AMS Special Session on Advances in Mathematical Finance and Optimization I

A Lead-Lag Analysis of Intraday and Overnight Returns

Jiwon Jung
Ph.D. Student
Department of Statistics
Purdue University

PURDUE UNIVERSITY®

March 18, 2023

Collaborators

Prof. Kiseop Lee
Department of Statistics
Purdue University

Prof. Tim S.T. Leung
Department of Applied Mathematics
University of Washington

No Break Time

Figure 1. Stock markets trading hours [1]

- Stock markets with non-overlapping trading hours
- Lagged correlations between intraday and overnight markets [2]

Table of Contents

Introduction

- A lead-lag effect refers to the relationship between two financial assets, where one asset's price movement predicts the price movements of the other.
  - Li, Liu, Wang, *et al.* [3] and Li, Wang, Sun, *et al.* [4] propose a statistically principled definition of the "lead-lag effect."

[1] American Depositary Receipts (ADRs) are certificates that represent shares of non-US companies in US dollars.

Introduction

- A lead-lag effect refers to the relationship between two financial assets, where one asset's price movement predicts the price movements of the other.
  - Li, Liu, Wang, *et al.* [3] and Li, Wang, Sun, *et al.* [4] propose a statistically principled definition of the "lead-lag effect."
- The existing literature investigates a potential correlation between ADR-SPY pairs of non-overlapping market hours.[1]
  - Kang and Leung [2] showed that ADRs' returns are affected by both the US market and the home market.

[1] American Depositary Receipts (ADRs) are certificates that represent shares of non-US companies in US dollars.

Introduction

- A lead-lag effect refers to the relationship between two financial assets, where one asset's price movement predicts the price movements of the other.
  - Li, Liu, Wang, *et al.* [3] and Li, Wang, Sun, *et al.* [4] propose a statistically principled definition of the "lead-lag effect."
- The existing literature investigates a potential correlation between ADR-SPY pairs of non-overlapping market hours.[1]
  - Kang and Leung [2] showed that ADRs' returns are affected by both the US market and the home market.
- Hawkes processes model self-exciting properties in diverse fields:
  - finance: Hawkes [5], Da Fonseca and Zaatour [6]
  - insurance: Swishchuk, Zagst, and Zeller [7], Jung, Lee, and Xu [8]
  - epidemiology: Browning, Sulem, Mengersen, *et al.* [9]

[1] American Depositary Receipts (ADRs) are certificates that represent shares of non-US companies in US dollars.

Goal

- Our goal is to construct a statistical method that effectively quantifies the **lead-lag effect** in multiple aspects, including the **direction**, the **strength**, and the **momentum** in intraday and overnight returns of a pair of stocks.

Data preparation

- We will conduct a lead-lag analysis on two Exchange-Traded Funds (ETFs) that are actively traded in the US market and have non-overlapping home market hours:

[2] https://finance.yahoo.com

Data preparation

- We will conduct a lead-lag analysis on two Exchange-Traded Funds (ETFs) that are actively traded in the US market and have non-overlapping home market hours:
  - **SPY (SPDR S&P 500 ETF Trust)** tracks the performance of 500 large capital U.S. stocks.

[2] https://finance.yahoo.com

Data preparation

- We will conduct a lead-lag analysis on two Exchange-Traded Funds (ETFs) that are actively traded in the US market and have non-overlapping home market hours:
  - **SPY (SPDR S&P 500 ETF Trust)** tracks the performance of 500 large capital U.S. stocks.
  - **FXI (iShares China Large-Cap ETF)** tracks the performance of 50 large capital Chinese equities that trade at the Hong Kong Stock Exchange.

[2] https://finance.yahoo.com

Data preparation

- We will conduct a lead-lag analysis on two Exchange-Traded Funds (ETFs) that are actively traded in the US market and have non-overlapping home market hours:
  - **SPY (SPDR S&P 500 ETF Trust)** tracks the performance of 500 large capital U.S. stocks.
  - **FXI (iShares China Large-Cap ETF)** tracks the performance of 50 large capital Chinese equities that trade at the Hong Kong Stock Exchange.
- The price data are obtained via the Yahoo Finance API[2].
- The time period in our analysis: 2011–2022.

