

Robust Detection of Lead-Lag Relationships in Lagged Multi-Factor Models

Yichi Zhang^{1,4}, Mihai Cucuringu^{1,2,4,7}, Alexander Y. Shestopaloff^{5,6}, Stefan Zohren^{3,4,7}

¹*Department of Statistics, University of Oxford*

²*Mathematical Institute, University of Oxford*

³*Department of Engineering, University of Oxford*

⁴*Oxford-Man Institute of Quantitative Finance, University of Oxford*

⁵*School of Mathematical Sciences, Queen Mary University of London*

⁶*Department of Mathematics and Statistics, Memorial University of Newfoundland*

⁷*The Alan Turing Institute*

September 20, 2023

Abstract

In multivariate time series systems, key insights can be obtained by discovering lead-lag relationships inherent in the data, which refer to the dependence between two time series shifted in time relative to one another, and which can be leveraged for the purposes of control, forecasting or clustering. We develop a clustering-driven methodology for robust detection of lead-lag relationships in lagged multi-factor models. Within our framework, the envisioned pipeline takes as input a set of time series, and creates an enlarged universe of extracted subsequence time series from each input time series, via a sliding window approach. This is then followed by an application of various clustering techniques, (such as K-means++ and spectral clustering), employing a variety of pairwise similarity measures, including nonlinear ones. Once the clusters have been extracted, lead-lag estimates across clusters are robustly aggregated to enhance the identification of the consistent relationships in the original universe. We establish connections to the multireference alignment problem for both the homogeneous and heterogeneous settings. Since multivariate time series are ubiquitous in a wide range of domains, we demonstrate that our method is not only able to robustly detect lead-lag relationships in financial markets, but can also yield insightful results when applied to an environmental data set.

Keywords: High-dimensional time series; Lead-lag relationships; Unsupervised learning; Clustering; Financial markets

Contents

1	Introduction	3
2	Connections to Multireference Alignment (MRA)	5
3	Model setup	6
3.1	Description	7
3.2	Notation	7
4	Methodology	8
5	Synthetic data experiments	12
5.1	Setup	12
5.2	Results	13
5.3	Simulation	15
6	Financial data experiments	18
6.1	Data description	18
6.2	Data pre-processing	19
6.3	Benchmark	19
6.4	Trading strategies	20
6.5	Performance evaluation	21
6.6	Results	23
7	Robustness analysis	28
8	CO₂ emissions data	33
8.1	Data description	33
8.2	Pipeline	33
8.3	Results	33
9	Conclusion	35
A	Appendix	41
A.1	Synthetic data experiment: spectral clustering	41
A.2	CO ₂ emissions plots	42
A.3	Futures data set details	44

1 Introduction

When observed over time, natural physical systems frequently produce data recorded as high-dimensional, nonlinear time series, which are ubiquitous in a wide range of domains. Numerous contributions, covering different aspects of their analysis, have been made. For example, Cont (2001) discussed financial time series with a focus on various statistical properties, including distributional and tail properties, extreme fluctuations, etc. Cartea, Donnelly, and Jaimungal (2018) constructed a metric for assessing volume imbalance in the limit order book sourced from the Nasdaq exchange, and demonstrated it is a good predictor of the sign of the next market order. Cao, Chen, and Hull (2020) utilized neural networks to analyze volatility surface movements based on daily call options data on the S&P 500 index. Drinkall, Zohren, and Pierrehumbert (2022) introduced a new technique that integrates time-series language models based on transformers into the field of infectious disease modelling. Sokolov et al. (2022) proposed a supervised machine learning framework for analyzing the impact of environmental, social and governance (ESG) factors on fund flows in US-domiciled equity mutual funds, and assessed whether sustainability has excess predictive power on fund-level flows as compared to benchmark driven by non-ESG factors. To detect change-points in nonlinear time series regression, Cui, Yang, and Zhou (2021) utilized a density-weighted anti-symmetric kernel function and identified the presence of change-points within the state domain, rather than the time domain. Furthermore, Vuletić, Prenzel, and Cucuringu (2023) examined the feasibility of utilizing generative adversarial networks to forecast financial time series probabilistically, and to learn from asset co-movements while addressing problems associated with mode collapse.

