DETECTING LEAD-LAG RELATIONSHIPS IN STOCK RETURNS AND PORTFOLIO STRATEGIES*

Álvaro Cartea †‡ **Mihai Cucuringu** *§†¶ **Qi Jin***‡‖

June 9, 2024

Click here for the most recent version

ABSTRACT

We propose a method to detect linear and nonlinear lead-lag relationships in stock returns. Our approach uses pairwise Lévy-area and cross-correlation of returns to rank the assets from leaders to followers. We use the rankings to construct a portfolio that longs or shorts the followers based on the previous returns of the leaders, and the stocks are ranked every time the portfolio is rebalanced. The portfolio also takes an offsetting position on the SPY ETF so that the initial value of the portfolio is zero. Our data spans from 1963 to 2022, and we use an average of over 500 stocks to construct portfolios for each trading day. The annualized returns of our lead-lag portfolios are over 20%, and the returns outperform all lead-lag benchmarks in the literature. There is little overlap between the leaders and the followers we find and those that are reported in previous studies based on market capitalization, volume traded, and intra-industry relationships. Our findings support the slow information diffusion hypothesis; i.e., portfolios rebalanced once a day consistently outperform the bidiurnal, weekly, bi-weekly, tri-weekly, and monthly rebalanced portfolios.

Keywords: Return prediction, Lead-lag relationships, Ranking, Lévy-area, Clustering

JEL classification: G11, G12, G14, G17

---

*We thank Andrew Alden, Torben Andersen, Álvaro Arroyo, Patrick Chang, Fayçal Drissi, Anthony Ledford, Slavi Marinov, Sean Myers (discussant), Roberto Renò, and Harrison Waldon for helpful comments and feedback. We are grateful to audience at Man AHL, J.P. Morgan, GSA Capital, and Oxford Asset Management for comments. We are grateful to audience at the OMI Machine Learning and Financial Econometrics workshop and the Eastern Finance Association Annual Meeting for helpful comments.

†Oxford-Man Institute of Quantitative Finance, University of Oxford

‡Mathematical Institute, University of Oxford

§Department of Statistics, University of Oxford

¶The Alan Turing Institute, London, UK

‖Corresponding author; Email: qi.jin@st-annes.ox.ac.uk

1 Introduction

Changes in stock prices of some firms tend to follow those of other firms. This relationship between stock prices is often referred to as a *lead-lag* relationship. Detecting lead-lag relationships among a large set of stocks is not straightforward. The extant literature uses ad-hoc methods to select leaders and followers, and employs these two sets of stocks in investment strategies to evaluate the economic significance of the lead-lag relationship. For example, Lo and MacKinlay (1990) assume that large market capitalization stocks lead small market capitalization stocks. They build equal-weighted portfolios within each quantile of market capitalizations and use the cross-autocorrelation between the five portfolios to evaluate the trading performance of the lead-lag relationship. Empirical evidence suggests that firm size (Lo and MacKinlay (1990)), trading volume (Chordia and Swaminathan (2000)), institutional ownership (Badrinath et al. (1995)), and other firm characteristics contribute to the lead-lag identity of a stock.

Empirically, however, many lead-lag relationships change over time and often cannot be explained by sorting stocks on a single firm characteristic. [7] This observation motivates that it is necessary to detect, instead of assume and then verify, lead-lag relationships.

Our objective is to find lead-lag relationships without explicitly assuming a link between firm characteristics and lead-lag relationships; instead, we develop a data-driven method that employs stock returns to identify leaders and followers, and we show that the lead-lag relationships we find are economically significant. We achieve this in three steps. First, we design an algorithm that identifies the direction and strength of the lead-lag relationship between the returns of two stocks. Second, we propose a framework that uses state-of-the-art algorithms to rank stocks from leaders to followers based on the pairwise relationships. Third, we construct a zero-cost portfolio to assess the returns predictability of the leaders over the followers, and we measure the economic significance of the portfolio's performance.

Specifically, in the first step we design a method to score the lead-lag relationship between pairs of assets. The sign of the score indicates which of the two assets is more likely the leader, and the magnitude of the score quantifies the strength of the lead-lag relationship. We compute the pairwise score for all combinations of asset pairs, and use the pairwise scores to construct a skew-symmetric matrix, which we refer to as the lead-lag matrix. In this paper, we introduce three scoring methods. Two methods use cross-correlations to capture linear lead-lag relationships; and the third is based on the Lévy-area of pairwise asset returns, which captures both linear and nonlinear lead-lag relationships.

In the second step, we use the lead-lag matrix to sort the stocks from "most likely to be a leader" to "most likely to be a follower". First, we compute the mean of each column of the skew-symmetric lead-lag matrix and order the columns according to their mean, from highest to lowest. A priori, stocks in columns with high means are more likely to be leaders, and those in columns with low means are more likely to be followers.

Finally, after identifying a set of leaders and a set of followers, we use the sign of the previous return of the set of leaders as a signal to buy or to sell an equal-weighted portfolio of the followers. The portfolio is financed by taking an offsetting position on the SPY ETF to finance the portfolio, so the value of the portfolio at inception is zero. We rank the

[7] See Appendix A for motivating examples.

stocks and rebalance the portfolio every day, and compare the portfolio's performance with those of the lead-lag-based portfolios proposed in Lo and MacKinlay (1990), Chordia and Swaminathan (2000), and Hou (2007). For robustness and comparisons, we also study portfolios that are rebalanced at bidiurnal, weekly, bi-weekly, tri-weekly, and monthly frequencies — stocks are ranked at the same frequency as that of the the portfolio rebalances.

In line with the extant literature, our study uses the profitability of the lead-lag portfolio to evaluate the economic significance of the lead-lag relationship between pairs and between groups of stocks. The profitability of the portfolio has two advantages as an evaluation metric. One, portfolio profitability provides a direct comparison with the existing literature. Two, as suggested by Kelly et al. (2022), directly considering portfolio performance is more suitable than using the R^2 of a linear regression. A portfolio can generate significant economic profits even if the predicted R^2 of a regression on returns is negative because the variance of forecasts can heavily influence the predicted R^2.

We use data from 1963 to 2022 to compare the performance of our lead-lag portfolios with those of the portfolios built with the lead-lag relationships identified in the literature. We use only the stocks with large market capitalization to ensure that our portfolios are tradable. Specifically, in each trading day, our analysis includes approximately 550 stocks which are the top 25% largest stocks by market capitalization in NYSE, NASDAQ, and AMEX. Our results show that the portfolios based on the lead-lag relationships we detect outperform those proposed in the extant literature that are based on market capitalization, trading volume, or sector identity. Specifically, the lead-lag relationships we detect lead to portfolios with annualized returns of over 20% and Sharpe ratios above 2. The annualized return and Sharpe ratio of our portfolios are, respectively, three times and approximately twice those of the portfolios proposed in the extant literature. The performance of our portfolios is robust to various choices of parameter values, ranking algorithms, and selection of sets of tradable stocks. In particular, we find that the identity of a stock as a leader or follower changes frequently over time. A lead-lag relationship found at any particular time window rarely remains profitable after 6 months. Our results are economically significant after accounting for industry standard transaction costs and trading frictions. The return of our portfolios cannot be explained by the Fama–French risk factors, illiquidity of stocks, or lead-lag relationships proposed in the literature. In particular, there is little overlap between the leaders and the followers that we identify and those identified by the literature; thus, our method detects fundamentally different lead-lag relationships from those identified in the literature.

In addition, we study one source of lead-lag relationships in stock prices. The literature suggests that asynchronous reactions in stock prices to common information cause the lead-lag relationships; they refer to this conjecture as the hypothesis of slow information diffusion.[8] The sluggishness in reaction to common information depends on various features, including relatively small firm size (Lo and MacKinlay (1990)), relatively low trading volume (Chordia and Swaminathan (2000)), relatively low institutional ownership (Badrinath et al. (1995)), and many other types of market features.[9]

[8] Lo and MacKinlay (1990), Brennan et al. (1993), and Badrinath et al. (1995) argue that lead-lag effects exist because some firms react more slowly to common information than others; Hou (2007) is the earliest paper we found that explicitly calls this the "slow information diffusion hypothesis".

[9] The literature proposes the "limited investor attention" hypothesis to explain why some stocks take longer than others to react to common information.

Our results lend strong support to the slow information diffusion hypothesis. We perform a temporal analysis of our lead-lag portfolios to study the speed of information diffusion. If the slow information diffusion hypothesis is valid, then the economic significance of the lead-lag relationships we identify would gradually diminish as the time gap between identifying the lead-lag relationship and executing the trades to build the portfolio increases. In our study, we vary the rebalancing frequency of the lead-lag portfolios and the frequency of detecting lead-lag relationships from daily to bidiurnal, weekly, bi-weekly, tri-weekly, and monthly. Our results show that as the rebalancing frequency decreases, the performance of the lead-lag portfolios decreases. In the 1960s and early 1970s, the performance of the portfolios with daily rebalancing is indistinguishable from that of the portfolios rebalanced at slower frequencies. However, over time, the daily portfolio gradually outperforms the monthly portfolio, and then the tri-weekly, the bi-weekly, and the weekly rebalanced portfolios from the late 1970s onward.

Our findings contribute to the literature on return predictability, specifically the strand that explores lead-lag relationships among stocks. After Lo and MacKinlay (1990), there has been a growing literature on the lead-lag relationships between two portfolios of stocks. In contrast to our approach, the extant literature does not develop methods to detect which stocks are leaders and which are followers; instead, it assumes which stocks lead and which follow, and uses portfolios to evaluate the lead-lag relationships. Badrinath et al. (1995) study the lead-lag relationship between groups of stocks with higher institutional ownership (i.e., leaders) and lower institutional ownership (i.e., followers), while Chordia and Swaminathan (2000) study the lead-lag relationship where groups of stocks with larger trading volume lead stocks with smaller trading volume. Brennan et al. (1993) verify that stocks with higher investment analyst coverage lead those with lower coverage. Hou (2007) shows lead-lag relationships between large and small market capitalization stocks within each industry, while Frazzini and Cohen (2008) and Menzly and Ozbas (2009) assume lead-lag relationships between groups of stocks with economic links specified by consumer-supplier relationships from financial reports data. More recently, Parsons et al. (2020) assume lead-lag relationships between groups of stocks from different sectors that are co-headquartered, and Huang et al. (2022) show that stocks with frequent, gradual price updates tend to lead those with infrequent, sharp price updates.

We also contribute to a more general literature that uses the cross-section of stock returns to predict asset price movements. For example, DeMiguel et al. (2014) build a vector autoregression (VAR) model that regresses asset returns on the cross-section of previous returns of all other stocks and build arbitrage portfolios. Kelly et al. (2023) provide a theoretical framework on building optimal portfolios based on the cross-section of stock returns, and Yan and Yu (2023) study cross-stock momentum portfolios based on the framework. We contribute to this literature by introducing a framework to rank stocks and then use the returns of the high-ranked stocks to predict the returns of the low-ranked stocks.

In contrast to the strand of finance literature on lead-lag relationships, the literature in network science and graph-based machine learning develops data-driven methods to detect lead-lag relationships and applies them to networks of financial assets. For example, Bennett et al. (2022) use clustering algorithms to classify stocks into several groups and study lead-lag relationships between the stock groups. Li et al. (2021) use the number of days for which the magnitude of returns of one stock is similar to that of the previous return of another stock to construct a network and fit power-law distributions to determine leaders and followers.[10] This strand of the literature uses financial data to test their proposed methods, but does not study the economic principles of lead-lag relationships, see also Shi et al. (2023) and Zhang et al. (2023).

To the best of our knowledge, our proposed methodology is the first data-driven method in the finance literature designed to detect lead-lag relationships and to verify the slow information diffusion hypothesis in stock returns on daily or slower frequencies. However, several other lines of work explore lead-lag relationship detection for time series in fields such as statistics, econometrics, and machine learning. We fill the gap in this literature by designing a data-driven method customized to financial time series. A well-known parametric method in econometrics to detect lead-lag relationships is the Granger Causality test, Granger (1969). Scherbina and Schlusche (2018) use this method to test return lead-lag pairs among individual stocks, and prove pairwise return predictability upon news release, while Basnarkov et al. (2020) use this method to test for lead-lag relationships in foreign exchange markets. Similarly, although not designed for lead-lag relationships, the Sargan–Hansen test can be used to test lead-lag relationships when the link is considered as an instrumental causal relationship, see Sargan (1958). This strand of the literature assumes a linear relationship between the leaders and the followers.

While studies in the lead-lag literature assume linear relationships between leaders and followers, studies in other strands of finance suggest that nonlinear relationships should be considered. Works such as Campbell and Cochrane (1995); Bansal and Yaron (2004); He and Krishnamurthy (2013); DeMiguel et al. (2021), uncover nonlinear relationships in asset returns across various markets, frequencies, and asset classes; in particular, Pohl et al. (2018) suggest that linear approximations to nonlinear models can lead to considerable errors in the model predictions of asset returns. DeMiguel et al. (2021) also show that by accounting for nonlinearity, investors can construct portfolios of funds that incur positive alpha whereas linear methods incur near-zero alpha. Altogether, these works illustrate the need to account for nonlinearity in lead-lag relationships. Our approach is fundamentally different from previous methods in econometrics because we do not assume a linear relationship between the time series of returns, and because our model captures nonlinear dependencies that can be parameterized by piecewise continuous functions.

We use the Lévy-area between each pair of stock returns to find nonlinear lead-lag relationships. We derive that, under mild assumptions, when the return of one stock is modeled as a continuous function of the previous return of another stock, the Lévy-area between the returns of these two stocks reflects the direction and strength of the lead-lag relationship between them. For more detail on Lévy-area in machine learning, data streams, and quantitative finance, see Lyons (2014); Gyurkó et al. (2013); Chevyrev and Kormilitzin (2016).

To the best of our knowledge, the work of Bennett et al. (2022) is the only predecessor that uses Lévy-area of stock return time series for lead-lag relationships; specifically, it applies Lévy-area and Hermitian clustering (Cucuringu et al. (2019)) to form clusters of leaders and followers, and explores network effects and predictability among the clusters

[10]Some other works use methods in network science to build graphs with intra-day data and inspect lead-lag relationships, see Chester Curme and Kenett (2015), Basnarkov et al. (2020), and Buccheri et al. (2019).

they construct. Different from our work, Bennett et al. (2022) do not build on the slow information diffusion hypothesis and do not rank individual stocks from leaders to followers.

In the machine learning and statistics literature, there are data-driven methods for time series lead-lag detection. One of the most common and intuitive methods is to compute lagged correlation or cross-correlation between pairs of time series. Cross-correlation is a class of lagged correlation coefficients between two time series. Numerous works in finance use this method. For example, Chan (1992) studies intra-day lead-lag relationships between the cash market and stock index futures market. In the computer science literature, Sakurai et al. (2005) use lagged correlation to develop the so-called data-stream mining method. For univariate regression models, the R^2 value of a regression is equivalent to the square of the correlation coefficient between the dependent variable and the independent variable after normalizing the variables. Therefore, using cross-correlation as lead-lag detection method is equivalent to conducting univariate linear regressions and using R^2 as a measurement criterion of lead-lag relationships.

On the other hand, works including those of Badrinath et al. (1995); Chordia and Swaminathan (2000); Hou (2007); Frazzini and Cohen (2008); Menzly and Ozbas (2009) use the R^2 of the Granger causality test to verify the direction of the lead-lag relationships they study. While it is easier to establish statistical significance with the Granger causality test, directly using cross-correlation better aligns with the scope of this paper because we do not assume the direction of lead-lag relationships and we wish to explore the dynamics of lead-lag relationships. Another popular method for lead-lag detection is the spectral method in Hause (1971). If two time series are related by a lead-lag relationship, then the cross-spectral density function will exhibit a peak at a certain frequency corresponding to the lag between the two series. This method is most useful when the time series are driven by cyclical or periodic patterns. However, in finance, the spectral method is less useful and not frequently employed for lead-lag detection because stock returns exhibit more complex and often nonlinear relationships.

Finally, in our study of the slow information diffusion hypothesis, we are the first to study lead-lag relationships across various sampling frequencies. Most papers study lead-lag relationships at a monthly frequency (Badrinath et al. (1995), Menzly and Ozbas (2009), Frazzini and Cohen (2008), Parsons et al. (2020), Huang et al. (2022)), and others study weekly lead-lag relationships (Lo and MacKinlay (1990), Chordia and Swaminathan (2000), Hou (2007)). Few papers study daily lead-lag relationships, e.g., Badrinath et al. (1995), Brennan et al. (1993), and Chordia and Swaminathan (2000). Hou (2007) argues that daily data may induce problems caused by non-synchronous updates of price information, which is mainly caused by the low liquidity of some stocks. In our study, we only use relatively large market capitalization stocks; therefore, we minimize instances of non-synchronous price updates. The scarcity of studies employing daily data may be attributed to the prevalence of previous research that employs firm characteristics, which are available only at weekly or monthly intervals. With our approach, one can investigate frequencies from daily to monthly because our method only requires asset prices, which are available at much higher frequencies than that of firm characteristics data.

