発想に基づくものである。つまり、 X が Y に対して先行しているとき、 Y を X の価格変動に応じて取引したときの利益または損失が $R(t; X_t, Y_t)$ である。具体的には、 $R(t; X_t, Y_t)$ の第一項は、時刻 s_i で X の価格が上昇(下落)したときに、時刻 s_i で \hat{Y} を買い(売り)、時刻 s_{i+1} で \hat{Y} を売る(買う)戦略のリターンを表している。一方で、第二項は、時刻 s_i で Y の価格が上昇(下落)したときに、時刻 s_i で \hat{X} を買い(売り)、時刻 s_{i+1} で \hat{X} を売る(買う)戦略のリターンを表している。図 1 が我々のアルゴリズムの概念図である。これを用いると、Lead-Lag パラメタ θ は次のように計算することができる。

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} R(T, X_t, Y_{t+\theta}) \quad (2)$$

NAPLES は、上記の 3 つの問題を解決する。すなわち、(1) 非同期観測の時系列に対して計算可能であり、(2) 推定量は単に N 項の合計であるため、 $O(N)$ のオーダーで計算可能であり、(3) 取引戦略のリターンであるため、各時点ごとに Lead-Lag を推定でき、Lead-Lag の時変的な性質を捉えることができる。

さらに、 $R(t; X_t, Y_t)$ の性質を表す下記の定理を示す。

Theorem 1 二つの資産価格 X, Y が相関 ρ 、Lead-Lag パラメタ θ を持つ幾何ブラウン運動に従うとする。さらに観測時刻 s_i と t_j が常に等間隔 Δ であると仮定する。このとき、 $R(T; X_t, Y_t)$ は次を満たす。

$$E[R(T; X_t, Y_t)] = \begin{cases} L \arcsin\left(\frac{\rho|\theta|}{\Delta}\right) & 0 < |\theta| < \Delta \text{ のとき} \\ L \arcsin\left(\frac{\rho(2\Delta - |\theta|)}{\Delta}\right) & \Delta < |\theta| < 2\Delta \text{ のとき} \\ 0 & \text{それ以外の場合} \end{cases}$$

ただし l は $\theta > 0$ のとき $m-1$ で、 $\theta < 0$ のとき $n-1$ であるような定数であり、 $L = \frac{l}{\pi} \text{sign}(\theta)$ である。

この定理の証明を一般化した定理とともに Appendix に示す。

3.1 提案手法の定性的解釈

前節では提案手法の数学的性質を明らかにした。この節では提案手法の具体例を計算することによってそ

の原理を定性的に明らかにする。Theorem 1から等間隔に観測される場合において我々の指数の期待値は以下のようになることがわかっている。

$$E[R(T)] = 1_{0 < |\theta| < \Delta} \frac{l}{\pi} \text{sign}(\theta) \arcsin\left(\frac{\rho|\theta|}{\Delta}\right) + 1_{\Delta < |\theta| < 2\Delta} \frac{l}{\pi} \text{sign}(\theta) \arcsin\left(\frac{\rho(2\Delta - |\theta|)}{\Delta}\right).$$

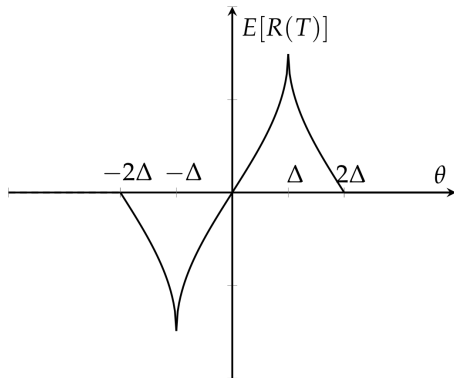


図 2: θ を変化させたときの $E[R(T)]$ の変化

図 2 は、Lead-Lag 効果を表す θ のみを動かしたときの我々の推定量の値である。その定性的性質を考察する。まず、この手法において絶対値が 2Δ よりも大きなラグを検出することは、 $E[R(T)]$ が常に 0 であるためできない。これはこの手法の技術的限界を示すわけではなく、 X_t, Y_t という系列があったときにも $|\theta| > 2\Delta$ であったとしても $X_t, Y_{t+\alpha}$ について $R(T)$ を計算することで $|\theta - \alpha| < 2\Delta$ とすることで Lead-Lag 効果の検出が可能になるということを示している。これによって $|\theta - \alpha| > 2\Delta$ を満たすような α に関しては全く反応がないので検出の精度を上げることができる。