Key insights regarding high-dimensional time series can be obtained by discovering latent structures. An example of a latent structure is lead-lag relationships, which are widely observed and found in the realms of finance [Tolikas (2018), Buccheri, Corsi, and Peluso (2021), Miori and Cucuringu (2022), Bennett, Cucuringu, and Reinert (2022), Albers et al. (2021), Ito and Sakemoto (2020), Yao and H.-Y. Li (2020), Y. Li et al. (2021)], the environment [De Luca and Pizzolante (2021), Wu et al. (2010)], and biology [Runge et al. (2019)]. For example, Miori and Cucuringu (2022) explored lead-lag relationships within data-driven macroeconomic regimes by clustering the performance of diverse asset class indices in time relative to each other. Bennett, Cucuringu, and Reinert (2022) constructed a directed network for encoding pairwise lead-lag relationships between time series of equity prices in the US equity market, in order to detect pairs of lead-lag clusters that exhibited a high pairwise directed flow imbalance. Another finance-related application study was derived from Albers et al. (2021), who analyzed the existence and strength of lead-lag relationships between pairs of Bitcoin markets. The importance and potentially high impact of this problem are broadly recognized; however, to date, there has been limited progress on the robust detection of lead-lag relationships in high-dimensional time series.

Clustering is widely used as part of the analysis of time series [Zolhavarieh, Aghabozorgi, and Teh (2014)]. For example, Lu, Reinert, and Cucuringu (2023) developed a similarity measure between equities based on co-occurrence of trades from Lu, Reinert, and Cucuringu (2022) and used spectral clustering algorithms [Shi and Malik (2000), Ng, Jordan, and Weiss (2001), Cucuringu,

Davies, et al. (2019)] to detect dynamic communities within US equity markets. In particular, subsequence time series clustering, as one of time series clustering, groups similar subsequences into the same cluster. This type of clustering is used for detecting structures or patterns, and is typically used as a subroutine in rule discovery [Uehara and Shimada (2002), Sarker et al. (2003)], indexing [C.-S. Li, Yu, and Castelli (1998), Radhakrishnan, Wilson, and Loizou (2000)], classification [Cotofrei and Stoffel (2002), Vijay and Nanda (2023)], prediction [Tino, Schittenkopf, and Dorffner (2000), Schittenkopf, Tiño, and Dorffner (2002)], and anomaly detection [Yairi, Kato, and Hori (2001), Dong et al. (2023)]. However, despite its wide use, Lin, Keogh, and Truppel (2003) reported a surprising result: clustering subsequences extracted from a single time series via a sliding window is meaningless. They illustrated that all previous results involving clustering subsequences of a single time series were inaccurate since the resultant cluster centers appeared to be a form of sine waves, regardless of the initial patterns in the input data [Keogh and Lin (2005)]. Afterwards, several lines of work [Madicar et al. (2013), Rakthanmanon et al. (2011), Rodpongpun, Niennattrakul, and Ratanamahatana (2012)] proposed solutions to the aforementioned meaningless results of subsequence time series clustering, and achieved meaningful time series clusters. In contrast, we take the approach of clustering every subsequence, extracted via a sliding window from a *set* of time series.

In this paper, we develop a clustering-driven methodology to enable the robust detection of lead-lag relationships in high-dimensional time series. In the proposed pipeline, we are given as input a set of n time series. With this in mind, we create an enlarged universe \mathcal{U} of $N = n \times h$ time series by extracting h subsequence time series via a sliding window from each input time series. To this enlarged universe \mathcal{U} , we apply various clustering techniques (e.g., K-means++ and spectral clustering) by employing various pairwise similarity measures between subsequence time series. The underlying clusters are then leveraged for the purpose of discovering the latent lead-lag relationships. In essence, the clustering step can be construed as an initial denoising step, in which we group together data relevant to estimating a subset of the lead-lag effects. Once the clusters have been extracted, the lead-lag estimates within each cluster are aggregated across clusters to enhance the identification of consistent relationships in the original universe. Our main contributions are summarized as follows.

Summary of main contributions.

1. We introduce a computationally scalable pipeline for the robust detection of lead-lag relationships in high-dimensional time series.
2. We demonstrate that our proposed methodology can reliably detect lead-lag relationships in a range of factor model-based simulated high-dimensional time series.
3. In a financial market setting, we leverage the detected lead-lag relationships to construct a profitable trading strategy and show that our method outperforms the benchmark in most of cases.
4. We apply our method to a data set of CO2 emissions and demonstrate that it achieves results consistent with the literature.