The remainder of the paper is organized as follows. Section 2 presents the mathematical setup to identify lead-lag relationships and provides the main mathematical models we use. Section 3 describes the data and discusses results.

Section 4 provides robustness checks and considers alternative specifications of numerical experiments. Section 5 concludes and the appendix collects proofs and results.

2 Mathematical Model and Problem Setting

Here, we identify and verify lead-lag relationships in a set of stocks in two steps: identify potential leaders and followers, and construct a lead-lag portfolio to verify the lead-lag relationship.

2.1 Identifying pairwise lead-lag relationships

For each asset n, and for each time $t = 1, 2, \cdots, T$, we standardize the asset returns at time t so that their mean and standard deviation are zero and one, respectively, over the rolling window from $t - 30$ to t. The standardized returns are denoted by $\tilde{R}_{n,t}$, where $n \in 1, 2, \cdots, N_t$. Here, N_t denotes the number of tradable stocks at time t, and the units of t can vary from a day to a month. At each time t, we aim to identify a set of leaders n_l^t and a set of followers n_f^t, such that the returns of the leaders at time $t - \ell$ are employed to predict the returns of the followers at time t; here, ℓ is a positive integer with the same units as those of t.

To demonstrate that the returns of leader stocks predict the returns of follower stocks, we first consider the linear model

$$\tilde{R}_{p,t} = \beta_\ell \, \tilde{R}_{q,t-\ell} + \epsilon_t \, . \tag{1}$$

Here, β_ℓ is the coefficient of the lead-lag relationship where asset q with previous standardized return $\tilde{R}_{q,t-\ell}$ is used to predict the standardized return $\tilde{R}_{p,t}$ of asset p, and ϵ_t is random noise. A sufficient condition for the leader stocks to predict follower stocks is that the coefficient β_ℓ in the regression analysis is statistically significant. We use this regression model as a baseline because β_ℓ is the Pearson correlation between the standardized returns $\tilde{R}_{p,t}$ and $\tilde{R}_{q,t-\ell}$.

We develop a framework for lead-lag detection methods that uses pairwise lead-lag relationships to rank a set of assets from "most likely to be a leader" to "most likely to be a follower". First, consider a pairwise lead-lag detection method that determines which of the two assets p and q is "more-likely" the leader and which is "more-likely" the follower. The method scores each asset pair p, q to obtain at least one of the following: (i) the direction of the lead-lag relationship through the sign of the score, and (ii) the magnitude of the score to determine the strength of the lead-lag relationship. At each time t, we compute all pairwise scores and construct an N_t by N_t lead-lag matrix $\mathbf{C_t}$ where the $i,\ j$ entry is the pairwise lead-lag score for asset i and asset j. Next, we employ four algorithms on the lead-lag matrix $\mathbf{C_t}$ to rank the assets from most likely to be a leader to most likely to be a follower.

In the body of this paper, we use the column averages of the lead-lag matrix $\mathbf{C_t}$ to rank the stocks; recall that each stock is represented by a column in the matrix. High mean values of columns in $\mathbf{C_t}$ are associated with leaders, and low mean values of columns are associated with followers. From the perspective of the slow information diffusion hypothesis, stocks with high column mean values are those for which information arrives sooner, and the stocks with low mean values are those that react slower to information. Appendix D introduces three alternative ranking methods we use in our robustness analysis.

Our framework encompasses the approach used by the benchmarks we employ. For example, to cast Lo and MacKinlay (1990) in our framework, we use the difference in the market capitalization of two companies as lead-lag score, i.e., the entries of the lead-lag matrix $\mathbf{C_t}$, and to rank companies by the column average of the lead-lag matrix $\mathbf{C_t}$.

Here, we describe our lead-lag detection methods and show that they fit into the above framework to identify leaders and followers. First, consider two correlation-based methods for which we define their lead-lag scores as operators on the vector space of standardized asset returns

$$C^1(\tilde{R}_{p,t}, \tilde{R}_{q,t}) = \operatorname*{arg\,max}_{\ell \in [-T,T]} Corr(\tilde{R}_{p,t},\, \tilde{R}_{q,t-\ell})\,, \tag{2}$$

$$C^2(\tilde{R}_{p,t}, \tilde{R}_{q,t}) = \pm\frac{1}{T}\max\left\{\sum_{\ell\in[-T,0]} Corr(\tilde{R}_{p,t},\, \tilde{R}_{q,t-\ell})\,,\, \sum_{\ell\in[0,T]} Corr(\tilde{R}_{p,t},\, \tilde{R}_{q,t-\ell})\right\}. \tag{3}$$

Here, $Corr(\tilde{R}_{p,t},\, \tilde{R}_{q,t-\ell})$ is the Pearson correlation between the returns $\tilde{R}_{p,t}$ and $\tilde{R}_{q,t-\ell}$. Recall that $\tilde{R}_{p,t}$ denotes standardized returns, so their mean is zero and their variance is one; therefore, the sign of the score C^2 determines the direction of the lead-lag relationship and the magnitude of C^2 quantifies the strength of the lead-lag relationship; on the other hand, it is only guaranteed that the sign of the score C^1 determines the direction of the lead-lag relationship. However, C^1 performs equally well as C^2 on simulated data because when asset p leads many other assets, the value of $\sum_q C^1(\tilde{R}_{p,t}, \tilde{R}_{q,t})$ will be large and vice versa. When computing C^1, many of the components in the resulting lead-lag matrix $\mathbf{C}$ will be zero; therefore, C^1 provides a sparser alternative to C^2 when we construct the lead-lag matrix $\mathbf{C}$, which makes C^1 more storage space efficient than C^2.

Next, we use the concept of Lévy-area to detect lead-lag relationships in linear models, e.g., (1), and when there are nonlinear and higher-order lead-lag relationships between the leader and the follower. The concept of Lévy-area is linked to Signatures from rough path theory (see Appendix B for more details).

For a two-dimensional random process $X_t = (X_t^i, X_t^j)$, consider its coordinate iterated integrals over a time interval (s,t), defined as

$$\begin{aligned} S(X)^{i,j}_{s,t} &= \int_{s<a<t} S(X)^i_{s,a}\, dX^j_a \\ &= \int_{s<a<t}\int_{s<b<a} dX^i_b\, dX^j_a\,, \end{aligned} \tag{4}$$

which we use to compute the Lévy-area

$$A^{L\acute{e}vy}_{i,j} = \frac{1}{2}(S(X)^{i,j}_{s,t} - S(X)^{j,i}_{s,t}) \tag{5}$$

between X_t^i and X_t^j. We show, under mild assumptions, that the Lévy-area between normalized returns $\tilde{R}_{n,t}$ and $\tilde{R}_{m,t}$ can detect lead-lag relationships for the more complex model

$$\tilde{R}_{n,t} = \beta_\ell \, f(\tilde{R}_{m,t-\ell}) + \epsilon_t \,. \tag{6}$$

Here, f is any continuous function, and when $f(R) = R$ we obtain the linear model (1).[11]

If we assume that the auto-correlation of a stock's return is very small compared with the variance of the stock's return, and if we assume that all derivatives of f are non-negative, we obtain the following theorem:

Theorem 1. *Assume X_t^i and X_t^j are two independent random processes with zero mean, unit variance, and symmetric distribution, and both satisfy* (6) *over a time interval $[s,t]$. Then, the sign of the Lévy-area $A_{i,j}^{Lévy}$ between X_t^i and X_t^j is the same as the sign of ℓ if and only if $\ell = \pm 1$. In addition, if $\ell = \pm 1$ and the third derivative of the function f exists, there is a constant K such that for all pairs (i,j), we have that $\mathbb{E}|A_{i,j}^{Lévy} - K\beta_\ell| = \frac{M}{6}\,\beta_\ell\,\mathbb{E}[\sum_{s<a<t} f^{'''}(\xi_{a-1}^j)(X_{a-1}^j)^4]$ for some constant M and $|\xi_{a-1}^j| < |X_{a-1}^j|$.*

For a proof, see Appendix C.

Figure 1 shows the cumulative sum of two simulated time series. The simulation is designed so that one series leads and the other one follows.[12] In this example, the Lévy-area correctly identifies the lead-lag relationship while the cross-correlation-based linear methods fail.

For the leading time series $\{x_t\}$ and the following time series $\{f(x_{t-1})\}$, both of length 500, C^1 and C^2 correctly identify which one is the leader with an accuracy of 63.1% and 52.5%, respectively, in 50000 Monte Carlo simulations; in contrast, Lévy-area correctly identifies the lead-lag relationship with an accuracy of 94.5%.

The above theorem shows that one can identify both linear and nonlinear pairwise lead-lag relationships using the Lévy-area between each pair of stocks. Despite the Lévy-area method is only applicable to lags of size 1, one can use multi-step standardized returns instead of one-step standardized returns to circumvent this constraint, thus one can re-scale multi-step standardized returns to set the lag ℓ to 1 and focus on one-step lead-lag relationships. The approach of fixing the lag size is standard in the literature, e.g., Lo and MacKinlay (1990) consider only fixed lag sizes, where the lag size does not change across all pairs of leaders and followers. In addition, the efficacy of the Lévy-area to quantify the strength of the lead-lag relationship depends on the behavior of the function f and the standardized returns of the assets. Specifically, because the returns are standardized to have mean zero, if the third derivative of f is relatively small over the unit disc, then the Lévy-area provides an accurate estimate of the lead-lag strength between the two stocks.

[11]Formally, f is any member of $C^\infty(\mathbb{R}^n)$, which is the Banach space with infinity norm of all continuous functions defined on the space of n-dimensional real vectors.

[12]The leading time series x_t is generated from a normal distribution with mean 0 and standard deviation 1, and the following time series $f(x_{t-1})$ is linked to x_t through the function $f(x_t) = \frac{10}{1+\exp(-x_t)} + \frac{5}{1+\exp(10x_t)} - 7.5$.

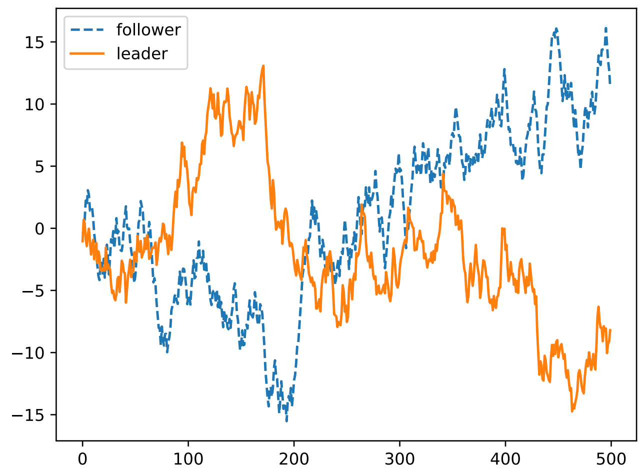


Figure 1: An example of a leading time series and a following time series that cannot be detected by linear methods.

2.2 Constructing Lead-lag portfolios

After sorting the assets from most likely to be a leader to most likely to be a follower, we use the raw returns of the leaders to predict the raw returns of the followers, and we construct a portfolio to evaluate the economic significance of the lead-lag relationship. Consider models (1) and (6), our portfolio compares the predictive power of the leaders on the whole US equity market with the predictive power of the leaders on the followers to assess if there is a significant lead-lag relationship between the leaders and the followers. At each time t, assume the top m stocks represent the set of leaders n_l^t and the bottom n stocks represent the set of followers n_f^t in the sorting, the raw returns of the m stocks at time $t - \ell$ should provide an indication of the raw returns of the n stocks at time t. In this section, we set the lag ℓ to 1 day, see Theorem 1. Furthermore, because we expect the predictability of the m leaders on the n followers to be stronger than that of the m leaders on the US equity market, the sign of the raw returns of the m leaders at time $t - \ell$ should be, statistically, the same as the sign of the difference between the raw return of the n follower stocks and the market return R_{mkt}.

Based on the above intuition, we build a portfolio to evaluate if the lead-lag relationship we detect is economically significant; see Algorithm 1 and Figure 2a. The algorithm uses the previous raw returns of the leaders as a signal to construct an equal-weight portfolio with the followers. Here, R_{t-1}^{Leader} and $R_t^{Follower}$ are the average raw returns of an equal-weighted portfolio with the leaders and the portfolio of the followers at time $t - 1$ and t, respectively. When the average of the raw returns of the leaders is positive, the algorithm buys the followers, and when the average of the raw returns of the portfolio of leaders is negative, the algorithm sells the followers. In step 3 of the algorithm, one trades the market to finance the portfolio. For example, when the average of the raw returns of the leaders is negative at time t, one short sells the followers and buys the market at time $t + 1$. Therefore, one way of interpreting the performance of the portfolio is by how much more the leaders predict the followers than they predict the US equity market.

Algorithm 1 Constructing a lead-lag portfolio

Input: Raw asset returns $R_{1,1:t}, \cdots, R_{N,1:t}$, Raw market return $R_{mkt,1:t}$
Parameter : Look-back window w, number of leaders m, number of followers n; $m + n \leq N$
while $t \geq w$ **do**

1. Compute skew-symmetric matrix $C_{t-w:t}$, whose entries are pairwise lead-lag scores

2. Sort by column average of $C_{t-w:t}$, pick top m assets as leaders, bottom n assets as followers

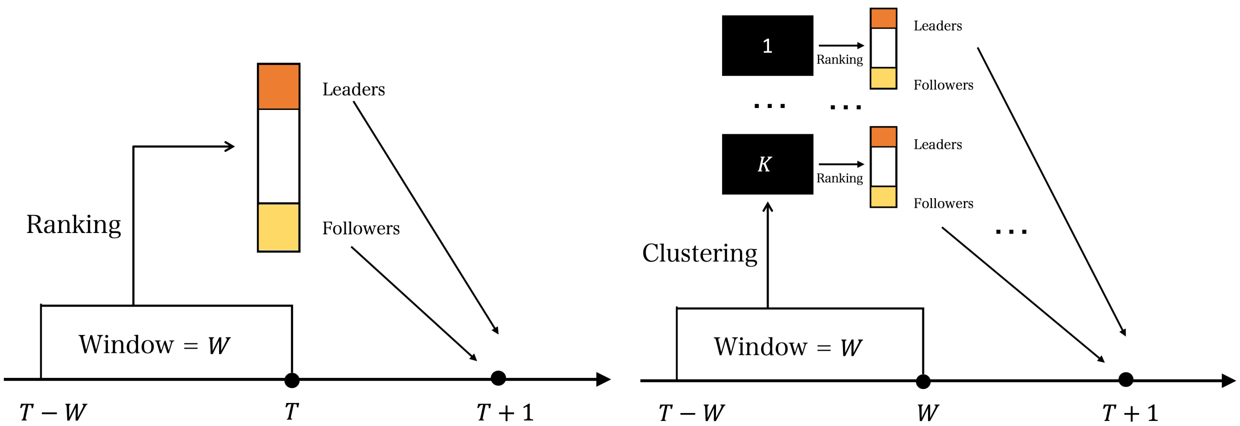
3. Compute R_{t-1}^{Leader} and $R_t^{Follower}$

if $R_{t-1}^{Leader} \geq 0$ **then**
  $R_t^{Portfolio} = R_t^{Follower} - R_{mkt,t}$
else
  $R_t^{Portfolio} = R_{mkt,t} - R_t^{Follower}$

An alternative to this approach is to trade the spread between the top and bottom quantiles of stocks after sorting, i.e., to finance the portfolio with the leaders instead of the market. For example, if at time t the average of the raw return of the portfolio of leaders is negative, one short sells the followers and buys the leaders at time $t + 1$. This alternative approach assumes that the predictability of the leader on the follower is greater than the autocorrelation of the leader. Compared with this alternative, our approach offers two primary advantages. First, the performance of our portfolio is interpretable; it measures the ability of the portfolio of leaders to predict the portfolio of followers instead of predicting the market. Second, it is more feasible to trade exchange-traded funds such as the SPY to finance the lead-lag portfolio. In particular, when short selling, it is more feasible to short sell the SPY ETF than an entire portfolio of leaders. In the context of financial time series; our approach to consider the performance of the portfolio is more suitable than those in the literature that apply linear regression models and observe R^2 values because the variance of forecasts can heavily influence the predictive R^2, see Kelly et al. (2022).