また、 $|\theta| < \Delta$ に関しては $E[R(T)]$ が単調増加している。これはラグが大きければ大きいほど早く先行する銘柄の動きを捉えることができ、次の Δ 秒後のブロックで売買を行うことによって利益を得ることができることを示している。一方で $|\theta| > \Delta$ に関しては $E[R(T)]$ の絶対値が単調減少している。これはラグが観測間隔よりも大きいときは、検出できたとしてもその売買を行うブロックが遅れた 2 つ遅れたブロックになってしまうことから、ブラウン運動の独立増分性からその先行者から得られる情報を活用できなくなることを示している。

一般的に Δ は秒からミリ秒やナノ秒単位に縮小する傾向があり、この関数は 2 つの正と負の δ 関数を組み合わせたような形になることが考えられる。そのような場合において、 α による探索を行うことでより良い Lead-Lag 効果の推定量を得ることが考えられる。一方で Δ が比較的大きな場合においても絶対値が Δ の中では

指数が単調増加することから良い推定量が得られる事がわかる。

4 実験と考察

本章では提案手法の有効性を示すために、人工データ及び実データによる実験を行う。本章では、先行研究と同様に、Hoffman-Rosenbaum-Yoshida の手法 (HRY) と Dobрева-Schaumburg の手法 (DS) との比較実験を行う [15, 13]。

4.1 人工データの問題設定

実験のために次のような確率微分方程式に従って生成したデータを用いる。

$$\begin{cases} X_t = x_0 \exp\left(\sigma_1 B_t^{(1)}\right) \\ Y_t = y_0 \exp\left(\rho\sigma_2 B_{t-\vartheta}^{(1)} + \sigma_2 (1 - \rho^2)^{1/2} W_{t-\vartheta}\right) \end{cases}.$$

ここで各実験の数値はそれぞれ、観測する時刻 $0 \leq t \leq 10000$ 、各資産の初期値 $x_0 = y_0 = 100$ 、各資産のボラティリティ $\sigma_1 = \sigma_2 = 1.0$ 、Lead-Lag パラメタ $\theta = 10$ 、共分散 $\rho \in \{1.0, 0.9, 0.8, 0.7, 0.6, 0.5\}$ とした。

4.2 Lead-Lag の検出力の実験

本実験では、Lead-Lag を検出できる能力について考察する。非同期観測の幾何ブラウン運動での人工データを用いた実験結果を示す。一般的に時系列の相関係数が低くなっているときの Lead-Lag 効果は測定することが難しいので本実験では相関係数を動かし、それ以外のパラメータは固定した場合での正答率を見る。ここでの共分散 $\rho \in \{1.0, 0.9, 0.8, 0.7, 0.6, 0.5\}$ とした。推定する Lead-Lag はある候補の中から選ばれるがその項は $\{-100, -99, \dots, 99, 100\}$ とした。Lead-Lag を検出できたか否かについては、Lead-Lag の真値である $\theta = 10$ から上下 2 の間である $\{8, 9, 10, 11, 12\}$ のうちのどれかが推定された場合それは正解であるとし、それ以外は不正解とした。実験は 1000 回乱数を発生させてその中で正答率を見るというセットを 1 セットとし、そのセットを 100 セット行う事によって、平均と標準偏差を計算した。

表 1 はその正答率を示す。表から HRY は共分散が 1 以外の場合においては極端に悪い性質を示していることがわかる。また DS はハイパーパラメータが最も良いものをチューニングして使っており、これは同期観測においては我々の手法を $\{0, 1\}$ に直しただけである。故に同様の結果が得られていることがわかる。

表 1: 非同期観測の人工データにおける Lead-Lag 効果推定

Methods	$\rho = 1.0$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.7$	$\rho = 0.6$	$\rho = 0.5$
HRY[17]	0.974 ± 0.004	0.944 ± 0.006	0.939 ± 0.006	0.931 ± 0.006	0.919 ± 0.008	0.898 ± 0.009
DS[13]	1.000 ± 0.000	0.984 ± 0.004	0.964 ± 0.005	0.939 ± 0.007	0.918 ± 0.007	0.887 ± 0.008
提案手法	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	0.999 ± 0.001	0.999 ± 0.001	0.994 ± 0.002