Paper outline. This paper is organized as follows. We first discuss connections between lead-lag detection and the multireference alignment (MRA) problem in Section 2. Section 3 describes the *lagged multi-factor model* and establishes the notation used in this paper. Section 4 describes our proposed method for detecting lead-lag relationships in high-dimensional time series. In Section 5, we validate our method on synthetic data sets from the lagged multi-factor model, and show that our method can reliably recover them. Since high-dimensional time series data are ubiquitous in finance, our application domain mainly targeted financial time series. Within this domain, each time series corresponds to the excess return time series of a specific financial instrument. We explore lead-lag relationships in the US equity, ETF, and futures markets in Section 6. We then move on to a robustness analysis in Section 7. To demonstrate the broader scope and applicability of our proposed algorithm, we also consider an application from the environmental sciences, looking at a data set of CO2 emissions in Section 8. Finally, we summarize our findings in Section 9, and discuss possible future research directions.

2 Connections to Multireference Alignment (MRA)

Our proposed framework for lead-lag detection has strong connections to the multireference alignment (MRA) problem, which we briefly describe here. MRA aims to estimate one signal from n cyclically and noisy shifted copies of itself [Bandeira et al. (2014)]. In the **homogeneous** setting, let $x \in \mathbb{R}^L$ be the unknown signal and let R_r be the cyclic shift operator by r . We are given n measurements of the form

$$y_j = R_{r_j}x + \varepsilon_j, \quad j = 1, \dots, n, \quad (1)$$

where the $\varepsilon_j \sim \mathcal{N}(0, \sigma^2 I)$ are i.i.d. white Gaussian noise. The goal of MRA is to then estimate the unknown signal x , up to a shift, in a regime with high-noise in which the shifts r_j are also unknown.

In the **heterogeneous** setting that arises in areas such as cryo-electron microscopy, there are

k distinct signals $x_1, \dots, x_k \in \mathbb{R}^L$ to be estimated. Each of the available n observations is derived from one of these k signals, yet the correspondence is unknown to the user. The model can be written as follows

$$y_j = R_{r_j} x_{v_j} + \varepsilon_j, \quad j = 1, \dots, n, \quad (2)$$

where the classes v_j and the shifts r_j are unknown, while the ε_j are i.i.d. Gaussian noise of variance σ^2 as before. The goal is to estimate the signals x_1, \dots, x_k , up to shifts and ordering. Figure 1 illustrates an example of nine observations x_i, y_i, z_i ($i = 1, 2, 3$), each derived from one of the three signals $x, y, z \in \mathbb{R}^6$ with a noisy cyclic shift.

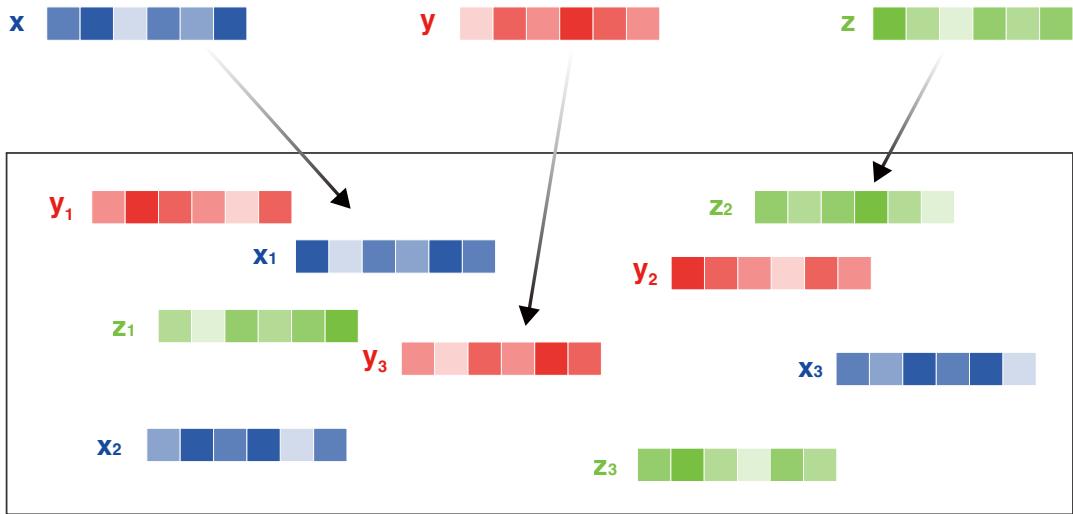


Figure 1: Nine observations derived from one of three signals with noisy cyclic shifts.