Some studies use characteristics including firm sector or the location of a firm's headquarter to group stocks into blocks. Next, they sort the stocks into leaders and followers within each of these blocks and construct lead-lag portfolios within each block of stocks as above, see, e.g., Hou (2007). To compare our approach with those in the papers that classify stocks into groups, we extend our methodology to include data-driven algorithms that partition stocks into groups. Specifically, we use state-of-the-art data-driven clustering algorithms, including spectral clustering (Ng et al. (2001)) and Hermitian clustering (Cucuringu et al. (2019)) that use information in the lead-lag matrix $\mathbf{C}$ to group stocks. Financially, these algorithms rearrange assets into groups via pairwise characteristics. We apply these clustering algorithms to the skew-symmetric lead-lag matrix $\mathbf{C}$ obtained with Algorithm 1, and employ the resulting clusters to construct our portfolios. This extension to our approach allows meaningful comparisons with the lead-lag relationships identified in the literature, and the approach enables comparisons between lead-lag relationships we detect and the relationship induced by pre-defined clusters such as sectors. See Figure 2b for a visual illustration of our clustering method.

From now on, we refer to the portfolio illustrated in Figure 2a as the "global" lead-lag portfolio, and we refer to portfolios illustrated in Figure 2b as the "clustered" lead-lag portfolio. In the global lead-lag portfolio, we rank the assets into most likely leaders and most likely followers, and we use the returns of the leaders to predict the returns of the followers. In the clustered portfolio, we first cluster the stocks into K blocks and construct lead-lag portfolios within each block as in Algorithm 1. Initially, we assume zero execution costs, and in Section 4.1 we include trading fees and frictions.



(a) Global portfolio

(b) Clustered portfolio

Figure 2: Global portfolio: Assets are ranked into most likely leaders to most likely followers, and the returns of the leaders are used to predict the returns of the followers, see Algorithm 1. Clustered portfolio: Assets are clustered into K blocks, and K lead-lag portfolios are constructed within each block as in Algorithm 1.

3 Empirical Results

3.1 Data

Asset prices and firm characteristics data are from the Center of Research in Security Prices (CRSP) daily returns database. The sample period is from July 1963 to December 2022. To be consistent with standard literature practices, we include NYSE, Amex, and NASDAQ stocks. For each trading day, to ensure that the trading positions we take are feasible, we use only the stocks in the top 25 percentile of market capitalization, which is given by the product of the price of the stock at the end of the day and the number of shares outstanding.

For all frequencies (e.g., bidiurnal, weekly, bi-weekly, monthly, etc.), returns are computed from daily returns as the cumulative product of one plus the end-of-day return of the assets.[13] For example, for a calendar week with five trading days and raw returns $R_1, R_2, \cdots, R_5$, the raw return of the week is $R_{week} = \prod_{i=1}^{5}(1 + R_i) - 1$, and the raw return of the week is standardized to compute the Lévy-area and cross-correlation between pairs of stock returns.

[13] Daily return for stock n at day t is computed by purchasing the stock on the most recent time previous to t when the stock had a valid price, see CRSP.

To study the composition of portfolios, we also include industry classification data. The industry classification information links each firm to a single, non-overlapping Fama–French 12 industries by its SIC code. The industries are (1) nondurables, (2) durables, (3) manufacturing, (4) energy, (5) chemicals, (6) business equipment, (7) telecommunications, (8) utilities, (9) shops, (10) healthcare, (11) finance, and (12) other.

3.1.1 Summary Statistics

Tables 11 to 15 in Appendix E provide summary statistics for the set of assets used in this paper, broken down by decade and by Fama–French sector labels for stocks. Table 11 shows the average number of stocks that are included in our study from 1963 to 2022, split by decade and industry. Recall that on each trading day we construct the portfolio with the stocks in the top 25 percentile by market capitalization. In addition, to compute lagged correlation and Lévy-area, we use the stocks that do not have missing returns in the previous 60 trading days. From left to right in the table, we see an increasing trend in the number of stocks included in our study, with an average of 143 in the 1960s to 593 in the 2010s. However, the increase in number of stocks is not monotonic and stocks are distributed unequally across industries. For nondurables, durables, manufacturing, and utilities, we observe a monotonic decrease in the number of stocks between the 1990s and the 2010s. This is likely correlated to the evolution of the global economy which caused the number of companies in these four industries to decline in the US. On the other hand, for business equipments, telecommunications, shops, healthcare, and finance stocks, we observe a peak in the number of stocks in the 2000s. We also observe the same trend in the total number of stocks. The count of stocks in several industries peaks in the 2000s and declines in the 2010s; this observation coincides with the 2008 financial crisis.

Table 12 shows the mean and standard deviation of stock prices, and Table 13 shows the mean and standard deviation of stock returns for the stocks we use to build portfolios. Overall, we observe a strong upward trend in the mean and the volatility of stock prices over time, but the average daily return and volatility of stock returns do not undergo considerable changes. The mean and volatility of stocks prices in the finance sector are much higher than those in other sectors. This is primarily caused by the high unit price of Berkshire Hathaway (BRK.A), which lists at approximately 500,000 dollars per share at the time this paper is written.[14] Meanwhile, for stock returns, we observe an increase in volatility and decrease in mean returns during the 2000s. This is likely correlated to the dotcom bubble and the 2008 financial crisis. We observe this trend across all Fama–French industries.

Tables 14 and 15 show the mean and the standard deviation of the volume and of the turnover ratio for the stocks, respectively. There is a clear upward trend in both volume traded and turnover ratio (defined as the quotient of the daily volume of shares traded to the shares outstanding) for stocks, indicating an increase in trading activity. In particular, average daily volume for each stock increased by over five times from the 1970s to the 1980s, over three times from the 1980s to the 1990s, and over five times from the 1990s to the 2000s. Similarly, the turnover ratio of stocks nearly doubles between 1970s and 1980s, 1980s and 1990s, and 1990s and 2000s. This is likely correlated to the emergence

[14]The mean price of stocks in the finance sector is similar to that of other sectors after removing BRK.A.

of algorithmic trading in the 1980s and the 1990s, and to the development of high-frequency trading in the 2000s.[15] Another possible reason for the change in trading volume and turnover ratio is the decimalisation of the US market in 2001, which reduced spreads and thus lowered execution costs, see Bessembinder (2003). On the other hand, the average turnover of stocks does not undergo large changes from the 2000s to the 2010s (13.6 and 13.2 respectively). Hence, the increase in share volume traded from the 2000s to the 2010s is driven by the increase in shares outstanding. The observed trends in market liquidity are consistent across all industries.

3.2 Evidence of Daily Lead-lag Relationship

In this section, we build lead-lag portfolios to investigate lead-lag relationships. We rank the stocks and rebalance the portfolios at a daily frequency because this is the most granular frequency at with one can compare our results with the benchmarks. The benchmark portfolios use firm characteristics that are available at daily frequencies, e.g., market capitalization and trading turnover. We employ Algorithm 1 to construct the portfolios and provide empirical evidence that the data-driven lead-lag portfolios outperform the baseline lead-lag portfolios.

For our data-driven lead-lag portfolio, the Lévy-area and cross-correlation between pairwise stock returns are computed over a 60-day lookback window; the stocks we include on each trading day are the top 20% (i.e., leaders) and bottom 20% (i.e., followers); and the market return is that of the SPY ETF. Recall that we consider the top quantile of all stocks in terms of market capitalization to construct our portfolio. Appendix F presents the results for alternative choices of look-back windows and various choices of the number of leaders and followers.[16] Additionally, when comparing our portfolios with those in the literature that group stocks into small blocks, we choose spectral clustering to form 12 clusters, which is the same number of Fama–French industries. Appendix F presents the results for alternative choices of clustering methods.

To compare our findings with those in Lo and MacKinlay (1990), Chordia and Swaminathan (2000), and Hou (2007), which are our main benchmarks, we adapt their lead-lag portfolio construction procedures to align with that in Algorithm 1. Specifically, for Lo and MacKinlay (1990), we sort stocks by market capitalization (i.e., the product of the shares outstanding on the previous trading day and the previous day's price), and we use the top 20% stocks as leaders and the bottom 20% stocks as followers to construct the portfolio. We use the SPY ETF to finance all portfolios, as in our data-driven lead-lag portfolios. For Chordia and Swaminathan (2000), we use the same steps as above, except that instead of market capitalization, we use turnover to rank the stocks from leaders to followers. Similarly, for Hou (2007), we group stocks into their respective Fama–French industries and sort them by market capitalization within each group. Next, we use the approach of Lo and MacKinlay (1990) to construct and finance lead-lag portfolios for each industry; therefore, on each trading day, the lead-lag portfolio for Hou (2007) consists of 12 portfolios with zero net initial capital

[15] Program trading became widely used in the 80s in trading between the SP 500 equity and futures markets (McGowan (2010)). Electronic trading venues known as electronic communications networks (ECNs) emerged in the late 1990s and allowed trading of financial products outside of the traditional stock exchanges (McGowan (2010)). In the US, market share of high-frequency trading increased from less than 20% in 2005 to around 60% in 2009 (Zaharudin et al. (2022)).

[16] Most of the literature splits stocks into quantiles. They use the top quantile as leaders and bottom quantile as followers. In Appendix F, we use the top 40% of the stocks as leaders and the bottom 40% of stocks as followers.

(i.e., hold long and short positions of the same dollar value on assets in each portfolio initially). Appendix F provides results for alternative choices of the percentage of stocks that are considered leaders and followers, and Appendix D provides the results for alternative choices of ranking algorithms.

Table 1: Performance of various lead-lag portfolios

Panel A: Global Lead-Lag Portfolios					
	Compound Return (%)	Return (bps/day)	Daily Vol. (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	19.69	7.14	0.47	2.37	16.90
Avg Cross-Cor	27.97	9.79	0.70	2.21	28.64
Lévy-area	24.87	8.82	0.59	2.38	24.67
Market Cap	6.15	2.37	0.50	0.75	63.40
Turnover	7.60	2.91	0.34	1.37	13.62
Panel B: Clustered Lead-lag Portfolios					
	Compound Return (%)	Return (bps/day)	Daily Vol. (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	14.52	5.39	0.38	2.23	18.21
Avg Cross-Cor	19.01	6.91	0.49	2.23	23.49
Lévy-area	17.50	6.40	0.42	2.43	12.28
Industry	15.75	5.81	0.43	2.18	42.29

Table 1 shows the main results for various lead-lag portfolios. Panel A reports the performance of the data-driven lead-lag portfolios compared with those of Lo and MacKinlay (1990) and Chordia and Swaminathan (2000), denoted by "Market Cap" and "Turnover", respectively. Panel B shows the comparison between clustered lead-lag portfolios from the data-driven methods and the portfolio from Hou (2007), which is denoted "Industry". In both panels, we evaluate the portfolios with five criteria: annualized return, daily return, daily volatility, portfolio Sharpe ratio, and maximum drawdown of the portfolio. The return of the portfolio reflects the predictive power of the lead-lag relationship; volatility and Sharpe ratio show the risks of lead-lag portfolios; and maximum drawdown indicates when lead-lag relationships temporarily break down.

The data-driven lead-lag portfolios outperform the benchmarks on almost all categories except for volatility, see Panel A in Table 1. The daily returns for the portfolios range from 2.37 basis points (bps) per day for the Market Cap portfolio to over 9 bps per day for the average cross correlation portfolio. The Sharpe ratios of our data-driven portfolios are all above two and the annualized returns are above 20%. In particular, the annualized return of the Lévy-area portfolio is 24.87% over the span of 60 years. Additionally, we compute criteria that are sensitive to risk to understand further the performance of the portfolios. The Sharpe ratios of the data-driven portfolios are higher than those of the two benchmark portfolios based on firm-characteristic-driven lead-lag identifications. This shows that the methods that rank stocks based on pairwise lead-lag relationships capture economically significant lead-lag relationships. In particular, the Sharpe ratio of the Lévy-area portfolio is the highest, and the annualized return of the Lévy-area portfolio is the second highest among our three data-driven methods reported in Panel A of Table 1. In Appendix F we show that the Sharpe ratio and annualized return of the Lévy-area portfolio are also the highest when we propose an alternative choice of proportion of leaders and followers to construct the portfolios. Overall, our results show that the Lévy-area portfolio outperforms the cross-correlation-based methods.

Panel B of Table 1 compares the performances of various clustered lead-lag portfolios. The performances of the Max Cross-Cor, Avg Cross-Cor, Lévy-area, and Industry portfolios are economically significant, with returns ranging from 5.39 to 6.91 bps/day. The data-driven portfolios Max Cross-Cor, Avg Cross-Cor, and Lévy-area, outperform the Industry portfolio for most criteria, including annualized return and Sharpe ratio. In particular, compounded return of the Lévy-area portfolio is 17.5%, with a relatively low volatility of 0.42%. In terms of criteria that account for risk exposures, the data-driven portfolios also perform better than the baseline Industry portfolio, with higher Sharpe ratios and lower historical maximum drawdowns. In particular, the Sharpe ratio for the Lévy-area portfolio is 2.43, which is at least 0.2 higher than that of any other portfolio. This suggests that the data-driven methods used to construct the lead-lag portfolios capture economically significant lead-lag relationships, with the Lévy-area method being particularly effective.

3.3 Composition of Lead-lag Relationship

In this subsection, we study the composition of the data-driven lead-lag portfolios. We employ a three-fold approach to analyze the characteristics of leaders and followers and to compare the composition of the lead-lag portfolios with that of the benchmarks. First, we use the Jaccard score and the overlap coefficient to compute the similarity between data-driven portfolios and the benchmark portfolios. Second, we use the Adjusted Rand Index (ARI) to compute the similarity between the clusters formed for the data-driven portfolios and the underlying Fama–French sector labels of the stocks, see Hubert (1985). Third, we use a permutation test to assess if, for the leaders and followers, there is a statistically significant difference between the average market capitalization, volume, turnover, unit price, number of shares outstanding, and return. The goal of these three analyses is to understand and quantify the overlap between the constituents of the data-driven lead-lag portfolios found in our study and those reported in the literature.

In the first analysis, for each day and for each strategy, we compute the pairwise Jaccard score and the overlap coefficient for the stocks traded by each strategy (i.e., the followers identified by each trading strategy). Figure 3 shows the average Jaccard score and Overlap coefficient. In the comparison, we include portfolios constructed based on Lo and MacKinlay (1990) and Chordia and Swaminathan (2000); recall they are denoted MktCap and Turnover, respectively, and our three data-driven methods are denoted Lévy, MaxCor, and AvgCor. To compare the similarity between our data-driven strategies and the benchmark, we also include a strategy which trades the union of stocks traded in MktCap and Turnover, which is denoted "Combined".

The Jaccard coefficient between the Combined strategy and the data-driven strategies is approximately 15%, and the respective overlap coefficients are around 35%, which shows that approximately 35% of the stocks traded in any one of

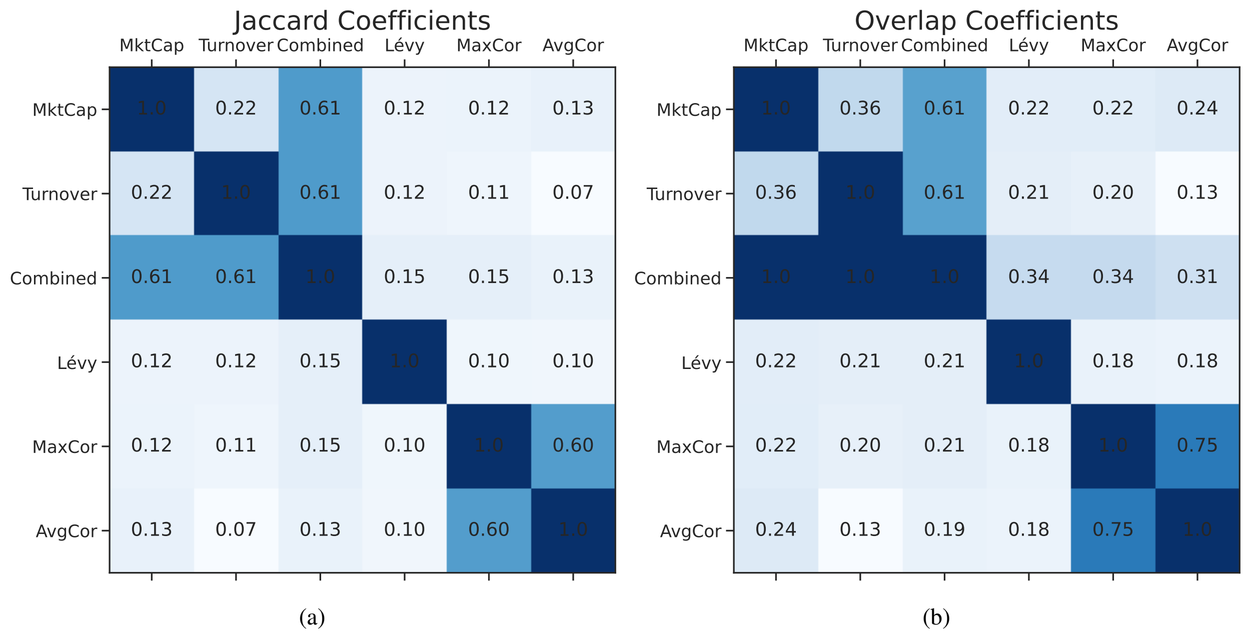


Figure 3: (a) Jaccard coefficients among lead-lag portfolios; a higher value of the Jaccard coefficient represents higher similarity (b) Overlap coefficients among lead-lag portfolios; a higher value of the overlap coefficient represents higher similarity.

the data-driven strategies is also traded by at least one of the baseline strategies that uses firm characteristics to identify leaders and followers, see Figure 3. A Jaccard score of 60% and an overlap coefficient of 75% between the MaxCor and the AvgCor portfolios show that these two data-driven benchmarks are similar. However, the low similarity between the correlation-based portfolios and the Lévy-area portfolio shows that Lévy-area, which detects both linear and nonlinear lead-lag relationships, constructs a considerably different portfolio when compared with all other portfolios, both the data-driven and those proposed by the literature.