[2] https://finance.yahoo.com

Data preparation

- Daily returns $r(t)$ on day t:

$$r(t) = \frac{P_t^{close} - P_{t-1}^{close}}{P_{t-1}^{close}} \tag{1}$$

- Intraday $r_{ID}(t)$ and overnight returns $r_{ON}(t)$ on day t:

$$r_{ID}(t) = \frac{P_t^{close} - P_t^{open}}{P_t^{open}} \tag{2}$$

$$r_{ON}(t) = \frac{P_t^{open} - P_{t-1}^{close}}{P_{t-1}^{close}} \tag{3}$$

Modeling lead-lag effects of daily returns

$$X_t \rightarrow Y_{t+1}$$

$$X_t = \text{the leader return on day } t$$
$$Y_t = \text{the lagger return on day } t$$

Modeling lead-lag effects of daily returns

$$X_t \to Y_{t+1}$$

$$X_t = \text{the leader return on day } t$$
$$Y_t = \text{the lagger return on day } t$$

- Li, Liu, Wang, *et al.* [3] identifies a lead-lag day as a day satisfying the following criterion for a fixed small $0 < \Delta < 1$:

$$\begin{cases} (1-\Delta)X_t \leq Y_{t+1} \leq (1+\Delta)X_t & \text{if } X_t \geq 0 \\ (1+\Delta)X_t \leq Y_{t+1} \leq (1-\Delta)X_t & \text{if } X_t < 0 \end{cases} \quad (4)$$

  - daily returns
  - the same movement direction
  - multiplicative errors

Modeling lead-lag effects of ID and ON returns

$$X_{2t-1} \rightarrow Y_{2t}$$

$$\begin{aligned} X_{2t-1} &= \text{the leader (SPY) intraday return on day } t \\ Y_{2t} &= \text{the lagger (FXI) overnight return on day } t \end{aligned}$$

Modeling lead-lag effects of ID and ON returns

$$X_{2t-1} \rightarrow Y_{2t}$$

$$\begin{aligned} X_{2t-1} &= \text{the leader (SPY) intraday return on day } t \\ Y_{2t} &= \text{the lagger (FXI) overnight return on day } t \end{aligned}$$

- We further model the lagged correlation by simple linear regression of the two returns:

$$Y_{2t} = \gamma X_{2t-1} + \epsilon_t \tag{5}$$

  - intraday and (following) overnight returns
  - The reversal direction is allowed.
  - additive errors

Modeling lead-lag effects of ID and ON returns

- Using the least squares estimator, we obtain the estimates;

$$\hat{Y}_{2t} = \hat{\gamma} X_{2t-1}, \;\; t = 1, 2, 3, \ldots \tag{6}$$

- Reversal patterns; Strong reversal during the financial crisis periods

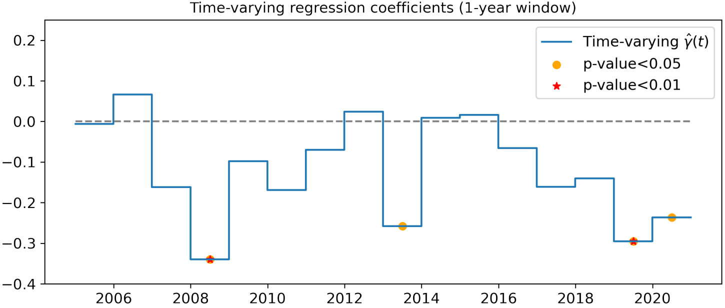

Figure 2. Exploration of the regression coefficient γ

Modeling lead-lag effects of ID and ON returns

Construction of a lead-lag process

We model this lead-lag effect as a discrete-time **counting process** with **direction** and **strength** components:

$$N(t) = \sum_{k=1}^{t} \underbrace{\mathbb{1}(Y_{2k} \cdot \hat{Y}_{2k} > 0)}_{\textbf{Direction}} \underbrace{\mathbb{1}(|Y_{2k} - \hat{Y}_{2k}| < \delta)}_{\textbf{Strength}} \tag{7}$$

for a small fixed $\delta > 0$ and up to date $t = 1, 2, 3, \ldots, T$.