The work of [Perry et al. (2019)] introduced the first known procedure to provably achieve signal recovery in a low signal-to-noise ratio (SNR) regime for heterogeneous MRA. Moreover, [Boumal et al. (2018)] considered the application of MRA to the 2D class averaging problem in cryo-EM, aiming to achieve classification, alignment and averaging concurrently in a single pass, without using any of these steps explicitly. The authors of [Boumal et al. (2018)] also proposed to use signal characteristics that are invariant under translations, and used them to recover the original signal. In the spirit of MRA, we will consider an analogous single-pass approach for robust detection of lead-lag relationships.

3 Model setup

In this section, we will introduce the standard and lagged versions of the *multi-factor model*, which we will assume as a model for our time series data. The lagged multi-factor model will be used to validate our method on a synthetic data scenario before proceeding to the real-world applications. The gist of these models is to represent a time series as a (noisy) superposition of factors with varying exposures to each factor. We will also summarize the notation used in this paper.

3.1 Description

We first recall the standard multi-factor model for a multivariate time series

$$X_i^t = \sum_{j=1}^k B_{ij} f_j^t + \epsilon_i^t \quad i = 1, \dots, n; \quad t = 1, \dots, T, \quad (3)$$

where X_i^t is the time series i (e.g., the excess return of a financial asset) at time t , k is the number of factors, B_{ij} is the exposure of time series i to factor j , f_j^t is the factor j at time t , and ϵ_i^t is the noise at time t , with variance σ^2 . Furthermore, n is the number of time series, and T is the total number of time steps.

In this paper, we focus on the *lagged* version of the multi-factor model, which can be written as

$$X_i^t = \sum_{j=1}^k B_{ij} f_j^{t-L_{ij}} + \epsilon_i^t \quad i = 1, \dots, n; \quad t = 1, \dots, T, \quad (4)$$

where the only difference compared to the standard multi-factor model is the addition of L_{ij} , the lag at which time series i is exposed to factor j . Thus, $f_j^{t-L_{ij}}$ is the value of factor j at time $t - L_{ij}$.

We introduce two main settings in the lagged multi-factor model (4), as follows.

- **Single Membership:** Each time series has a lagged exposure to a single factor. We consider the following two main categories.
 - **Homogeneous Setting:** The model only has one factor, i.e. $k = 1$.
 - **Heterogeneous Setting:** The model has more than one factor, i.e. $k \geq 2$. However, each time series is exposed only to a single factor.
- **Mixed Membership:** Each time series is allowed to have a lagged exposure to more than one factor, hence it is mixed. The model contains at least two factors, i.e. $k \geq 2$.

In this paper, our goal will be the inference of L_{ij} in the lagged multi-factor model, focusing on the single membership setting. We do not focus on the inference of the unknown coefficient matrix B and factors f . As shown later, the inference of L_{ij} alone is of practical importance in certain applications, e.g. finance. We leave the mixed membership setting for future research.

3.2 Notation

We first introduce the definition of a time series, subsequence time series and sliding window.

- **Time Series:** A time series $X_i = X_i^1, \dots, X_i^T$ is an ordered set of T real-valued variables.
- **Subsequence Time Series (STS):** Given a time series $X_i = X_i^1, \dots, X_i^T$ of length T , a STS Y_i^z of time series X_i is a sample of length $q < T$ of contiguous positions from X_i starting at z , that is, $Y_i = X_i^z, \dots, X_i^{z+q-1}$ where $1 \leq z \leq T - q + 1$.
- **Sliding Window:** Given a set of time series X_i^t ($i = 1, \dots, n; t = 1, \dots, T$), and a user-defined STS length of q , an enlarged universe matrix U of N STS can be built by sliding a window shifted by s across X_i^t . The size of the enlarged universe matrix U is N by q .

We summarize the remaining notation conventions used in the paper in Table 1.

Table 1: Variables and their description as used in our paper.