Next, we use spectral clustering and Hermitian clustering to compute the ARI between clusters formed with the data-driven portfolios and the underlying Fama–French industry labels. By doing this, we compare the clusters that we find with the Fama–French industry labels, used by Hou (2007), and therefore we study the overlap between our portfolios and those of Hou (2007). We split time from 1963 to 2003 into sub-periods and group the stocks into clusters to compare the clusters with the underlying stock sectors. First, we split time into 1000-day windows and form a new matrix with the sum of the lead-lag matrices in each window. In this step, we include stocks that are both in the top quantile in terms of market capitalization and are actively traded over the previous 1000 days. Next, we use spectral clustering and Hermitian clustering to split the new matrix into 12 clusters and use the ARI to compute the similarity between the formed clusters and the underlying sector labels. Panel A of Table 2 presents the results.

Panel A in Table 2 shows that the similarity between the clusters for lead-lag portfolios and the Fama–French industry labels is low, with the exception of the Hermitian clusters for the Lévy-area portfolio. The data-driven portfolios arrange the stocks into groups that are almost unrelated to the industry labels; however, these data-driven clusters show stronger lead-lag relationships within each cluster than those within the Fama–French industries.

Table 2: Permutation test and adjusted Rand index analysis on Lead-lag portfolios

Panel A: Average adjusted Rand index between lead-lag portfolios and the Fama–French industry labels

	Average ARI Spectral (%)	Average ARI Hermitian (%)
Max Cross-Cor	1.00	5.40
Avg Cross-Cor	3.17	3.19
Lévy-area	7.71	20.34

Panel B: Characteristics for leaders and followers in the Lévy-area portfolio

	Leader Avg Percentile (%)	Follower Avg Percentile (%)	Permutation test p-value
Market Cap	54.73	56.15	0.031
Volume	50.53	53.25	< 0.001
Turnover	38.82	41.28	0.025
Return	50.07	50.44	< 0.001
Price	56.65	55.31	0.103
Shares Outstanding	63.28	66.35	< 0.001

The Lévy-area portfolio with Hermitian clustering is the only case in panel A, Table 2, that is similar to the Fama–French 12 sectors with an ARI of approximately 20%. Figure 4 provides a visualization of clusters in the Lévy-area portfolio for the 1000-day window from 2019 to 2022. In the figure, each colored short dash represents a stock where the horizontal position represents cluster membership, and the vertical position represents sector membership. The black vertical lines represent the boundaries between each cluster.

Figure 4 shows that Hermitian clustering detects six clusters that are quite similar to the underlying sector labels. The six clusters are mostly Energy stocks, Finance stocks, Business Equipments, Shops, Utilities, and Healthcare stocks, respectively. However, it is difficult to determine if the remaining six clusters and the sectors are similar.

The third analysis contains the permutation tests on the Lévy-area portfolio; see results in Panel B, Table 2. On each trading day, we compute the percentile score for each stock's average Lévy-area, market capitalization, turnover, volume, price, number of shares outstanding, and return for all stocks traded on that day. Next, we take the average of these percentile scores for each stock over its respective lifespan of active trading and sort all stocks that are ever traded (approximately 5000) by their average percentile of their Lévy-area. We take the top quantile as leaders and the bottom quantile as followers, and we use the permutation test to verify if there is a statistically significant difference between the mean of market capitalization, turnover, volume, and return of the leaders and the followers. Panel B in Table 2 shows that there is a statistically significant difference in average market capitalization, share volume traded, turnover, and return between the historical leader stocks and the follower stocks.

On average, stocks with a historical propensity to lead tend to have larger market capitalization, larger share volume traded, larger turnover, and high daily returns compared with stocks that are historically more often identified as

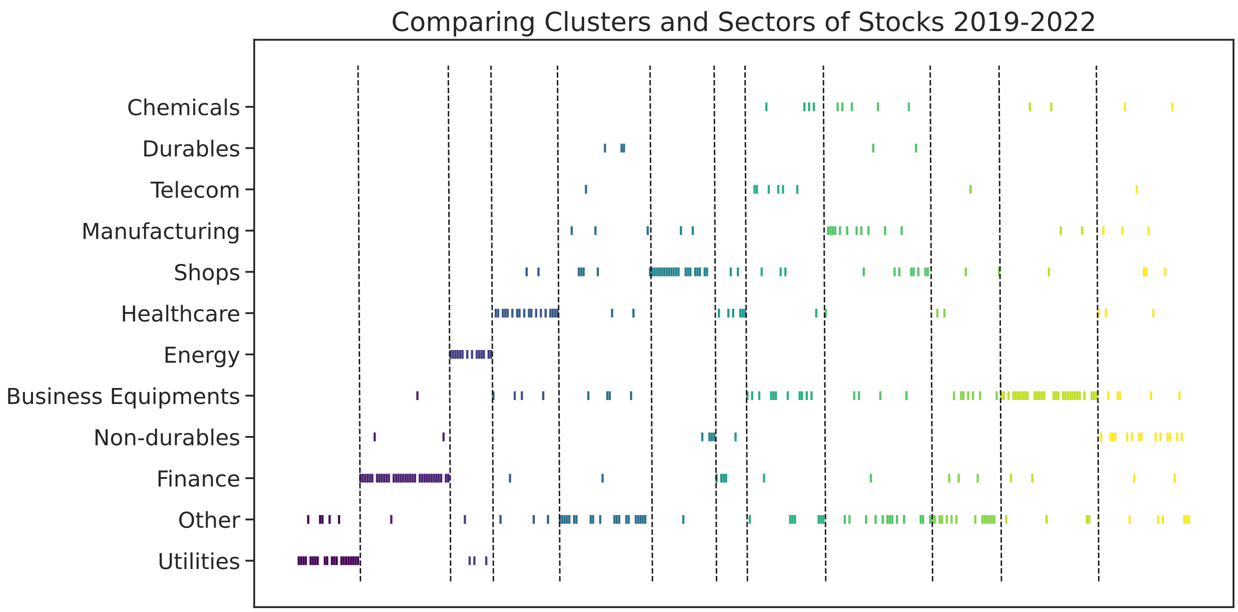


Figure 4: Clusters formed by Hermitian clustering on the Lévy-area matrices from 1 January 2019 to 31 December 2022. The area between black vertical dashes represents each cluster formed with Hermitian clustering. During this period, 376 stocks are traded.

followers. This difference is significant at the 5% confidence level for market capitalization and turnover, and it is significant at the 0.1% confidence level for volume and return. The differences in market capitalization, volume, and turnover of the leaders and the followers coincide with results in the literature; however, the extant literature does not discuss the differences in the returns for the leaders and followers. The discussion of why the leaders have higher historical average returns is beyond the scope of this paper and we defer this to future work. Results for alternative choices of portfolios are included in Appendix G.

3.4 Sector lead-lag characteristics

In this section, we study if some sectors tend to contain more leaders and others contain more followers. Specifically, on each day, we use the ranking we infer from the lead-lag matrix to compute the percentile rank of each stock on each trading day. Then, we compute the average percentile rank for each Fama–French sector over time to understand the composition of each sector. We also inspect how the sector lead-lag identities evolve over time to understand if there is a change in the dynamics of sector-wise lead-lag relationships from 1963 to 2022.

Table 3 presents the average percentile ranks of stocks in various sectors for different lead-lag matrix constructions from 1963 to 2022. The average percentile rank provides insights into the leader or follower status of each sector, with a lower rank indicating a higher proportion of leaders and a higher rank indicating a larger proportion of followers.

The results highlight that the inter-industry lead-lag relationships are predominantly significant for the lead-lag relationship detected by the Lévy-area method. Sectors such as Manufacturing, Chemicals, and Energy exhibit much

Table 3: Average percentile rank of stocks in each sector for various lead-lag matrix constructions

	Lévy-area	MaxCor	AvgCor
Manufacturing	47.6	50.1	50.1
Chemicals	48.5	50.4	50.4
Energy	49.4	50.4	50.4
Telecom	49.5	50.4	50.2
Healthcare	49.6	49.9	50.0
Utilities	49.8	50.7	50.5
Non-durables	49.9	50.2	50.1
Shops	50.1	50.6	50.5
Finance	50.3	50.0	50.0
Other	50.4	49.8	49.8
Durables	51.1	50.1	50.2
Business equipments	55.4	50.1	50.0

lower average percentile ranks, suggesting a higher proportion of leader stocks in these sectors. This coincides with the observation in Frazzini and Cohen (2008) that lead-lag relationships exist between customer-supplier linked firms with the supplier firms being leaders. On the other hand, the Business Equipments sector stands out with a notably higher average percentile rank, indicating a larger number of follower stocks within the sector. This implies that the stocks in Business Equipments tend to lag behind the other sectors. Overall, the Lévy-area method appears to detect inter-industry lead-lag relationships and finds sectors that tend to persistently contain leaders or followers.

In contrast, the cross-correlation-based lead-lag matrix constructions, including MaxCor and AvgCor, do not show significant differences in percentile rank across sectors. This suggests that the follower and the leader dynamics found by these methods are less persistent across sectors. The sectors demonstrate relatively similar average percentile ranks, indicating a balanced mix of leader and follower stocks within these sectors. This result likely shows that the cross-correlation-based methods do not account for the inter-industry lead-lag effects found by the Lévy-area method.

To further study the inter-industry lead-lag relationships detected by the Lévy-area method, we perform another experiment where we first compute the average lead-lag percentile rank of stocks in each sector for every year from 1963 to 2022 and then sort the sectors by their average lead-lag percentile rank from smallest to largest and record the rank of each sector, see Figure 5. We observe a significant shift in lead-lag identity across sectors over time. Business Equipments consistently lags other sectors from 1963 to around 2000, but does not rank at the bottom from 2000 onward; Finance exhibits strong leadership from 2017 onwards, but oscillates between leading and lagging prior to 2017. Utilities gradually shifts from a consistent leader in the 1970s to a consistent follower in the 1990s. Thus, the lead-lag relationships across sectors change, and the relationship constantly evolves over time.

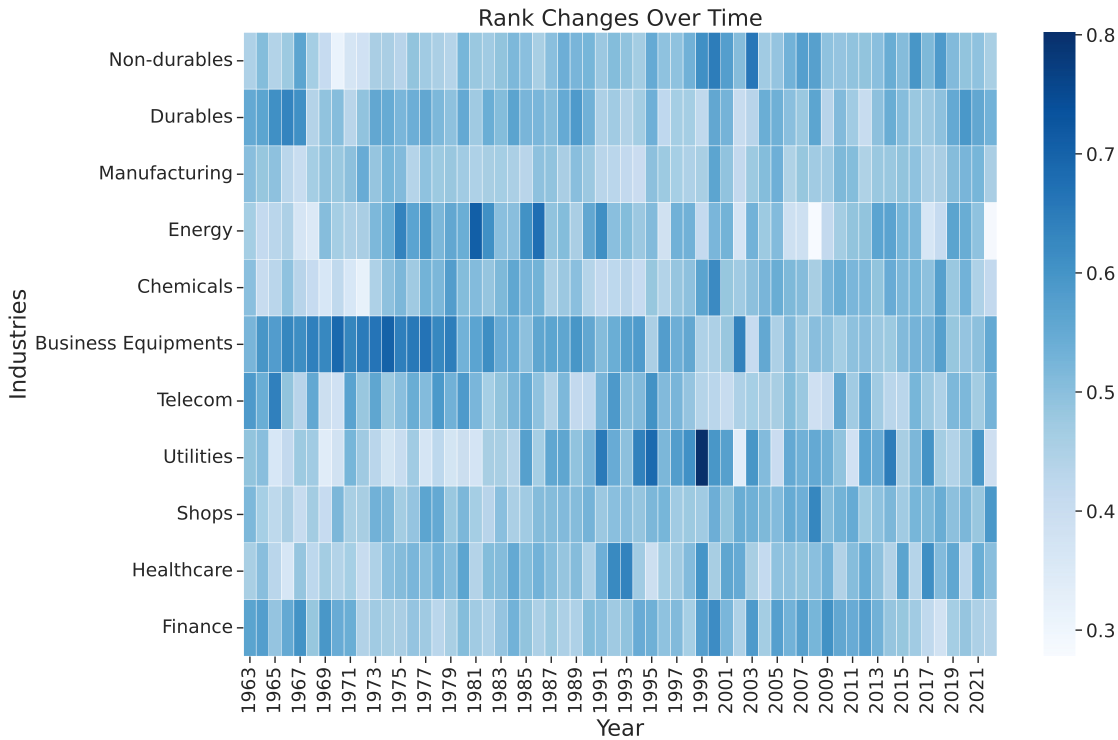


Figure 5: Change in sector rank of average lead-lag score for the Lévy-area lead-lag matrix construction from 1963 to 2022.

3.5 Lead-lag Relationship for Various Time Scales

In this section, we investigate the lead-lag relationships for various time frequencies, which is a new contribution to the existing literature. Specifically, we explore the daily, bidiurnal, weekly, bi-weekly, tri-weekly, and monthly time frequencies to gain insight into the persistence of the lead-lag relationships. By doing so, we provide a more comprehensive understanding of these relationships and how their temporal dynamics evolve from 1963 to 2022.

3.5.1 Performance of Lead-lag portfolios for Various Time Scales

Most of the extant literature including Chordia and Swaminathan (2000), Hou (2007), Frazzini and Cohen (2008), Parsons et al. (2020), and Huang et al. (2022) hypothesize that lead-lag effects are consequences of the slow information diffusion. If this hypothesis holds, then lead-lag relationships should disappear as we reduce the time frequency with which we detect the relationship. Therefore, one would expect the performance of the lead-lag portfolios to decrease as the frequency with which one ranks leaders and followers decreases.

To analyze the lead-lag relationships at different frequencies, we aggregate returns from daily to other frequencies and compute the pairwise lead-lag scores as before. Recall that, to compute weekly lead-lag relationships, we compute weekly raw returns $R_{t-5:t} = \Pi_{j\in[0,\cdots,5]}(1 + R_{t-j}) - 1$ and relabel the return $R_{t-5:t}$ as R_{t_w}, where t_w is the w^{th}

natural week of the year. Next, we normalize the computed weekly returns to feed into the steps in Algorithm 1 to detect lead-lag relationships.