4.3 Lead-Lag の推定における収束速度の比較

本実験では人工データの観測時間を増やすことによって各アルゴリズムのが非同期観測の条件下においてどのような速度で収束するのかを見る。各モデルとパラメータはと次のようにして行う。共分散 $\rho = 0.9$ 観測間隔は平均 10 標準偏差 2 の正規分布に従うとした。推定する Lead-Lag はある候補の中から選ばれるがその項は $\{-100, -99, \dots, 99, 100\}$ とした。本実験においては観測する時刻 $t = 0, 1, \dots, T$, の最大値である T を動かすことによって、各アルゴリズムが収束するまでにどれだけの時間を要するかを調べる。 T に関しては、 $T \in \{10^{2.5}, 10^{2.6}, 10^{2.7}, \dots, 10^{4.9}, 10^{5.0}\}$ を用いる。 T が大きくなるにつれて相対的に観測間隔は小さくなるため、これは HRY の収束の仮定を満たすことにもなる。実験に関しては 1,000 回実験を行った際の推定値の平均絶対誤差をもって評価する。つまり、乱数を $n = 1,000$ 回発生させ、 i 番目の実験において各アルゴリズムによって推定された Lead-Lag を $\hat{\theta}_i$ とし、真値を θ_i とするとき、そのアルゴリズムの平均絶対誤差は $\frac{1}{n} \sum_{i=1}^n |\hat{\theta}_i - \theta_i|$ となる。

図 3はその収束の様子を示すグラフである。図からわかる通り、HRY は収束速度について早いだが、正確な値に収束することはない。DSはそれに比べると良い値に収束するが、それでも収束が遅いことがわかる。一方で提案手法は DS と同様の収束速度を保ちながら、有意に正確な値を推定することができる。

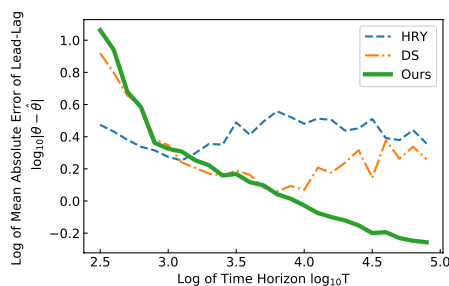


図 3: 各アルゴリズムの収束の比較

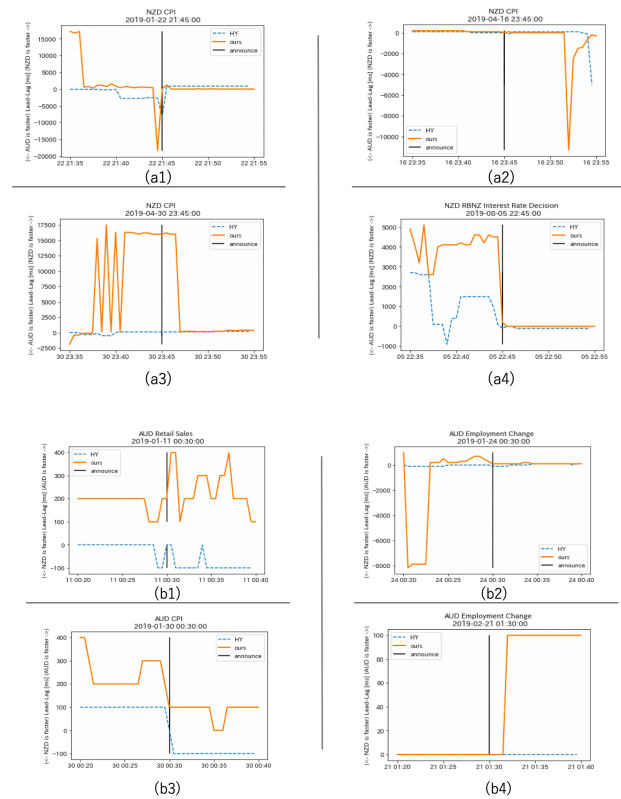


図 4: 重要なマクロ指標発表前の Lead-lag パラメータ。

4.4 実データにおける Lead-Lag の推定

本節では、提案手法が実データにおいても有効であることを示す。本実験では、AUD/USD と NZD/USD という 2 つの地理的に近く似た通貨とされる通貨ペアを用いた。ただし、AUD はオーストラリアドル、USD はアメリカドル、NZD はニュージーランドドルの略称である。我々は、2019 年 1 月から 2019 年 9 月 30 日までのデータを Dukascopy からダウンロードした¹。