Variables	Description
k	Total number of factors
K	Total number of clusters
n	Total number of time series
N	Total number of STS in the enlarged universe
m	Number of lags
M	Maximum number of lags
t	Time
T	Length of time series
q	Length of STS
h	Number of STS in each time series
\mathcal{U}	Universe set, it contains N STS with length of q
l	Length of sliding window
σ	Noise
s	Value of sliding window shift
δ	Forward looking horizon windows
p	Past looking horizon windows
$X_{n \times T}$	Time series matrix, where X_i^t is the time series i at time t
X_i	Time series i
Y_i^z	STS from time series i starting at z
$B_{n \times k}$	Loading matrix, where B_{ij} is the exposure of time series i to factor j
$f_{k \times 1}$	Vector of returns of the k factors
f_j^t	Value of returns of factor j at time t
$\epsilon_{n \times 1}$	Vector of noise
$L_{n \times k}$	Lag matrix, where L_{ij} is the lag of time series i to factor j
$U_{N \times q}$	Universe matrix, where U_{ij} is the STS i at time j
$C_{N \times N}$	Correlation matrix, where C_{ij} is the correlation of STS i and j
$P_{N \times N}$	Sparse matrix, where P_{ij} is assigned the weight of edge that connects STS i and j
$G_{N \times N}$	Similarity matrix, where G_{ij} is the similarity of STS i and j
ϕ_d	Cluster d
$\{X_i, X_j\}$	Pair of time series i and j
$\Delta_d\{X_i, X_j\}$	Set of the relative lags between all pairs of time series i and j in cluster d
$V_{i \times j}$	Voting matrix, where $v_{i \times j}$ is the number of lags between STS of i and j in the same cluster
$\gamma\{X_i, X_j\}$	The estimated value of the relative lags between all pairs of time series i and j in all clusters
$\Gamma_{n \times n}$	Lead-lag matrix, where Γ_{ij} is the lead or lag value between time series i and j
$E_{n \times n}$	Error matrix, where E_{ij} is the error of lead or lag value between time series i and j
$\Psi_{n \times n}$	Ground truth lead-lag matrix, where Ψ_{ij} is the ground truth of lead or lag value between time series i and j

4 Methodology

In this section, we present our methodology to infer unknown lags L_{ij} in detail, before applying it to synthetic, financial, and environmental data set.

We consider a set of time series $X_{n \times T}$ as our input. STS with length q are extracted from each time series X_i by a sliding window shifted by s . Therefore, the number of STS for each time series in the ensemble is $h = \frac{T-q}{s} + 1$, and the total number of STS from $X_{n \times T}$ is $N = n \cdot h$. An enlarged universe matrix $U_{N \times q}$ is constructed by collecting all STS, which is shown in Figure 2. Note that $U_{N \times q}$ contains exactly the same information as $X_{n \times T}$.

After the STS extraction step, we employ various clustering techniques to group similar STS