Table 4: Performance of lead-lag portfolios for various frequencies

Panel A: Lévy-area portfolio for various frequencies

	Compound Return (%)	Return (bps/period)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Daily	24.87	8.82	0.59	2.37	24.67
Bi-Daily	11.02	8.30	0.76	1.23	24.17
Weekly	12.33	23.09	1.34	1.22	26.65
Bi-Weekly	8.37	31.94	1.69	0.95	19.59
Tri-Weekly	7.30	42.9	2.01	0.87	24.56
Monthly	7.81	65.8	2.56	0.86	22.50

Panel B: MaxCor Portfolio for various frequencies

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Daily	19.69	7.14	0.48	2.36	16.90
Bi-Daily	13.36	9.96	0.66	1.70	32.64
Weekly	13.32	24.84	1.09	1.62	28.48
Bi-Weekly	9.51	36.12	1.47	1.24	23.16
Tri-Weekly	8.86	50.64	1.80	1.15	16.95
Monthly	6.91	61.20	2.35	0.86	21.5

Panel C: AvgCor Portfolio for various frequencies

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Daily	27.97	9.79	0.70	2.21	28.64
Bi-Daily	16.46	12.10	0.96	1.41	52.67
Weekly	16.41	30.19	1.56	1.38	37.26
Bi-Weekly	12.63	47.31	1.99	1.19	33.79
Tri-Weekly	11.65	65.80	2.39	1.13	24.91
Monthly	9.76	85.36	3.02	0.94	19.52

Table 4 presents performances of the Lévy-area, MaxCor, and AvgCor portfolios for daily, bidiurnally, weekly, bi-weekly, tri-weekly, and monthly frequencies. Panels A, B, C in Table 4 show the results for Lévy-area, MaxCor, and AvgCor portfolios, respectively.

We observe an overall downward trend on portfolio performances as the rebalance frequency decreases from daily to monthly. The drop in annualized return and Sharpe ratio from the daily level to the bidiurnal level is the largest, and the subsequent performance drops from bidiurnal to lower frequencies is smoother and less pronounced. We also observe an overall upward trend in portfolio volatility as rebalance frequency decreases.

The results in Table 4 support the slow information diffusion hypothesis. The strength of lead-lag relationship erodes as the time between ranking of stocks and rebalancing the portfolios becomes longer. This result suggests that the market slowly absorbs the lead-lag relationship induced by the slow diffusion of information, and the residual lead-lag relationship at a monthly frequency becomes economically insignificant.

3.5.2 Slow Information Diffusion Hypothesis and the Speed of the Market

Table 4 shows that the lead-lag relationship is strongest when we detect the lead-lag profiles and rebalance the portfolios both on daily frequencies. Portfolio performance weakens as the frequency of lead-lag detection and of portfolio rebalancing decreases. Thus, have daily lead-lag relationships consistently outperformed lead-lag relationships at lower frequencies throughout the entire historical period?

In this section, our study points to the time when the daily data-driven lead-lag portfolios first outperform the lower-frequency portfolios. As Table 4 suggests, there is a point in time beyond which the daily lead-lag portfolios outperform all lower-frequency lead-lag portfolios; however, in the earlier years of our study, the daily lead-lag portfolio does not consistently outperform lower-frequency portfolios. For example, the daily Lévy-area portfolio begins to outperform the weekly Lévy-area portfolio in the mid to late-1970s; see Figure 6.

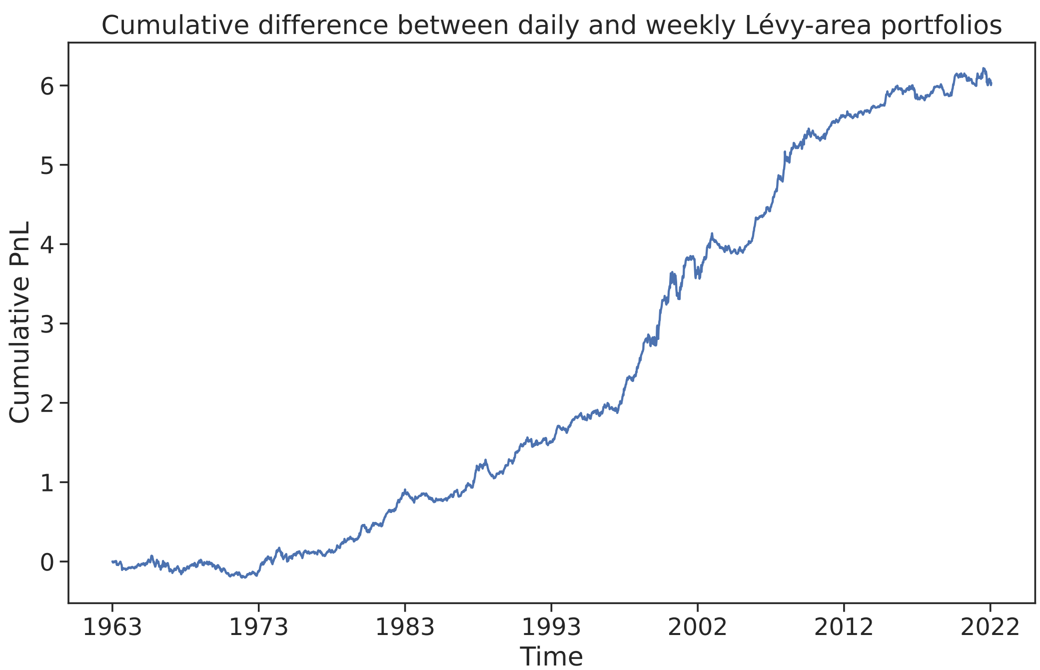


Figure 6: Cumulative sum of the difference between the daily lead-lag Lévy-area portfolio returns and the weekly lead-lag Lévy-area portfolio returns

To further explore how the frequency of lead-lag detection and rebalancing affects the returns of the three data-driven lead-lag portfolios, we record the first time these lead-lag portfolios outperform each of their corresponding lower-frequency portfolios. In particular, we track the initial occurrence of a cumulative excess return of 50% and 100% in the daily portfolio compared with the lower-frequency portfolios in Table 5. threeparttable

Table 5: Dates when daily lead-lag portfolios first outperform lower frequency lead-lag portfolios†

Panel A: First occurrence of a cumulative excess return of 50% between daily and lower-frequency portfolios

	Raw	Bidiurnally	Weekly	Bi-weekly	Tri-weekly	Monthly
Lévy-area	1968.07	1973.04	1981.09	1980.09	1978.10	1971.02
Max Cross-Corr	1968.03	1974.07	1996.02	1982.07	1975.08	1970.08
Avg Cross-Corr	1968.02	1974.01	1995.06	1985.02	1982.08	1973.12

Panel B: First occurrence of a cumulative excess return of 100% between daily and lower-frequency portfolios

	Raw	Bidiurnally	Weekly	Bi-weekly	Tri-weekly	Monthly
Lévy-area	1973.04	1974.09	1987.10	1981.11	1981.03	1974.12
Max Cross-Corr	1971.04	1990.08	1999.08	1985.09	1983.02	1974.07
Avg Cross-Corr	1970.06	1982.12	1996.02	1987.10	1984.07	1981.08

We use format YYYY.MM to express year and month in this table.

Table 5 presents the first month after January 1963 when the cumulatively gained profit of daily lead-lag portfolios is at least 50% and 100% above the cumulative returns of each of the lower-frequency portfolios. We also record the first month when the daily lead-lag portfolio cumulatively earns over 50% and 100% profit under the "Raw" column.

With the exception of bidiurnal lead-lag portfolios, we observe a clear trend in the time when the daily lead-lag portfolios start to outperform the lower-frequency portfolios. The daily portfolios cumulatively outperform the monthly portfolios first, starting from the 1970s, and then outperform the tri-weekly, bi-weekly, and weekly portfolio in chronological order during the 1980s and the 1990s. For bidiurnal portfolios, however, we observe that the daily lead-lag portfolios outperform the bidiurnal portfolios much sooner than the daily portfolios outperform the weekly, and the tri-weekly portfolios.

Results in Table 5 lend further support to the slow information diffusion hypothesis, which asserts that lead-lag relationships are caused by the different speeds with which information is impounded into the leaders and the followers. In such cases, it takes between a week and a month for most of the information to diffuse to the followers in the early 1970s; over time, some information is impounded on the next day of trading whereas some residual information still takes weeks to trickle down to the followers. Over time, more information is absorbed and reflected in the prices of the followers on the next trading day and less relevant information is left to be captured at the lower frequencies, which causes the daily rebalanced portfolio to outperform the lower frequency portfolios. The timings of changes in relative portfolio performances coincide with times when technological advancements enabled faster trading activities such as electronic execution and high-frequency trading, see McGowan (2010).

On the other hand, the bidiurnal frequency presents a unique dynamic. The literature does not compare performance of portfolios rebalanced at daily and bidiurnally, weekly, or monthly periods. Smith and Desormeau (2006) and Dennis et al. (1995) explore rebalancing periods over much slower frequencies such as annually or once every two years. We conjecture that if an investor possesses the technological capability to trade on a bidiurnal basis, it is likely that they can trade daily. As a result, market participants, especially informed traders who can trade at higher frequencies may find little incentive to trade bidiurnally because trading daily allows them to benefit from information asymmetries quicker than trading of lower frequencies. Consequently, the daily portfolios outperform the bidiurnal portfolio faster than they outperform the lower-frequency portfolios; we leave the discussion of this hypothesis to future work.

4 Robustness Analysis

In this section, we present the results of a number of robustness checks and alternative specifications in our study. The appendix includes further robustness analysis.

4.1 Execution Cost and Trading Turnover

Here, we explore the impact of trading frictions on the performance of the lead-lag portfolios. We study performances of both the data-driven portfolios and the benchmark portfolios as a function of trading frictions. To estimate the execution costs for a given portfolio, we first estimate the execution costs for trading one dollar worth of equity in the US market and then multiply this value by the turnover of the portfolio. The turnover of a portfolio that trades $d_1^t, \cdots, d_n^t$ dollar value in stocks $s_1, \cdots, s_n$ at time t is defined as

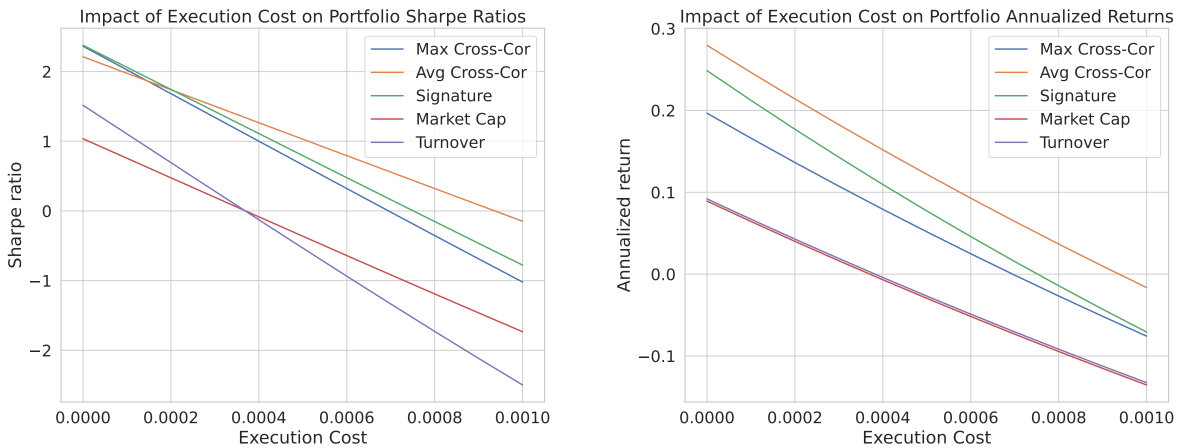
$$\text{TVR}_t = \frac{\sum_{i=1}^n |d_i^t - d_i^{t-1}|}{\sum_{i=1}^n |d_i^{t-1}|}, \tag{7}$$

where $|d_i^t - d_i^{t-1}|$ is the change in dollar value held by the portfolio on asset s_i in between rebalances.

For each portfolio, we compute its daily turnover and explore the portfolio's performance as a function of per-dollar execution cost. Figure 7a shows Sharpe ratios of the portfolios as a function of execution costs, and Figure 7b shows the returns of the portfolios as a function of execution costs. For each strategy, Table 6 presents the mean and standard deviation of its daily turnover ratio.

Table 6: Turnover ratio of various portfolios

	Average turnover ratio (%)	Average change in portfolio composition (%)	Proportion of sign flips (%)
Max Cross-Cor	102.5	22.6	44.3
Avg Cross-Cor	104.5	26.7	44.1
Lévy-area	111.3	23.6	50.4
Market Cap	95.4	1.21	44.2
Turnover	94.4	1.22	43.3



(a) (b)

Figure 7: (a) Sharpe ratio of portfolios as a function of per-dollar execution costs. Per-dollar execution cost is the cost incurred when performing a single trade on a portfolio of one dollar. (b) Annualized return of different portfolios as a function of per-dollar execution costs.

Figures 7a and 7b show that the data-driven lead-lag portfolios are significantly more robust to execution costs than the benchmark portfolios in terms of Sharpe ratio and annualized return. The Market Cap portfolio and the Turnover portfolio deliver losses when execution costs on each dollar traded is larger than 4 bps;[17] on the other hand, the data-driven portfolios are profitable when execution costs are 7 bps or lower, and their Sharpe ratio remains above 1 when execution costs are 4 bps.

We analyze the turnover ratio of the data-driven portfolios and of the benchmark portfolios in Table 6. We split the turnover ratio of the portfolios into two components, change in portfolio composition and sign flips. When the sign of the signal changes (i.e., the sign of the return of the leaders change), the turnover ratio of the portfolio is 2 regardless of the changes in the composition of the portfolio; on the other hand, when there is no sign flip, the turnover ratio of the portfolio is the proportion of the followers that change from the previous day's portfolio. Table 6 reports the average daily turnover ratio, average proportional change in portfolio composition, and the proportion of days when the sign of the signal flips.

Table 6 shows that ratio of turnover of the the data-driven portfolios is higher than that of the Market Cap portfolio and the Turnover portfolio. The composition of the data-driven lead-lag portfolios tends to change more than the baseline portfolios, but their proportion of sign flips is similar. Figures 7a and 7b show that the higher turnover ratio is not sufficient to offset the difference in performance between the data-driven portfolios and the baseline portfolios.

Finally, there is no bid-offer spread during the close auction, so our portfolios are not affected by the bid-offer spread that one would face during regular trading hours. Specifically, during the closing auction, buy and sell orders are collected and at a specified closing time, these orders are matched and executed at a single close price.

[17]Currently, industry standard on trading a portfolio on large cap US equities is less than 3 bps per dollar. The fees for taking short positions is around 1 bps per dollar.

4.2 Exposure to illiquidity premium and Fama–French Factors

Amihud (2002) suggests that stock illiquidity positively affects stock excess return, and this phenomenon is referred to as illiquidity premium. In this section, we explore if the lead-lag portfolios we construct can be explained by the illiquidity premium of stocks and by other common risk factors.

We measure the illiquidity of stocks as the daily ratio of absolute stock return to its dollar volume, see Amihud (2002). Next, we construct a portfolio that consists of a long position in stocks with the highest 20% illiquidity measure and a short position in stocks with the lowest 20% illiquidity measure; we denote by $\mathrm{R}_t^{\mathrm{liq}}$ the returns of such portfolio for each day t.

We use linear models to explore if the illiquidity portfolio adds value in explaining the returns of our lead-lag portfolio, and we also explore if the returns of our lead-lag portfolio can be explained by the benchmark lead-lag portfolios discussed in section 3.2, and by the Fama–French five factors. Consider the models

$$\mathrm{R}_t^{\text{Lévy}} = \alpha + \beta_{cap}\,\mathrm{R}_t^{\mathrm{cap}} + \beta_{tvr}\,\mathrm{R}_t^{\mathrm{tvr}} + \epsilon_t, \tag{M1}$$

$$\mathrm{R}_t^{\text{Lévy}} = \alpha + \beta_{liq}\,\mathrm{R}_t^{\mathrm{liq}} + \beta_{cap}\,\mathrm{R}_t^{\mathrm{cap}} + \beta_{tvr}\,\mathrm{R}_t^{\mathrm{tvr}} + \epsilon_t, \tag{M2}$$

$$\mathrm{R}_t^{\text{Lévy}} = \alpha + \beta_{liq}\,\mathrm{R}_t^{\mathrm{liq}} + \beta_{cap}\,\mathrm{R}_t^{\mathrm{cap}} + \beta_{tvr}\,\mathrm{R}_t^{\mathrm{tvr}} + \sum_{\mathrm{Fama5}} \beta_{FF5}\,\mathrm{R}_t^{\mathrm{FF5}} + \epsilon_t, \tag{M3}$$

where $\mathrm{R}_t^{\text{Lévy}}$ is the return of the Lévy-area lead-lag portfolio on a daily rebalancing frequency, $\mathrm{R}_t^{\mathrm{cap}}$ and $\mathrm{R}_t^{\mathrm{tvr}}$ are returns of the two benchmark lead-lag portfolios based on market capitalization and stock turnover ratios, and $\mathrm{R}_t^{\mathrm{FF5}}$ are the returns of the five-factor Fama–French portfolios.