また、オーストラリアとニュージーランドの重要な指標発表については、“Economic calendar Investing.com Forex (2011-2019)”²を用いた。図 4.3 が重要な指標発表前後の Lead-Lag パラメータを示している。

¹<https://www.dukascopy.com/>

²<https://www.kaggle.com/devorvant/economic-calendar>

4.5 考察

本論文では、Lead-Lag 効果推定のための指数 NAPLES を提案した。実験 4.2 から、単なる共分散推定量よりも、符合を用いた DS と提案手法がより効果的であることがわかる。実験 4.3 から、定性的に定理 1 を考察が示すように提案手法は Lead-Lag に鋭敏に反応するため、収束が良く精度も良いことがわかる。実験 4.4 からは、以下のような考察が得られる。

- Lead-lag 効果は重要な指標発表前後で変化する。
- NAPLES は HRY よりも早く Lead-Lag 効果を推定できる (a1, a2, b4)。
- NAPLES と HRY が示す Lead-Lag 効果は符合としては同じである。
- NAPLES は HRY が発見できない Lead-Lag 効果を発見できる。 (a3 and b2 and b4)。

5 まとめ

本論文では、高頻度取引における Lead-Lag 効果の推定における 3 つの課題である、(1) タイムスタンプの非同期性、(2) 計算量の膨大性、(3) Lead-Lag 効果の時変性を解決するための Lead-Lag 効果の新しい推定量である NAPLES を提案し、推定法の数学的な性質を与えるとともにその解釈を行った。また人工データ・実データを用いた実験により、提案手法の有効性を示すとともに、マクロ指標の発表前にリードラグの推定を行い、その考察を行った。今後の課題としては、提案手法の統計的性質を明らかにすること、および具体的な裁定取引戦略への応用により収益性を検証することが挙げられる。

参考文献

- [1] Hamad Alsayed and Frank McGroarty. Ultra-high-frequency algorithmic arbitrage across international index futures. *Journal of Forecasting*, Vol. 33, No. 6, pp. 391–408, jun 2014.
- [2] Torben G. Andersen and Tim Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, Vol. 4, No. 2-3, pp. 115–158, June 1997.
- [3] Bruno Biais, Thierry Foucault, and Sophie Moinas. Equilibrium fast trading. *Journal of Financial Economics*, Vol. 116, No. 2, pp. 292–313, May 2015.
- [4] Markus Bibinger, Nikolaus Hautsch, Peter Malec, and Markus Reiss. Estimating the spot covariation of asset prices—statistical theory and empirical evidence. *Journal of Business & Economic Statistics*, Vol. 37, No. 3, pp. 419–435, December 2017.
- [5] Nicolas P.B. Bollen, Michael J. O’Neill, and Robert E. Whaley. Tail wags dog: Intraday price discovery in VIX markets. *Journal of Futures Markets*, Vol. 37, No. 5, pp. 431–451, aug 2016.
- [6] Bradford Case, Yawei Yang, and Yildiray Yildirim. Dynamic correlations among asset classes: REIT and stock returns. *The Journal of Real Estate Finance and Economics*, Vol. 44, No. 3, pp. 298–318, March 2010.
- [7] Andrea Ceron, Luigi Curini, and Stefano M. Iacus. First- and second-level agenda setting in the twittersphere: An application to the italian political debate. *Journal of Information Technology & Politics*, Vol. 13, No. 2, pp. 159–174, mar 2016.
- [8] Alain P. Chaboud, Benjamin Chiquoine, Erik Hjalmarrsson, and Clara Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, Vol. 69, No. 5, pp. 2045–2084, Sep 2014.
- [9] Kalok Chan. A further analysis of the lead-lag relationship between the cash market and stock index futures market. *Review of Financial Studies*, Vol. 5, No. 1, pp. 123–152, January 1992.
- [10] Kohei Chiba. Estimation of the lead-lag parameter between two stochastic processes driven by fractional brownian motions. *Statistical Inference for Stochastic Processes*, Vol. 22, No. 3, pp. 323–357, January 2019.
- [11] Tarun Chordia and Bhaskaran Swaminathan. Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance*, Vol. 55, No. 2, pp. 913–935, 2000.
- [12] Frank de Jong and Theo Nijman. High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance*, Vol. 4, No. 2-3, pp. 259–277, June 1997.
- [13] Dobrislav Dobrova and Ernst Schaumburgb. High-frequency cross-market trading: Model free

measurement and applications, 2017. Working paper.