Modeling lead-lag effects of ID and ON returns

Momentum of the lead-lag process

- A stair-case pattern is often observed in some time steps.
- We quantify this cascading effect of $N(t)$ by a Hawkes process

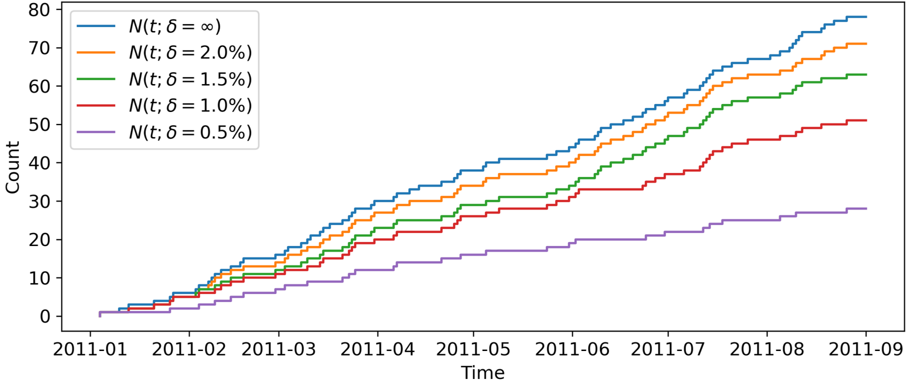

Figure 3. Lead-lag counting processes

Modeling the momentum by a Hawkes process

Hawkes process

- A Hawkes process is a self–exciting counting process.
- An event occurrence increases the probability of the occurrence of another event.

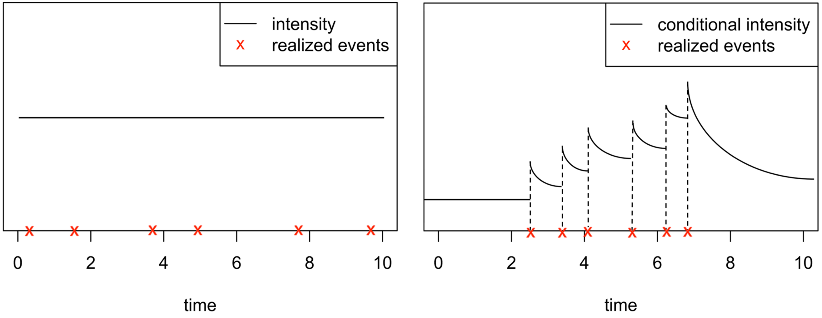

Figure 4. Poisson (left) and Hawkes (right) processes

Modeling the momentum by a Hawkes process

The conditional intensity of the lead-lag process

We use a discrete-time Hawkes process to model the lead-lag intensity:

$$\lambda(t;\mathcal{F}_{t-1}) = E[N(t) - N(t-1)|\mathcal{F}_{t-1}] \quad (8)$$

$$= \underbrace{\mu(t)}_{\text{background intensity}} + \underbrace{\alpha \int_0^t g(s - t_i) dN(s)}_{\text{self-exciting intensity}} \quad (9)$$

where $\mu(t)$ is the background intensity (exogenous effect term) and $g(t)$ is the self-excitation kernel (endogenous effect term).

Modeling the momentum by a Hawkes process

- **Choice of $\mu(t)$**: For simplicity, we choose a piecewise constant function as a background intensity for a given r_{thresh};

$$\mu(t) = \mu + \mu_x I(X_{t-1} \geq r_{thresh}) + \mu_y I(Y_{t-1} \geq r_{thresh}) \tag{10}$$

- **Choice of $g(t)$**: We use a geometric excitation kernel as a discrete counterpart of the exponential kernel [9];

$$g(t - t_i) = \beta(1 - \beta)^{t - t_{i-1}}, \quad t > t_i \tag{11}$$

Modeling the momentum by a Hawkes process

- **Choice of δ:** A small δ fits better to a Hawkes process.