Table 7: Regression of Lévy-Area lead-lag portfolio against various benchmarks

	Intercept [bps]	β_{liq}	β_{cap}	β_{tvr}	β_{Mkt}	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}	$\mathrm{R}^2[\%]$
M1	$10.1^{\ddagger}$	-	$0.401^{\ddagger}$	$-0.026^{\dagger}$	-	-	-	-	-	12.3
	(22.26)		(45.08)	(-1.99)						
M2	$11.0^{\ddagger}$	$-1.495^{\ddagger}$	$0.397^{\ddagger}$	-0.016	-	-	-	-	-	12.5
	(22.8)	(-5.38)	(44.48)	(-1.24)						
M3	$11.0^{\ddagger}$	$-1.462^{\ddagger}$	$0.402^{\ddagger}$	-0.022^{*}	$0.0001^{\dagger}$	-0.00008	-0.0001	-0.0002	-0.0002	12.6
	(22.8)	(-5.25)	(44.71)	(-1.70)	(2.56)	(-0.85)	(-1.05)	(-1.77)	(-1.41)	

Table 7 reports results from regressions of daily returns for M1, M2, and M3. Data are from CRSP, and returns are measured from close-to-close. The t-statistics are reported in parentheses, and statistical significance at the 1%, 5%, and 10% level is indicated by ‡, †, and *, respectively.

There is little difference in the intercept and the R^2 across the three models. Overall, we observe that on average, around 10 to 11 basis points of the daily returns of the Lévy-Area lead-lag portfolio cannot be explained by the risk factors and benchmark portfolios we include in the models. Notably, the Lévy-Area lead-lag portfolio exhibits a positive exposure to the market capitalization lead-lag portfolio, indicating a propensity to select smaller-sized stocks. On the other hand, it demonstrates a negative exposure to the Turnover ratio lead-lag portfolio and the illiquidity premium portfolio, suggesting a preference for more liquid stocks and for stocks with higher turnover ratios.

Results for M2 and M3 show that the returns of the Lévy-Area lead-lag portfolio are not statistically significantly explained by the Fama–French five factors after considering lead-lag relationship related to size and liquidity. We observe no significant difference in the intercept and R^2 between the two models, and the coefficients on the Fama–French five factors are both small in magnitude and statistically insignificant.

Furthermore, comparisons between M1 and M2 imply some collinearity between the illiquidity premium portfolio and the Turnover ratio lead-lag portfolio. The addition of the illiquidity premium portfolio to the regressions results in a 0.1% increase in R^2, and the coefficient on the Turnover ratio lead-lag portfolio becomes statistically insignificant. These findings suggest that the illiquidity premium portfolio does not contribute significantly to explaining the returns of the Lévy-Area lead-lag portfolio beyond the information provided by the two benchmark lead-lag portfolios included in the analysis.

4.3 Alternative Selection of Sets of Stocks

In this subsection, we show that the lead-lag relationship we detect is robust to alternative selection of the set of stocks. In our data-driven baseline results, the set of stocks in a given trading day consisted of the top 25 percentile in terms of market capitalization in that day. This choice ensures that the portfolios we construct can be traded at scale in the market.

To assess the robustness of our results for a different set of stocks, we select those that rank at the top 25 percentile in terms of both volume traded and turnover ratio. We obtain a smaller set of stocks than the one used above because volume and turnover ratio vary more than market capitalization. On average, this new set of stocks contains approximately one third of the stocks that we use in the main results above. Table 8 reports the results for this alternative set of stocks.

Panel A of Table 8 compares the performance of the lead-lag portfolios without clustering. The daily returns for the portfolios range from 1.80 bps per day for the Market Cap portfolio to over 17 bps per day for the Lévy-area portfolio. The data-driven portfolios outperform the two benchmark portfolios for all criteria except for volatility and maximum drawdown. The performances of portfolios with the alternative set of stocks show the same relative hierarchy as in the main results presented above. In particular, the Lévy-area portfolio achieves 51.85% annualized return. The Sharpe ratio of the data-driven portfolios is higher than those of the two portfolios based on firm characteristics. This shows that the data-driven methods capture economically significant lead-lag relationships in the alternative set of stocks. In particular, the Sharpe ratio and annualized returns of the Lévy-area portfolio are the highest among the three data-driven methods in panel A of Table 8.

Panel B of Table 1 compares the performances of various clustered lead-lag portfolios. The performances of the Max Cross-Cor, Avg Cross-Cor, and Lévy-area portfolios are economically significant, with returns ranging from 9.18 to

Table 8: Performance of lead-lag portfolios in the alternative set of stocks

Panel A: Global Lead-Lag Portfolios

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	35.57	12.08	1.90	1.01	82.47
Avg Cross-Cor	48.69	15.75	2.09	1.20	76.44
Lévy-area	51.85	16.59	2.08	1.27	91.31
Market Cap	4.62	1.80	1.77	0.16	99.60
Turnover	21.48	7.72	1.23	1.00	52.49

Panel B: Clustered Lead-lag Portfolios

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	26.03	9.18	1.08	1.35	46.28
Avg Cross-Cor	27.61	9.68	1.16	1.33	36.10
Lévy-area	31.45	10.86	1.18	1.46	40.53
Industry	4.55	1.77	1.39	0.21	87.55

10.86 bps/day. However, the Industry portfolio is not economically significant, with a return of less then 2 bps per day. The data-driven portfolios, Max Cross-Cor, Avg Cross-Cor, and Lévy-area, outperform the Industry portfolio in all categories. In particular, the compounded return of the Lévy-area portfolio is 31.45%, and its Sharpe ratio of 1.46 is the highest. This suggests that the data-driven methods used in constructing the lead-lag portfolios capture economically significant lead-lag relationships, with the Lévy-area method being particularly effective, and are robust to this alternative set of stocks.

When compared with the results in Table 1, the performance of the portfolios for the alternative set of stocks is less economically significant. Although the annualized returns of portfolios in panel A of Table 8 are almost twice the value of those in Table 8, the volatility of returns of portfolios for the alternative set of stocks is almost four times the volatility of the portfolios in Table 1. Therefore, the Sharpe ratio results in Table 8 are half the value of those shown in Table 1. This result shows that including more stocks benefits the performance of lead-lag portfolios, while the gap in performance between the data-driven portfolios and the portfolios based on firm characteristics remains robust to the selection of the set of stocks.

Furthermore, the benchmark portfolios with the alternative set of stocks perform worse than those with the set of stocks discussed above; meanwhile, the data-driven portfolios still remain economically significant. This shows that the data-driven methods in lead-lag detection are robust and transferable to various sets of assets.

4.4 Intermediate Lead-lag Relationships

In this section, we show that the lead-lag relationship we find is robust to considering the so-called "intermediate lead-lag relationships".

Recall that when identifying leaders and followers among the assets, we apply ranking methods to sort the assets from most likely to lead to most likely to follow and construct a portfolio that uses the returns of assets that rank at the top to predict the returns of the assets that rank at the bottom. While this method provides a criterion to evaluate lead-lag relationships, it neglects assets that rank in the middle of the sorting. Consider, for example, a set of three assets a, b, c, where asset b ranks in the middle, follows a, and leads c. In this case, asset b exhibits an "intermediate lead-lag relationship" as it is not included in the lead-lag portfolio, but possesses lead-lag relationships with other assets. It is also possible that asset b has no relationship with a or c.

Here, we extend the lead-lag detection mechanisms we designed above to study intermediate lead-lag relationships. For each trading day, we rank the assets into quantiles, denoted by $Q1, Q2, Q3, Q4$, respectively. Instead of using the previous return of $Q1$ to predict the return of $Q4$ as in Algorithm 1, we construct three portfolios in which we sequentially use $Q1$ to predict $Q2$, $Q2$ to predict $Q3$, and $Q3$ to predict $Q4$. We finance each of the three portfolios with a position in the SPY ETF and take the average of the three portfolios to construct the final portfolio, see Figure 8.

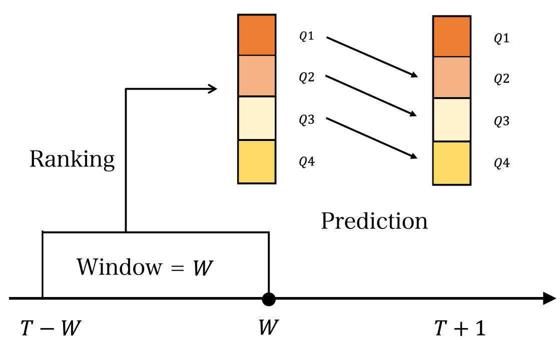


Figure 8: Portfolio construction of lead-lag portfolio with intermediate lead-lag relationships

The portfolio illustrated in Figure 8 considers the intermediate lead-lag relationships, so measuring the performance of the portfolio can help to understand if these relationships are economically significant. Table 9 reports the performance of the Lévy-area, Max Cross-Cor, and Avg Cross-Cor, respectively.

Table 9: Intermediate Lead-lag Relationships

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	16.1	5.91	0.40	2.37	15.8
Avg Cross-Cor	14.8	5.48	0.35	2.52	14.1
Lévy-area	20.4	7.36	0.40	2.89	14.7

The table shows that the lead-lag portfolios are economically significant after considering intermediate lead-lag relationships. However, compared with the results without considering intermediate lead-lag relationships in Table 1 where the top 20% of stocks are considered leaders and the bottom 20% of stocks are considered followers, the portfolios in Table 9 show lower return, but obtain lower volatility, higher Sharpe ratio, and lower maximum drawdown. Thus, the results show that our method is robust to considering intermediate lead-lag relationships and that intermediate lead-lag relationships do exist in the equities market.

5 Conclusion

This paper presented a method to detect linear and nonlinear lead-lag relationships in the US equity market. In contrast to the extant literature, which uses firm characteristics such as market capitalization, trading turnover, and trading volume to select leaders and followers, our method employs the Lévy-area between pairs of stock returns to infer which one in the pair is more likely the leader, and to quantify the strength of this relationship. We constructed a portfolio that uses the previous returns of the leaders to determine positions on the followers; and showed that they generate economically significant performances that outperform all benchmarks in the literature. The performance of our portfolios is robust to various alternative specifications in algorithms, hyperparameters, and data sets.

The performance of the lead-lag portfolios we construct cannot be fully explained by lead-lag effects generated by market capitalization, trading volume, or intra-industry; there is little overlap between the composition of our lead-lag portfolios and that of the benchmarks. Our results show that accounting for nonlinearities is key to determine the lead-lag relationships.

The lead-lag relationships we find change over time. The leader-follower identity of stocks in various sectors changes several times between 1963 and 2022. This finding further supports the necessity of data-driven lead-lag detection methods that capture dynamically evolving lead-lag relationships.

Finally, we examined the performance of our portfolios across various rebalancing frequencies, and the results provided empirical support to confirm the slow information diffusion hypothesis. Specifically, the performance of portfolios decreases as both the ranking and the rebalancing are performed less frequently.

Our research leads to various future directions of work. One, explore lead-lag relationship on intra-day data, for alternative markets, and in alternative asset classes. Such a study might explain the relative differences in performance of lead-lag portfolios that are rebalanced at various frequencies over time. Two, understand the analytical properties of the nonlinear relationship between leaders and followers with methods such as smoothing or other non-parametric approaches. This direction can help to understand the source and composition of lead-lag relationships from the perspective of lagged factor exposures or other asset pricing models. Finally, we intentionally kept the construction of portfolios simple in this paper; future work is to study advanced portfolio construction methods that optimize the portfolio positions based on the lead-lag relationships we uncover.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56.

Badrinath, S. G., Kale, J. R., and Noe, T. H. (1995). Of shepherds, sheep, and the cross-autocorrelations in equity returns. *The Review of Financial Studies*, 8(2):401–430.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Basnarkov, L., Stojkoski, V., Utkovski, Z., and Kocarev, L. (2020). Lead–lag relationships in foreign exchange markets. *Physica A: Statistical Mechanics and its Applications*, 539:122986.

Bennett, S., Cucuringu, M., and Reinert, G. (2022). Lead-lag detection and network clustering for multivariate time series with an application to the us equity market.

Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis*, 38(4):747–777.

Brennan, M. J., Jegadeesh, N., and Swaminathan, B. (1993). Investment Analysis and the Adjustment of Stock Prices to Common Information. *The Review of Financial Studies*, 6(4):799–824.

Buccheri, G., Corsi, F., and Peluso, S. (2019). High-frequency lead-lag effects and cross-asset linkages: A multi-asset lagged adjustment model. *Journal of Business Economic Statistics*, 39:1–28.

Campbell, J. Y. and Cochrane, J. H. (1995). By force of habit: A consumption-based explanation of aggregate stock market behavior. Working Paper 4995, National Bureau of Economic Research.

Chan, K. (1992). A further analysis of the lead–lag relationship between the cash market and stock index futures market. *The Review of Financial Studies*, 5(1):123–152.

Chester Curme, Michele Tumminello, R. N. M. H. E. S. and Kenett, D. Y. (2015). Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 15(8):1375–1386.

Chevyrev, I. and Kormilitzin, A. (2016). A primer on the signature method in machine learning.

Chordia, T. and Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *Journal of Finance*, 55:913–935.

Cucuringu, M. (2015). Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and semidefinite programming synchronization.

Cucuringu, M., Li, H., Sun, H., and Zanetti, L. (2019). Hermitian matrices for clustering directed graphs: insights and applications. *CoRR*, abs/1908.02096.

De Bacco, C., Larremore, D. B., and Moore, C. (2018). A physical model for efficient ranking in networks. *Science advances*, 4(7):eaar8260.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., and A. P. Santos, A. (2021). Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock Return Serial Dependence and Out-of-Sample Portfolio Performance. *The Review of Financial Studies*, 27(4):1031–1073.

Dennis, P., Perfect, S., Snow, K., and Wiles, K. (1995). The effects of rebalancing on size and book-to-market ratio portfolio returns. *Financial Analysts Journal*, 51:47–57.

Fogel, F., d'Aspremont, A., and Vojnovic, M. (2014). Serialrank: Spectral ranking using seriation. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Frazzini, A. and Cohen, L. (2008). Economic links and predictable returns. *Journal of Finance*, 63:1977–2011.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Gyurkó, L. G., Lyons, T., Kontkowski, M., and Field, J. (2013). Extracting information from the signature of a financial data stream.

Hause, J. C. (1971). Spectral analysis and the detection of lead-lag relations. *The American Economic Review*, 61(1):213–217.

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70.

Hou, K. (2007). Industry Information Diffusion and the Lead-lag Effect in Stock Returns. *The Review of Financial Studies*, 20(4):1113–1138.

Huang, S., Lee, C. M., Song, Y., and Xiang, H. (2022). A frog in every pan: Information discreteness and the lead-lag returns puzzle. *Journal of Financial Economics*, 145(2, Part A):83–102.

Hubert, L. Arabie, P. (1985). Comparing partitions. pages 193–218.

Kelly, B., Malamud, S., and Pedersen, L. H. (2023). Principal portfolios. *The Journal of Finance*, 78(1):347–387.

Kelly, B. T., Malamud, S., and Zhou, K. (2022). The virtue of complexity in return prediction. Working Paper 30217, National Bureau of Economic Research.

Li, Y., Liu, C., Wang, T., and Sun, B. (2021). Dynamic patterns of daily lead-lag networks in stock markets. *Quantitative Finance*, 21(12):2055–2068.

Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *The Review of Financial Studies*, 3(2):175–205.

Lyons, T. (2014). *Rough paths, signatures and the modelling of functions on streams*. Kyung Moon SA.

McGowan, M. (2010). The rise of computerized high frequency trading: Use and controversy. *Duke Law and Technology Review*, 16.

Menzly, L. and Ozbas, O. (2009). Market segmentation and cross-predictability of returns. *Journal of Finance*, 65:1555–1580.

Ng, A., Jordan, M., and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In Dietterich, T., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press.

Parsons, C. A., Sabbatucci, R., and Titman, S. (2020). Geographic Lead-Lag Effects. *The Review of Financial Studies*, 33(10):4721–4770.

Pohl, W., Schmedders, K., and Wilms, O. (2018). Higher order effects in asset pricing models with long-run risks. *The Journal of Finance*, 73(3):1061–1111.

Sakurai, Y., Papadimitriou, S., and Faloutsos, C. (2005). Braid: Stream mining through group lag correlations. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, SIGMOD '05, page 599–610, New York, NY, USA. Association for Computing Machinery.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3):393–415.

Scherbina, A. and Schlusche, B. (2018). Follow the Leader: Using the Stock Market to Uncover Information Flows between Firms. *Review of Finance*, 24(1):189–225.