- [14] Takaki Hayashi and Yuta Koike. Multi-scale analysis of lead-lag relationships in high-frequency financial markets, 2017.
- [15] Takaki Hayashi and Yuta Koike. Wavelet-based methods for high-frequency lead-lag analysis. *SIAM Journal on Financial Mathematics*, Vol. 9, No. 4, pp. 1208–1248, January 2018.
- [16] Takaki Hayashi and Nakahiro Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, Vol. 11, No. 2, pp. 359–379, April 2005.
- [17] M. Hoffmann, M. Rosenbaum, and N. Yoshida. Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli*, Vol. 19, No. 2, pp. 426–461, May 2013.
- [18] Nicolas Huth and Frédéric Abergel. High frequency lead/lag relationships—empirical facts. *Journal of Empirical Finance*, Vol. 26, pp. 41–58, 2014.
- [19] Ira Kawaller, Paul Koch, and Timothy W Koch. The temporal price relationship between s& p 500 futures and the s& p 500 index. *Journal of Finance*, Vol. 42, pp. 1309–29, 02 1987.
- [20] Yuta Koike. Inference for time-varying lead-lag relationships from ultra high frequency data. *SSRN Electronic Journal*, 2017.
- [21] Andrew W. Lo and A. Craig MacKinlay. An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, Vol. 45, No. 1-2, pp. 181–211, July 1990.
- [22] Victor Hugo Martinez and Ioanid Rosu. High frequency traders, news and volatility. *SSRN Electronic Journal*, 2012.
- [23] Sait Ozturk, Michel van der Wel, and Dick van Dijk. Intraday price discovery in fragmented markets. Tinbergen Institute Discussion Paper 14-027/III, Amsterdam and Rotterdam, 2014.
- [24] Bhaskar Sinha and Sumati Sharma. Lead - lag relationship in indian stock market: Empirical evidence. *SSRN Electronic Journal*, 2008.

- [25] Xiaoli Wang. An empirical analysis of lead-lag relationship among various financial markets. *Accounting and Finance Research*, Vol. 4, No. 2, mar 2015.

6 Appendix

次の定理を同期観測の場合に制限すれば直ちに主定理が得られる。

Theorem 2 2つの資産価格 X, Y が幾何ブラウン運動 $dX_t = \sigma_1 dW_t^{(1)}$, $X_0 = x_0$, $dY_t = \sigma_2 dW_t^{(2)}$, $Y_0 = y_0$ に従うとする。ただし、 $W_t^{(1)}, W_{t-\theta}^{(2)}$ は $W_t^{(1)}$ と独立な標準ブラウン運動 $W_t^{(3)}$ を用いて $W_{t-\theta}^{(2)} = \rho W_t^{(1)} + \sqrt{1-\rho^2} W_t^{(3)}$ とあらわされる相関係数が ρ の標準ブラウン運動である。先行遅延関係は θ であらわす。資産価格 X_t を $0 = s_0 < s_1 < s_2 < \dots < s_n$ という時刻で観測し、資産価格 Y_t を $0 = t_0 < t_1 < t_2 < \dots < t_m$ という時刻で非同期的に観測する。これに対して $\bar{t}_i := \min\{t_j \mid s_i \leq t_j\}$, $t_i := \max\{t_j \mid s_i < t_j\}$, $\bar{s}_j := \min\{s_i \mid t_j \leq s_i\}$, $s_j := \max\{s_i \mid s_i < t_j\}$ として、 s_i, t_j に一番近いような t_j, s_i のうち大きいものと小さいものを1つずつ選んで対応付して、 $T = \max\{s_n, t_m\}$ とおく。タイムスタンプについて次のような条件を仮定する。(1) 任意の i について $s_i < t_j < s_{i+1}$ を満たすような t_j は1個または0個しか存在しない。(2) 任意の j について $t_j < s_i < t_{j+1}$ を満たすような s_i は1個または0個しか存在しない。このとき $R(T)$ の期待値は次のようになる。