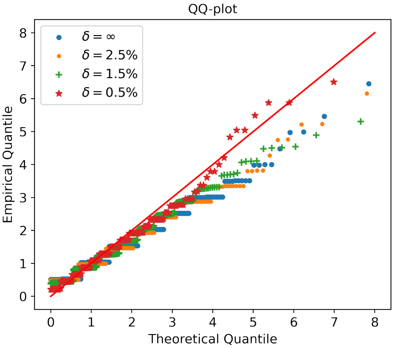

Figure 5. QQ-plot of fitted Hawkes models for different δ. This goodness-of-fit method is based on Chen [10]

Maximum likelihood estimates

- We have a set of parameter $\boldsymbol{\theta} = (\mu, \mu_x, \mu_y, \alpha, \beta)$ to estimate.
- We maximize the log-likelihood of the following form:

$$\log L(\boldsymbol{\theta}) = \sum_{t=1}^{T} \Bigg(\Delta N(t) \log \left(\mu(t) + \alpha \sum_{m=1}^{M} \beta (1-\beta)^{m-1} \Delta N(t-m) \right) - \left(\mu(t) + \alpha \sum_{m=1}^{M} \beta (1-\beta)^{m-1} \Delta N(t-m) \right) \Bigg) + K \tag{12}$$

Maximum likelihood estimates

	$\hat{\mu}$	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\alpha}$	$\hat{\beta}$	$-\log L(\hat{\theta})$
$\delta = 0.5\%$	0.2126	0.0000	0.0000	5.8543	0.0005	1373.4
$\delta = 1.0\%$	0.3291	0.0080	0.0000	1.3626	0.0027	1764.8
$\delta = 1.5\%$	0.4084	0.0214	0.0000	1.5053	0.0032	1970.5
$\delta = 2.0\%$	0.4639	0.0156	0.0000	1.6481	0.0000	2062.9
$\delta = \infty$	0.5074	0.0748	0.0000	1.4858	0.0000	2146.8

Table 1. Maximum likelihood estimates over the 10-year time period (2011–2020)

Conclusion and discussion

- We analyzed the lead-lag effect between intraday and overnight returns of the selected ETFs with non-overlapping trading hours.
- We proposed a statistical method to explore and quantify the lead-lag effect in these asynchronous markets.
- We plan to develop and analyze trading strategies using the lead-lag effects as signals.

References I

[1] P. Henn, "Trading hours: When can you trade stocks, currencies and crypto?" (2022), [Online]. Available: https://currency.com/trading-hours (visited on 03/14/2023).

[2] J. Kang and T. Leung, "Asynchronous ADRs: Overnight vs intraday returns and trading strategies," *Studies in Economics and Finance*, vol. 34, no. 4, pp. 580–596, 2017.

[3] Y. Li, C. Liu, T. Wang, and B. Sun, "Dynamic patterns of daily lead-lag networks in stock markets," *Quantitative Finance*, vol. 21, no. 12, pp. 2055–2068, 2021.

[4] Y. Li, T. Wang, B. Sun, and C. Liu, "Detecting the lead–lag effect in stock markets: Definition, patterns, and investment strategies," *Financial Innovation*, vol. 8, no. 1, p. 51, 2022.

[5] A. G. Hawkes, "Hawkes processes and their applications to finance: A review," *Quantitative Finance*, vol. 18, no. 2, pp. 193–198, 2018.

[6] J. Da Fonseca and R. Zaatour, "Correlation and lead–lag relationships in a hawkes microstructure model," *Journal of Futures Markets*, vol. 37, no. 3, pp. 260–285, 2017.

[7] A. Swishchuk, R. Zagst, and G. Zeller, "Hawkes processes in insurance: Risk model, application to empirical data and optimal investment," *Insurance: Mathematics and Economics*, vol. 101, pp. 107–124, 2021.

References II

[8] J. Jung, K. Lee, and M. Xu, "Modeling multi-state health transitions with hawkes processes," Working paper, 2023.

[9] R. Browning, D. Sulem, K. Mengersen, V. Rivoirard, and J. Rousseau, "Simple discrete-time self-exciting models can describe complex dynamic processes: A case study of covid-19," *PloS one*, vol. 16, no. 4, e0250015, 2021.

[10] Y. Chen, "Goodness-of-fit for fitting real data with hawkes processes,", 2016.

Questions & Answers

- Contact information: jung320@purdue.edu
- Personal webpage: `https://jiwon-jung.github.io/`