Shi, D., Calliess, J.-P., and Cucuringu, M. (2023). Multireference Alignment for Lead-Lag Detection in Multivariate Time Series and Equity Trading. *to appear in Proceedings of the Fourth ACM International Conference on AI in Finance (ICAIF 2023); SSRN:4560780*.

Smith, D. and Desormeau, W. (2006). Optimal rebalancing frequency for stock-bond portfolios. *Journal of Financial Planning*, 19:52–63.

Yan, J. and Yu, J. (2023). Cross-stock momentum and factor momentum. *Journal of Financial Economics*, 150(2):103716.

Zaharudin, K. Z., Young, M. R., and Hsu, W.-H. (2022). High-frequency trading: Definition, implications, and controversies. *Journal of Economic Surveys*, 36(1):75–107.

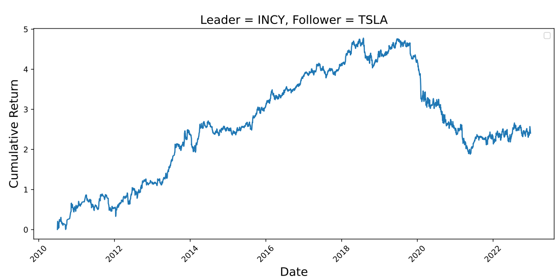
Zhang, Y., Cucuringu, M., Shestopaloff, A. Y., and Zohren, S. (2023). Dynamic Time Warping for Lead-Lag Relationships in Lagged Multi-Factor Models. *to appear in Proceedings of the Fourth ACM International Conference on AI in Finance (ICAIF 2023); SSRN:4572554*.

A Appendix 0: Notable lead-lag pairs as motivating examples

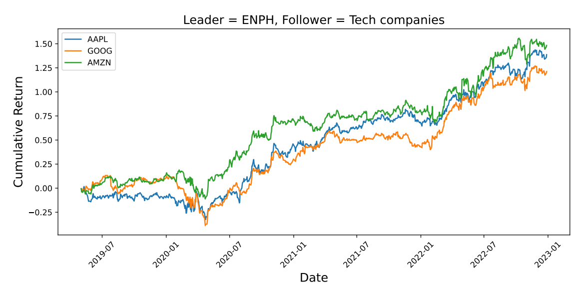
We present several motivating examples of pairwise lead-lag relationships that we detect. Some examples show the time-varying property of lead-lag relationships while others present examples of lead-lag relationships that cannot easily reconcile with what the literature discovered in the past.

For each proposed lead-lag pair, we consider a pairs trading strategies where one uses the previous time period return of the leader as an indicator to buy or sell the follower at the current time period.

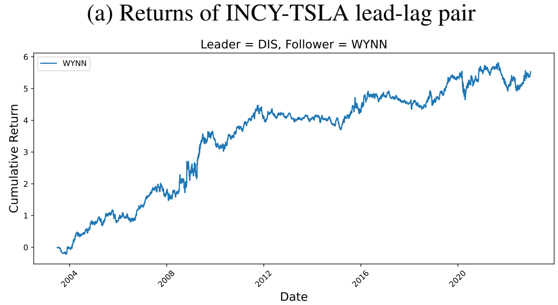
Figure 9 presents returns of pairs trading strategies that illustrate the time-varying property of lead-lag relationships. Example 9a shows that a ten-year lead-lag relationship between INCY, a pharmaceutical company and Tesla breaks in October 2019 when Tesla had a surprisingly good earnings announcement. Example 9b shows that a collection of lead-lag relationships between ENPH, a Californian energy company and several technology firms including AAPL, GOOG, and AMZN started to appear in mid-2020, coinciding COVID. These two motivating examples show that lead-lag relationships can appear and disappear between pairs of stocks, which suggests that it is necessary to detect, instead of assume, lead-lag relationships over rolling windows.



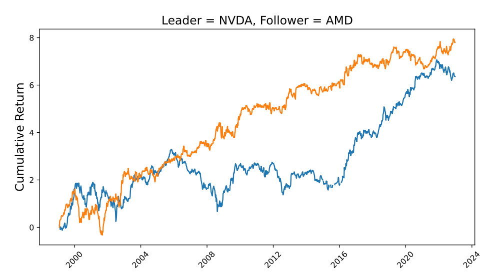
(a) Returns of INCY-TSLA lead-lag pair



(b) Returns of ENPH-Tech lead-lag pairs



(c) Returns of DIS-WYNN lead-lag pair



(d) Returns of NVDA-AMD lead-lag pair

Figure 9: Motivating examples of pairwise lead-lag relationships.

Example 9c shows that a twenty-year lead-lag relationship between Disney and WYNN, a casino and resort company. Wynn Resorts has larger market cap than Disney and there is no clear economic links between the two companies; therefore, one cannot easily explain this lead-lag pair using lead-lag relationships discovered in the literature. On the other hand, Example 9d shows a surprising lead-lag relationship between NVIDIA and AMD, two GPU companies. This example is built on a weekly frequency, i.e., the pairs trading strategy is executed only weekly. It is surprising that a simple weekly pairs trading strategy (orange) between two firms that are clearly economically linked is still profitable in 2023.

B Appendix 1: Introduction to Signature and Lévy-area

We provide a brief introduction to the concept of Signature and Lévy-area in this section, see Lyons (2014), Gyurkó et al. (2013), and Chevyrev and Kormilitzin (2016) for more details.

First, we define the concept of a real, continuous path.

Definition B.1. *A real, continuous path defined on the interval* $[0, T]$ *is a continuous function* $f : [0, T] \to \mathbb{R}$.

For a path $X_t = \left(X_t^1, \cdots, X_t^d\right)$ where each X_t^j is a real and continuous path, consider the d^k dimensional set I^0 with elements $(i_i, \cdots, i_k)$ with $k \geq 0$ and $i_m \in \{1, \cdots, d\}$ for $m = 1, \cdots, k$. The signature $S_{s,t}(X)$ of the path X over a time interval $[s, t]$ is a map from X to a sequence $\left(S(X)_{s,t}^I\right)_{I \in I^0}$, where

$$S(X)_{s,t}^I = \int_{s<u_1<u_2<\cdots<u_k<t} dX_{u_1}^{i_i}\, dX_{u_2}^{i_2} \cdots dX_{u_k}^{i_k}\,. \tag{A.1}$$

Any elements i_p, i_q in the set I^0 can be different or the same, and hence if the entries of $S_{s,t}(X)$ are iterated, one obtains expressions such as $S(X)_{s,t}^{1,1,2}$ or $S(X)_{s,t}^{1,2,2}$.

For a clearer intuition, note that the signature of X can be expanded as

$$S(X)_{s,t} = (1, S(X)_{s,t}^1,\, \cdots, S(X)_{s,t}^d,\, S(X)_{s,t}^{1,1},\, S(X)_{s,t}^{1,2}, \cdots)\,. \tag{A.2}$$

Here, for the first-order terms in $S(X)_{s,t}$, it is clear that $S(X)_{s,t}^i = X_t^i - X_s^i$, and by definition, the higher-order terms in $S(X)_{s,t}$ can also be iteratively calculated as

$$S(X)_{s,t}^{i,j} = \int_{s<a<t} S(X)_{s,a}^i\, dX_a^j\,. \tag{A.3}$$

The Lévy-area between two paths X_i and X_i is

$$A_{i,j} = \frac{1}{2}\left(S(X)_{s,t}^{i,j} - S(X)_{s,t}^{j,i}\right). \tag{A.4}$$

For a more intuitive understanding of Theorem 1, Figure 10 shows a two-dimensional continuous path with dimensions $X^1 = \{X_0^1, \cdots, X_3^1\}$ and $X^2 = \{X_0^2, \cdots, X_3^2\}$. The Lévy-area between X^1 and X^2 is $A^+ - A^-$ where A^+ and A^- are regions bounded by the path itself and the chord connecting the start and end of the path. In this example, we set $X^1 = \{0, 0.5, 2, 2.5\}$, $X^2 = \{0.5, 2, 2.5, 3.5\}$; i.e., we artificially set X^2 to lead X^1. The Lévy-area between X^1 and X^2 is negative, as expected from Theorem 1.

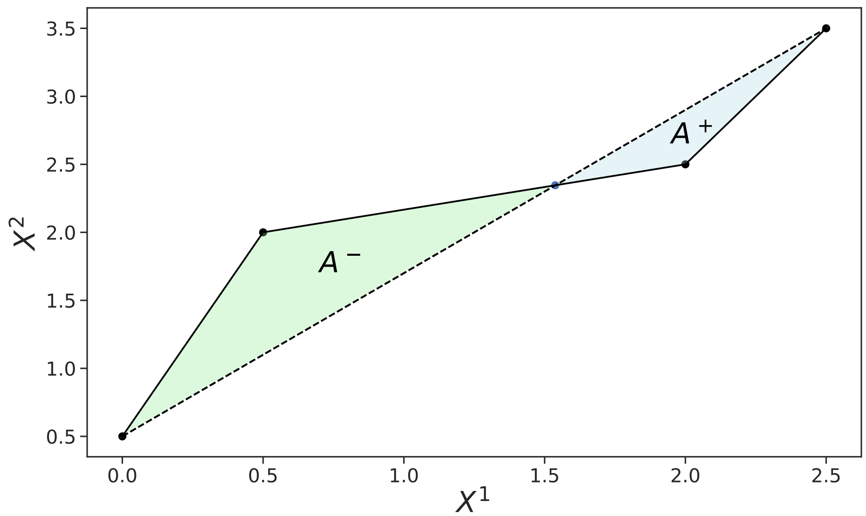


Figure 10: Illustration of the Lévy-area between two time series X^1 and X^2

Hence, for a multi-dimensional path X, we can assess and quantify the lead-lag relationship between two dimensions of X by calculating the signature and Lévy-area of the path along the two dimensions.

C Appendix 2: Proof of Theorem 1

Proof. Without loss of generality, assume $\ell > 0$. If $\ell < 0$, we re-write the regression equation by exchanging the order of X^i and X^j. Consider the Lévy-area between X_t^i and X_t^j

$$A_{i,j} = \frac{1}{2}\left(S(X)_{s,t}^{i,j} - S(X)_{s,t}^{j,i}\right) . \tag{B.1}$$

Next, write the right-hand side of (B.1) as

$$S(X)_{s,t}^{i,j} - S(X)_{s,t}^{j,i} = \int_{s<a<t}\int_{s<b<a} dX_b^i\, dX_a^j - \int_{s<a<t}\int_{s<b<a} dX_b^j\, dX_a^i\,. \tag{B.2}$$

Integrate the inner parts of the right-hand side of (B.2) to obtain

$$\int_{s<a<t}\int_{s<b<a} dX_b^i\, dX_a^j - \int_{s<a<t}\int_{s<b<a} dX_b^j\, dX_a^i = \int_{s<a<t}(X_a^i - X_s^i)\, dX_a^j - \int_{s<a<t}(X_a^j - X_s^i)\, dX_a^i\,. \tag{B.3}$$

Without loss of generality, assume that $X_s^i = X_s^j = 0$ (i.e., the return of assets at the start of time is 0), and hence we have

$$S(X)_{s,t}^{i,j} - S(X)_{s,t}^{j,i} = \int_{s<a<t} X_a^i\, dX_a^j - \int_{s<a<t} X_a^j\, dX_a^i\,. \tag{B.4}$$

The integration can be transformed into finite summations

$$
\begin{aligned}
S(X)^{i,j}_{s,t} - S(X)^{j,i}_{s,t} &= \sum_{s<a<t} X^i_a \Delta X^j_a - \sum_{s<a<t} X^j_a \, \Delta X^i_a \\
&= \sum_{s<a<t} X^i_a \, (X^j_a - X^j_{a-1}) - \sum_{s<a<t} X^j_a \, (X^i_a - X^i_{a-1}) \\
&= \sum_{s<a<t} -X^i_a \, X^j_{a-1} - \sum_{s<a<t} -X^j_a \, X^i_{a-1}
\end{aligned} \tag{B.5}
$$

because the processes X^i_t and X^j_t are discrete.

Substitute in the lead-lag equation $X^i_a = \beta_\ell \, f(X^j_{a-\ell}) + \epsilon$ into equation (B.5) to obtain

$$
\begin{aligned}
\sum_{s<a<t} -X^i_a X^j_{a-1} - \sum_{s<a<t} -X^j_a X^i_{a-1} = \sum_{s<a<t} X^j_a \, (\beta_\ell \, f(X^j_{a-\ell-1}) + \epsilon) \\
- \sum_{s<a<t} X^j_{a-1} \, (\beta_\ell \, f(X^j_{a-\ell}) + \epsilon) \, .
\end{aligned} \tag{B.6}
$$

Next, take expectation on both sides to reduce equation (B.6) to

$$
\mathbb{E}\left[\sum_{s<a<t} X^j_a \, X^i_{a-1} - \sum_{s<a<t} X^i_a \, X^j_{a-1}\right] = \mathbb{E}\left[\sum_{s<a<t} \beta_\ell \, f(X^j_{a-l}) X^j_{a-1} - \sum_{s<a<t} \beta_\ell \, f(X^j_{a-l-1}) \, X^j_a\right] . \tag{B.7}
$$

Here, consider $\sum_{s<a<t} \beta_\ell \, f(X^j_{a-\ell-1}) X^j_a \approx 0$ because we assume that the auto-correlation in the time series is very small compared to the variance of the time series. Therefore, it is clear that the expectation on the right-hand side is non-zero only if $\ell = 1$; hence, we consider $\mathbb{E}[\sum_{s<a<t} \beta_\ell \, f(X^j_{a-1}) X^j_{a-1}]$.

First, suppose f is a polynomial. Then for each of the terms of f we have that $\mathbb{E}[(X^j_{a-1})^{2k+1} X^j_{a-1}] \geq 0$ and $\mathbb{E}[(X^j_{a-1})^{2k} X^j_{a-1}] = 0$ for any positive integer k because $(X^j_{a-1})^{2k+1}$ is an odd function as X^j has mean zero and a symmetric distribution. The expectation $\mathbb{E}[\sum_{s<a<t} \beta_(X^j_{a-1}) X^j_{a-1}]$ is positive if $\beta_\ell \geq 0$ because we assume all derivatives of f are non-negative for any monic polynomial f.

Now, by the Stone–Weierstrass theorem, the set of polynomials is dense in the Banach space $C^\infty(\mathbb{R}^n)$, hence for any continuous function f, there exists a sequence of polynomials f_n such that the sequence converges to f under the infinity norm. Hence, we have shown that for all continuous functions f, the sign of $\mathbb{E}[\sum_{s<a<t} \beta_(X^j_{a-1}) X^j_{a-1}]$ and l is the same if and only if $\ell = 1$ when $\ell > 0$. And by symmetry, the sign of $\mathbb{E}[\sum_{s<a<t} \beta_(X^j_{a-1}) X^j_{a-1}]$ and ℓ are the same if and only if $\ell = -1$ when $\ell < 0$. This proves the first part of the theorem.

For the second part of the theorem, suppose f has third order derivatives at 0 (without loss of generality), then by Taylor's theorem there exists a polynomial $p(X^j_{a-1})$ such that it approximates $f(X^j_{a-1})$ with some third-order error term $\frac{M}{6} f^{'''}(\xi^j_{a-1})(X^j_{a-1})^3$ for some $|\xi^j_{a-1}| < |X^j_{a-1}|$, i.e.,

$$
\begin{aligned}
p(X^j_{a-1}) &= f(0) + f^{'}(0) \, X^j_{a-1} + \frac{1}{2} f^{''}(0) \; (X^j_{a-1})^2 + \frac{1}{6} f^{'''}(0) \, (X^j_{a-1})^3 \\
f(X^j_{a-1}) - p(X^j_{a-1}) &= \frac{M}{6} f^{'''}(\xi^j_{a-1}) \, (X^j_{a-1})^3 \, .
\end{aligned} \tag{B.8}
$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\sum_{s<a<t}\beta_\ell\, f(X^j_{a-1})X^j_{a-1}\right] &= \mathbb{E}\left[\sum_{s<a<t}\beta_\ell\,(f(0)+f^{'}(0)\,X^j_{a-1}+\frac{1}{2}f^{''}(0)\,(X^j_{a-1})^2+\frac{M}{6}f^{'''}(\xi^j_{a-1})\,(X^j_{a-1})^3)\,X^j_{a-1}\right]\\ &= \mathbb{E}\left[\sum_{s<a<t}\beta_\ell(f^{'}(0)(X^j_{a-1})^2+\frac{M}{6}\,f^{'''}(\xi^j_{a-1})\,(X^j_{a-1})^4)\right]\\ &= \mathbb{E}\left[(t-s)\,\beta_\ell\,f^{'}(0))+\mathbb{E}\left(\sum_{s<a<t}\frac{M}{6}\,\beta_\ell\,f^{'''}(\xi^j_{a-1})\,(X^j_{a-1})^4\right)\right]\,.\end{aligned} \tag{B.9}$$

The term $\mathbb{E}[(t-s)\,\beta_\ell\,f^{'}(0)]$ only depends on f, so if we fix f, we write this term as $C^0\beta_\ell$. The left-hand side of (B.9) is the expectation of the Lévy-area, and hence by rearranging the equation we have

$$\mathbb{E}\left[A^{Lvy}_{i,j}-C^0\,\beta_\ell\right]=\frac{M}{6}\,\beta_\ell\,\mathbb{E}\left[\sum_{s<a<t}f^{'''}(\xi^j_{a-1})\,(X^j_{a-1})^4\right]\,. \tag{B.10}$$

Hence, if f is chosen so that the third derivative is small relative to C^0, and if the historical returns of assets are much smaller than 1, the Lévy-area provides a good approximation to the lead-lag coefficient β_ℓ. □

D Appendix 3: Alternative Ranking Methods

Throughout the paper, we identified leaders and followers based on a column average ranking on the lead-lag matrix. Here, we use SpringRank by De Bacco et al. (2018), Serial Ranking by Fogel et al. (2014), and SyncRank by Cucuringu (2015) to construct the same portfolios as in the previous sections of the paper.