$$\begin{aligned} E[R(T)] &= \frac{1}{\pi} \sum_{i=1}^{n-1} 1_{s_{i-1} < \bar{t}_i + \theta < s_i} \arcsin\left(\frac{\rho(s_i - \bar{t}_i - \theta)}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}}\right) \\ &+ \frac{1}{\pi} \sum_{j=1}^{m-1} 1_{\bar{s}_j < t_j + \theta < s_{j+1}} \arcsin\left(\frac{\rho(t_j + \theta - \bar{s}_j)}{\sqrt{(s_{j+1} - \bar{s}_j)(t_j - t_{j-1})}}\right) \\ &- \frac{1}{\pi} \sum_{i=1}^{n-1} 1_{s_{i-1} < t_{i+1} + \theta < s_i} \arcsin\left(\frac{\rho(t_{i+1} + \theta - s_{i-1})}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}}\right) \\ &- \frac{1}{\pi} \sum_{j=1}^{m-1} 1_{\bar{s}_j < t_{j-1} + \theta < s_{j+1}} \arcsin\left(\frac{\rho(s_{j+1} - t_{j-1} - \theta)}{\sqrt{(s_{j+1} - \bar{s}_j)(t_j - t_{j-1})}}\right). \end{aligned}$$

Lemma 3 N, M をともに平均が0で分散が1の標準正規分布に従い、相関係数が ρ であるような確率変数であるとする。このとき次が成り立つ

$$P(N > 0, M > 0) = \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} = \frac{\arccos(-\rho)}{2\pi}.$$

Proof 1 (補題の証明) 2つ目の等式は明らかである。故に1つ目の等式を示す。これは *Box-Muller* 変換を用いればわかる。実際 V, W を独立な標準正規分布に従う確率変数とする。 U を $[-\pi, \pi)$ 上一様分布にしたが

う確率変数とする。このとき $R = \sqrt{V^2 + W^2}$ とおき $\phi = \arcsin \rho$ とおく。ここで

$$(N, M) = (\sqrt{2}(V \cos \phi + W \sin \phi), W),$$

$$(V, W) = (R \cos U, R \sin U)$$

が同時分布が同じという意味で成り立つことがわかる。よって以下のように補題が証明される。

$$\begin{aligned} P[N > 0, M > 0] &= P[\cos U \cos \phi + \sin U \sin \phi > 0, \sin U > 0] \\ &= P[U \in (\phi - \pi/2, \pi/2), U \in (0, \pi)] \\ &= \frac{1}{4} + \frac{\phi}{2\pi} = \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi} = \frac{\arccos(-\rho)}{2\pi}. \end{aligned}$$

Proof 2 (定理の証明) 最後の時刻である T まで観測しているので $R(T)$ は以下のようになる。

$$R(T) = \sum_{i=1}^{n-1} (b_{s_i}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i}) - \sum_{j=1}^{m-1} (b_{t_j}^{(Y)} \hat{X}_{t_{j+1}} - b_{t_j}^{(Y)} \hat{X}_{t_j}).$$

まず第 1 項 $\sum_{i=1}^{n-1} (b_{s_i}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i})$ について考える。 $b_{s_{i+1}}^{(X)}$ と $b_{s_i}^{(X)}$ が同符号であるときは $(b_{s_{i+1}}^{(X)} \hat{Y}_{s_{i+2}} - b_{s_{i+1}}^{(X)} \hat{Y}_{s_{i+1}}) + (b_{s_i}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i}) = (b_{s_{i+2}}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i})$ となり、または $b_{s_i}^{(X)}$ が 0 であるときは $(b_{s_{i+1}}^{(X)} \hat{Y}_{s_{i+1}} - b_{s_i}^{(X)} \hat{Y}_{s_i}) = 0$ となるので同符号または 0 となるような s_i を省略することによって、 $b_{s_{i+1}}^{(X)} = -b_{s_i}^{(X)}$ かつ $b_{s_i}^{(X)} \neq 0$ がすべての i に成り立つように s_i を取り直して計算しても $R(t)$ の数値は同じであるので、以下はそのように取り直した s_i を用いる。

また第 1 項は先述の通り

$$\hat{b}_{s_{i+1}}^{(Y)} := (\hat{Y}_{s_{i+1}} - \hat{Y}_{s_i}), \hat{b}_{t_{j+1}}^{(X)} := (\hat{X}_{t_{j+1}} - \hat{X}_{t_j})$$

と定義することによって

$$\sum_{i=1}^{n-1} b_{s_i}^{(X)} (\hat{Y}_{s_{i+1}} - \hat{Y}_{s_i}) = \sum_{i=1}^{n-1} b_{s_i}^{(X)} \hat{b}_{s_{i+1}}^{(Y)}$$

と書き直すことができる。ここで仮定より

$$\hat{b}_{s_{i+1}}^{(Y)} = \text{sign}(Y_{t_{i+1}} - Y_{\bar{t}_i}), \hat{b}_{t_{j+1}}^{(X)} = \text{sign}(Y_{s_{j+1}} - Y_{s_j}),$$

が成り立つので確率変数 $b_{s_i}^{(X)} \hat{b}_{s_{i+1}}^{(Y)}$ はこの 2 つの確率変数が同符号のとき 1、異符号のとき -1 となるような確率変数である。ゆえに、その期待値を計算するときは、同符号・異符号であるときの確率をそれぞれ計算して、

$$\begin{aligned} E[\sum_{i=1}^{n-1} b_{s_i}^{(X)} \hat{b}_{s_{i+1}}^{(Y)}] &= \sum_{i=1}^{n-1} E[b_{s_i}^{(X)} \hat{b}_{s_{i+1}}^{(Y)}] \\ &= \sum_{i=1}^{n-1} P((W_{s_i}^{(1)} - W_{s_{i-1}}^{(1)})(W_{t_{i+1}}^{(2)} - W_{\bar{t}_i}^{(2)}) > 0) \\ &\quad - \sum_{i=1}^{n-1} P((W_{s_i}^{(1)} - W_{s_{i-1}}^{(1)})(W_{t_{i+1}}^{(2)} - W_{\bar{t}_i}^{(2)}) < 0) \\ &= \sum_{i=1}^{n-1} 2P((W_{s_i}^{(1)} - W_{s_{i-1}}^{(1)})(W_{t_{i+1}}^{(2)} - W_{\bar{t}_i}^{(2)}) > 0) + 1 \end{aligned}$$

とすることができる。最後の確率を計算するために

$$p_{s_i, s_{i-1}} := (W_{s_i}^{(1)} - W_{s_{i-1}}^{(1)}), q_{t_{i+1}, \bar{t}_i} := (W_{t_{i+1}}^{(2)} - W_{\bar{t}_i}^{(2)})$$

という確率変数を定めて、

$$P(p_{s_i, s_{i-1}} q_{t_{i+1}, \bar{t}_i} > 0)$$

を計算する。 $W^{(1)}, W^{(2)}$ はブラウン運動であるので $p_{s_i, s_{i-1}}, q_{t_{i+1}, \bar{t}_i}$ はともに正規分布に従う。よってこの確率を計算するためには、補

題 3 から $p_{s_i, s_{i-1}}, q_{t_{i+1}, \bar{t}_i}$ の相関係数を計算すればよい。ここで、 $W^{(1)}, W^{(2)}$ は先行運行関係があるブラウン運動がであったので

$$\begin{aligned} q_{t_{i+1}, \bar{t}_i} &= (W_{t_{i+1}}^{(2)} - W_{\bar{t}_i}^{(2)}) \\ &= \rho(W_{t_{i+1}+\theta}^{(1)} - W_{\bar{t}_i+\theta}^{(1)}) \\ &\quad + \sqrt{1-\rho^2}(W_{t_{i+1}+\theta}^{(3)} - W_{\bar{t}_i+\theta}^{(3)}) \\ &= \rho p_{t_{i+1}+\theta, \bar{t}_i+\theta} + \sqrt{1-\rho^2} r \end{aligned}$$

と書ける。ただし、 $r = (W_{t_{i+1}+\theta}^{(3)} - W_{\bar{t}_i+\theta}^{(3)})$ であり p とは独立な確率変数である。ゆえに求めるべき共分散は共分散の双線形性と r の独立性によって

$$\begin{aligned} &\text{Cov}(p_{s_i, s_{i-1}}, q_{t_{i+1}, \bar{t}_i}) \\ &= \text{Cov}(p_{s_i, s_{i-1}}, \rho p_{t_{i+1}+\theta, \bar{t}_i+\theta} + \sqrt{1-\rho^2} r) \\ &= \rho \text{Cov}(p_{s_i, s_{i-1}}, p_{t_{i+1}+\theta, \bar{t}_i+\theta}) \\ &\quad + \text{Cov}(p_{s_i, s_{i-1}}, \sqrt{1-\rho^2} r) \\ &= \rho \text{Cov}(p_{s_i, s_{i-1}}, p_{t_{i+1}+\theta, \bar{t}_i+\theta}) \\ &= \rho \text{Cov}((W_{s_i}^{(1)} - W_{s_{i-1}}^{(1)}), (W_{t_{i+1}+\theta}^{(1)} - W_{\bar{t}_i+\theta}^{(1)})) \end{aligned}$$