In synthetic data simulations where we test the ability of these ranking algorithms to identify lead-lag relationships with various levels of noise, we observe that all three alternative ranking methods above can detect lead-lag relationships. In particular, SpringRank produces very similar rankings as those by the method we used in the main parts of this paper (i.e., ranking by column average); on the other hand, Serial Ranking and SyncRank are more sensitive to the level of noise, the size of the lag than SpringRank, and ranking by column average.

The observations in our synthetic data simulation suggest that ranking by column average is the best choice because of its interpretability and performance. Below we report the performance of alternative ranking methods on portfolios built with the top quantile of stocks in market capitalization on each trading day, which is the same data as that used to construct the portfolios reported in Panel A of Table 1.

Table 10: Performances of Alternative Ranking Methods

Panel A: SpringRank					
	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	19.7	7.13	0.48	2.37	16.9
Avg Cross-Cor	27.9	9.76	0.71	2.21	28.6
Lévy-area	24.9	8.82	0.59	2.38	24.7
Panel B: Serial Ranking					
	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	14.4	5.33	0.47	1.78	32.5
Avg Cross-Cor	13.5	5.04	0.48	1.68	27.5
Lévy-area	14.2	5.27	0.48	1.76	31.4
Panel C: SyncRank					
	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	12.7	4.77	0.49	1.57	28.33
Avg Cross-Cor	12.7	4.76	0.47	1.58	29.37
Lévy-area	11.5	4.33	0.48	1.41	30.48

Table 10 presents results for the same experiment as presented in Panel A of Table 1 with alternative specification of ranking methods. Panels A, B, C of table 10 show the results for data-driven portfolios when leaders and followers are identified with SpringRank, Serial Ranking, and SyncRank, respectively.

Table 10 shows that lead-lag relationships identified by the data-driven methods presented in this paper are robust under various ranking methods. While there is a difference between the results of Table 10 and Table 1, the portfolios still remain economically significant.

Portfolio performances using SpringRank are similar to the results in Table 1, while the results for Serial Ranking and SyncRank are not as good as those reported in Table 1 where the assets are ranked by column average. This is consistent with the results in the synthetic data simulations where SyncRank and Serial ranking are deemed less efficient than SpringRank and column average ranking in identifying lead-lag relationships.

E Appendix 4: Summary Statistics

Table 11: Average number of Firms Traded

Average number of firms traded						
	1963-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2023
nondurables	7.1	28.61	38.48	44.62	37.36	33.68
durables	10.5	16.7	17.5	20.11	14.8	11.94
manufacturing	38.26	70.21	75.36	81.49	71.65	49.63
energy	14.78	34.5	37.11	29.79	40.9	41.01
chemicals	11.52	20.85	25.98	22.69	17.99	21.36
business equipment	19.61	26.54	57.13	111.04	161.54	94.82
telecommunications	3.17	6.09	13.12	26.11	35.86	21.25
utilities	5.19	44.71	63.2	44.87	34.28	33.56
shops	4.99	19.87	39.45	70.37	77.5	67.2
healthcare	7.37	20.53	32.59	49.69	66.41	47.89
finance	2.99	33.97	66.67	96.6	105.23	74.45
others	17.62	34.03	48.06	53.11	74.91	94.1
all	142.9	356.59	514.64	650.49	738.42	592.56

Table 12: Mean (Volatility) of Price (Dollars) of Firms Traded

Mean (volatility) of price (dollars) of firms traded						
	1963-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2023
nondurables	49.23	34.72	41.82	43.85	43.08	72.72
	(22.79)	(21.47)	(19.96)	(22.93)	(35.27)	(71.86)
durables	65.44	33.08	40.46	43.55	42.66	65.52
	(30.73)	(21.29)	(22.58)	(18.8)	(24.11)	(43.24)
manufacturing	61.08	35.42	39.27	44.14	48.5	86.13
	(43.58)	(23.72)	(19.28)	(20.37)	(26.03)	(82.22)
energy	66.53	43.07	36.96	43.28	46.1	50.96
	(101.3)	(35.81)	(21.34)	(24.08)	(21.92)	(35.49)
chemicals	73.55	46.93	40.32	50.9	48.27	93.9
	(47.34)	(31.87)	(23.51)	(23.01)	(18.96)	(70.89)
business equipment	87.65	49.43	39.23	36.46	36.14	126.63
	(98.26)	(57.96)	(37.15)	(22.16)	(37.49)	(246.15)
telecommunications	55.8	34.59	54.01	52.26	31.67	67.37
	(24.14)	(18.91)	(55.79)	(72.23)	(21.97)	(107.26)
utilities	39.91	22.67	24.94	31.03	38.38	51.87
	(18.53)	(9.56)	(10.72)	(10.77)	(19.47)	(30.92)
shops	49.2	32.6	34.17	34.66	36.83	107.97
	(22.66)	(22.91)	(17.15)	(16.87)	(18.97)	(160.46)
healthcare	54.83	43.99	41.61	42.0	46.45	118.17
	(20.3)	(27.57)	(24.81)	(23.64)	(25.92)	(113.66)
finance	54.03	29.97	37.32	152.43	874.92	3352.74
	(29.95)	(16.26)	(17.97)	(2380.89)	(8546.08)	(30629.77)
others	57.35	29.82	32.53	38.36	47.58	122.28
	(29.39)	(21.91)	(19.13)	(23.12)	(57.12)	(307.53)
all	63.83	35.38	36.87	56.83	160.24	507.99
	(59.15)	(28.69)	(24.07)	(918.69)	(3239.38)	(10912.13)

Table 13: Mean (Volatility) of Daily Stock Returns (Percent)

Mean (volatility) of daily stock returns (percent)

	1963-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2023
nondurables	0.05 (1.64)	0.03 (2.1)	0.1 (2.11)	0.05 (1.87)	0.02 (2.12)	0.06 (1.92)
durables	0.04 (2.08)	0.0 (2.51)	0.08 (2.35)	0.08 (2.13)	-0.0 (2.97)	0.05 (2.28)
manufacturing	0.06 (1.83)	0.03 (2.25)	0.06 (2.28)	0.06 (2.24)	0.01 (3.05)	0.05 (2.26)
energy	0.07 (1.46)	0.05 (2.35)	0.05 (2.75)	0.02 (2.33)	0.04 (3.27)	0.03 (2.97)
chemicals	0.04 (1.46)	0.04 (2.02)	0.07 (2.15)	0.07 (1.73)	0.04 (2.46)	0.06 (1.95)
business equipment	0.1 (2.51)	0.02 (2.72)	0.04 (2.94)	0.1 (3.75)	-0.04 (4.87)	0.06 (2.5)
telecommunications	0.04 (1.77)	0.04 (2.2)	0.09 (2.14)	0.09 (2.75)	-0.07 (4.09)	0.05 (2.15)
utilities	0.05 (1.04)	0.04 (1.56)	0.07 (1.62)	0.04 (1.3)	0.03 (2.68)	0.05 (1.71)
shops	0.05 (1.7)	0.01 (2.61)	0.08 (2.47)	0.07 (2.67)	0.02 (2.8)	0.06 (2.36)
healthcare	0.06 (1.62)	0.02 (2.21)	0.07 (2.44)	0.07 (2.99)	0.03 (3.21)	0.06 (2.29)
finance	0.11 (2.08)	0.02 (2.3)	0.06 (2.16)	0.09 (2.2)	0.03 (3.01)	0.06 (2.18)
others	0.11 (2.37)	0.04 (2.82)	0.06 (2.64)	0.04 (2.75)	0.01 (3.61)	0.05 (2.93)
all	0.07 (1.93)	0.03 (2.3)	0.07 (2.36)	0.07 (2.64)	0.01 (3.57)	0.06 (2.42)

Table 14: Mean (Volatility) of Volume (Thousand Shares) Traded

Mean (volatility) of volume (thousand shares) traded

	1963-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2023
nondurables	15.89	24.53	168.69	504.54	2023.79	3574.74
	(25.75)	(31.66)	(330.68)	(903.89)	(2927.21)	(4489.26)
durables	30.76	40.15	263.06	766.86	4650.96	7021.39
	(34.79)	(51.52)	(427.55)	(1207.08)	(11069.88)	(17920.51)
manufacturing	16.83	29.08	154.66	423.7	2348.53	4474.21
	(19.65)	(38.63)	(263.95)	(743.63)	(5550.6)	(10580.17)
energy	17.61	39.63	232.54	579.24	2984.5	5660.98
	(25.34)	(50.18)	(421.07)	(655.04)	(4875.58)	(7504.94)
chemicals	12.09	31.23	175.51	430.95	2127.48	2876.99
	(14.81)	(36.93)	(314.65)	(590.71)	(2812.24)	(3522.38)
business equipment	27.47	38.93	244.0	1142.64	5988.36	6662.66
	(33.47)	(42.24)	(372.4)	(2231.66)	(12347.37)	(13189.96)
telecommunications	32.78	46.2	349.05	1096.04	6417.22	9018.87
	(40.32)	(61.41)	(813.14)	(1829.91)	(13353.32)	(15634.01)
utilities	10.33	24.93	182.96	317.95	1961.25	3572.77
	(9.76)	(33.29)	(886.38)	(411.83)	(2639.03)	(4873.39)
shops	17.81	33.44	211.61	582.8	2528.56	3402.12
	(25.55)	(45.71)	(334.01)	(872.33)	(3470.74)	(4607.55)
healthcare	11.89	32.18	207.85	662.55	2944.18	3691.78
	(14.47)	(39.65)	(278.74)	(900.29)	(5148.06)	(6760.03)
finance	17.18	29.42	182.96	457.31	3335.61	6541.94
	(30.36)	(40.17)	(290.99)	(704.05)	(11066.8)	(25201.02)
others	21.22	33.91	163.86	474.9	2307.92	4404.77
	(26.24)	(50.85)	(234.2)	(826.97)	(3657.88)	(8344.51)
all	19.42	31.8	196.67	634.27	3592.95	5019.04
	(26.0)	(42.36)	(449.73)	(1244.02)	(8640.04)	(12594.18)

Table 15: Mean (Volatility) of Daily Stock Turnovers

Mean (volatility) of daily stock turnovers

	1963-1969	1970-1979	1980-1989	1990-1999	2000-2009	2010-2023
nondurables	1.33 (3.02)	1.06 (1.59)	2.98 (4.93)	3.12 (5.28)	6.9 (9.99)	9.91 (13.75)
durables	2.68 (6.04)	1.44 (2.45)	3.55 (4.8)	3.82 (4.84)	10.31 (11.7)	12.32 (11.49)
manufacturing	2.03 (5.31)	1.65 (9.8)	3.34 (4.54)	5.2 (9.2)	13.8 (30.27)	13.2 (17.54)
energy	0.91 (2.39)	1.6 (2.71)	2.91 (3.81)	4.82 (7.11)	13.86 (13.07)	18.15 (19.28)
chemicals	0.97 (2.3)	1.17 (1.53)	3.16 (5.08)	3.12 (3.32)	7.43 (9.44)	9.11 (9.35)
business equipment	4.97 (8.81)	2.8 (5.05)	6.26 (8.09)	16.43 (22.37)	20.5 (25.05)	14.23 (21.51)
telecommunications	1.06 (1.6)	1.9 (4.49)	3.96 (6.78)	6.16 (12.19)	11.34 (16.35)	11.65 (13.62)
utilities	0.36 (0.33)	0.79 (0.98)	2.4 (7.41)	2.21 (2.93)	6.72 (7.56)	7.72 (6.64)
shops	1.55 (3.3)	1.9 (3.57)	4.29 (7.39)	6.75 (12.51)	12.71 (14.89)	15.24 (22.2)
healthcare	0.75 (1.06)	1.23 (2.64)	4.19 (6.22)	9.15 (15.82)	13.28 (20.7)	10.11 (17.85)
finance	2.95 (5.64)	1.36 (2.99)	3.41 (4.91)	4.03 (6.86)	8.52 (15.54)	9.28 (14.34)
others	3.78 (5.36)	2.3 (3.93)	4.32 (6.06)	6.97 (13.98)	16.5 (21.51)	18.04 (29.81)
all	2.32 (5.39)	1.57 (5.13)	3.73 (6.12)	7.11 (13.7)	13.61 (20.62)	13.24 (20.3)

F Appendix 5: Alternative Specification of Parameters in Lead-lag Portfolios

Table 16: Performances of Lead-lag Portfolios - Alternative Hyperparameters

Panel A: Global Lead-Lag Portfolios - 40% Leaders and Followers

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	18.36	6.70	0.44	2.43	18.9
Avg Cross-Cor	21.37	7.69	0.48	2.54	18.2
Lévy-area	24.26	8.62	0.51	2.70	17.8
Market Cap	0.08	0.03	0.46	0.01	91.75
Turnover	8.90	3.38	0.35	1.52	13.93

Panel B: Clustered Lead-lag Portfolios - Hermitian Clustering

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	15.73	5.80	0.38	2.38	12.46
Avg Cross-Cor	21.74	7.81	1.18	1.05	38.20
Lévy-area	24.31	8.64	0.39	3.43	9.19

Panel C: Clustered Lead-lag Portfolios - 40% Leader and Lagger

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	15.06	5.57	0.35	2.42	11.75
Avg Cross-Cor	17.23	6.31	0.40	2.52	20.39
Lévy-area	19.65	7.12	0.38	2.99	11.51
Industry	6.64	2.55	0.38	1.06	26.9

Panel D: Clustered Lead-lag Portfolios - 40% Leader and Lagger, Hermitian Clustering

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	16.08	5.92	0.37	2.51	14.77
Avg Cross-Cor	20.98	7.56	1.14	1.05	39.49
Lévy-area	22.70	8.12	0.37	3.45	10.07

Panel E: Global Lead-lag Portfolios - 20% Leader and Lagger, 30 Day Look-back Window

	Compound Return (%)	Return (bps/day)	Volatility (%)	Sharpe Ratio	Max Drawdown (%)
Max Cross-Cor	29.60	10.29	0.74	2.21	27.48
Avg Cross-Cor	22.72	8.13	0.51	2.19	16.91
Lévy-area	23.64	8.42	0.59	2.61	16.72

G Appendix 6: Composition of Lead-lag Portfolios

Table 17: Compositions of Lead-Lag Portfolios

Panel A: Characteristics for Leaders and Followers in the MaxCor Portfolio

	Leader Avg Percentile (%)	Follower Avg Percentile (%)	Permutation Test P-value
Market Cap	57.48	58.66	0.041
Volume	53.95	53.77	0.823
Turnover	39.09	37.95	0.145
Return	50.16	50.19	0.588
Price	57.22	58.87	0.045
Shares Outstanding	63.16	65.28	0.013

Panel B: Characteristics for Leaders and Followers in the AvgCor Portfolio

	Leader Avg Percentile (%)	Follower Avg Percentile (%)	Permutation Test P-value
Market Cap	57.99	59.26	0.019
Volume	53.67	53.61	0.936
Turnover	38.77	36.53	0.003
Return	50.17	50.19	0.759
Price	57.74	59.61	0.019
Shares Outstanding	63.28	66.35	<0.0001