となるので、ブラウン運動の独立増分性から

$$\text{Cov}(p_{s_i, s_{i-1}}, q_{t_{i+1}, \bar{t}_i}) = \begin{cases} \rho(s_i - \bar{t}_i - \theta) & s_{i-1} < \bar{t}_i + \theta < s_i \text{ のとき} \\ \rho(t_{i+1} + \theta - s_{i-1}) & s_{i-1} < t_{i+1} + \theta < s_i \text{ のとき} \\ 0 & \text{それ以外のとき} \end{cases}$$

であることがわかる。ゆえにこれらの確率変数の相関係数は以下のようになる。 $\text{Corr}(p_{s_i, s_{i-1}}, q_{t_{i+1}, \bar{t}_i}) =$

$$\begin{cases} \frac{\rho(s_i - \bar{t}_i - \theta)}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}} & s_{i-1} < \bar{t}_i + \theta < s_i \text{ のとき} \\ \frac{\rho(t_{i+1} + \theta - s_{i-1})}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}} & s_{i-1} < t_{i+1} + \theta < s_i \text{ のとき} \\ 0 & \text{それ以外のとき} \end{cases}$$

ここで $P(p_{s_i, s_{i-1}} q_{t_{i+1}, \bar{t}_i} > 0)$ を計算するために補題 3 を用いる。求めるべきは、 $P(p_{s_i, s_{i-1}} q_{t_{i+1}, \bar{t}_i} > 0) = P(p_{s_i, s_{i-1}} > 0, q_{t_{i+1}, \bar{t}_i} > 0) + P(p_{s_i, s_{i-1}} < 0, q_{t_{i+1}, \bar{t}_i} < 0) = 2P(p_{s_i, s_{i-1}} > 0, q_{t_{i+1}, \bar{t}_i} > 0)$ である。ここで補題から $P(p_{s_i, s_{i-1}} q_{t_{i+1}, \bar{t}_i} > 0) =$

$$\begin{cases} \frac{1}{\pi} \arcsin\left(\frac{\rho(s_i - \bar{t}_i - \theta)}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}}\right) + \frac{1}{2} & s_{i-1} < \bar{t}_i + \theta < s_i \text{ のとき} \\ \frac{1}{\pi} \arcsin\left(\frac{\rho(t_{i+1} + \theta - s_{i-1})}{\sqrt{(t_{i+1} - \bar{t}_i)(s_i - s_{i-1})}}\right) + \frac{1}{2} & s_{i-1} < t_{i+1} + \theta < s_i \text{ のとき} \\ \frac{1}{2} & \text{それ以外のとき} \end{cases}$$

が得られる。 $R(t)$ の第 2 項の

$$\sum_{j=1}^{m-1} (b_{t_j}^{(Y)} \hat{X}_{t_{j+1}} - b_{t_j}^{(Y)} \hat{X}_{t_j}) = \sum_{j=1}^{m-1} b_{t_j}^{(Y)} \hat{b}_{t_{j+1}}^{(X)}$$

についてもこれまでの議論を Y, X を置き換えて同様にブラウン運動の共分散を計算することによって

$$P(p_{s_{j+1}, s_j} q_{t_j, t_{j-1}} > 0) =$$

$$\begin{cases} \frac{1}{\pi} \arcsin\left(\frac{\rho(t_j + \theta - s_j)}{\sqrt{(s_{j+1} - s_j)(t_j - t_{j-1})}}\right) + \frac{1}{2} & s_j < t_j + \theta < s_{j+1} \text{ のとき} \\ \frac{1}{\pi} \arcsin\left(\frac{\rho(s_{j+1} - t_{j-1} - \theta)}{\sqrt{(s_{j+1} - s_j)(t_j - t_{j-1})}}\right) + \frac{1}{2} & s_j < t_{j-1} + \theta < s_{j+1} \text{ のとき} \\ \frac{1}{2} & \text{それ以外のとき} \end{cases}$$

が成り立つことがわかる。これらを足し合わせて定理が証明される。