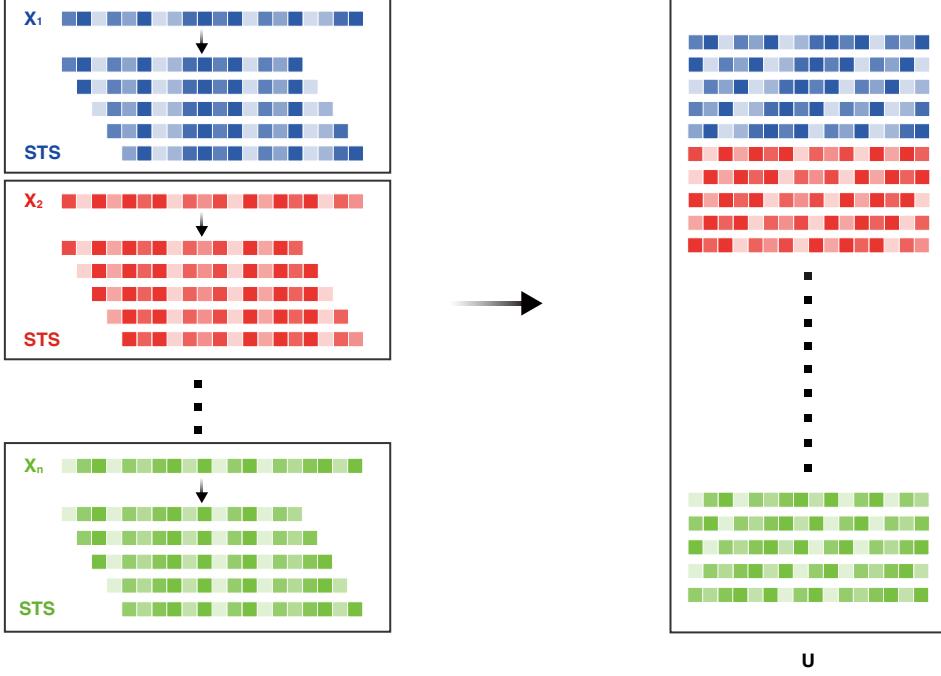


Figure 2: Left: STS are extracted from each time series via a sliding window approach. Right: An enlarged universe matrix $U_{N \times q}$ is constructed by stacking all the extracted STS from the input set of time series $X_{n \times T}$.

in the same cluster. One way is to apply K-means++ (KM) clustering [Arthur and Vassilvitskii (2006)]. We follow this approach in Algorithm 1, which initializes the cluster centers before proceeding with the standard K-means algorithm. With the KM initialization, a solution that is within $O(\log(k))$ of the optimal standard K-means solution is guaranteed.

Algorithm 1 : K-means++ Clustering

Input: Universe matrix $U_{N \times q}$, the total number of clusters K .

Output: Clusters $\phi_1, \phi_2, \dots, \phi_K$.

- 1: Randomly select an initial center ϕ_1 from U .
- 2: **Repeat** for $i \in 1, 2, \dots, K - 1, K$. Select the next center $\phi_i = x \in U$ with the probability

$$P(x) = \frac{D(x)^2}{\sum_{x' \in U} D(x')^2},$$

where x' is the closest center that has already been chosen and $D(x')$ is the distance to that center.

- 3: Continue with the standard K-means algorithm.
-

Algorithm 2 : Spectral Clustering

Input: Similarity matrix $G_{N \times N}$, the total number of clusters K .

Output: Clusters $\phi_1, \phi_2, \dots, \phi_K$.

- 1: Compute normalized Laplacian L .
 - 2: Compute the eigenvectors v_1, v_2, \dots, v_K corresponding to K smallest eigenvalues of L .
 - 3: Construct matrix $M \in \mathbb{R}^{N \times K}$ with v_1, v_2, \dots, v_K as columns.
 - 4: Form matrix $\tilde{M} \in \mathbb{R}^{N \times K}$ by normalizing row vectors of M to norm 1.
 - 5: Apply K-means clustering to assign rows of \tilde{M} to clusters $\phi_1, \phi_2, \dots, \phi_K$.
-

An alternative method we considered is that of spectral (SP) clustering, described in Algo-

rithm 2. Note that if the size $U_{N \times q}$ is prohibitively large due to a large number n of input time series, in order to speed up the computation of eigenvalues and eigenvectors, one could apply the K-nearest neighbours algorithm (KNN) on $U_{N \times q}$, which leads to the sparse matrix $P_{N \times N}$. The entries of the similarity matrix $G_{N \times N}$ are computed using a Gaussian kernel between neighbours from $P_{N \times N}$.

$$G_{ij} = \exp\left(-\|Y_i^a - Y_j^b\|^2 / (2\sigma^2)\right), \quad (5)$$

where the parameter $\sigma = 1/N$ by default. Finally, we apply SP clustering on $G_{N \times N}$.

We denote time series i and j by X_i and X_j respectively, and a pair consisting of them by $\{X_i, X_j\}$. For each cluster ϕ_d ($d = 1, \dots, K$), we consider all possible pairwise relative lags of X_i and X_j . We denote this set by $\Delta_d\{X_i, X_j\}$ and obtain it by computing the difference between starting indices of STS from X_i and X_j appearing in cluster d . Let $S_i = \{s_i^1, \dots, s_i^k\}$ and $S_j = \{s_j^1, \dots, s_j^l\}$ be the sets of STS starting indices of, respectively, X_i and X_j . To avoid double counting, we consider only $i < j$. This amount to the following

$$\Delta_d\{X_i, X_j\} = \Delta_d\{(Y_i^{s_i^1}, Y_j^{s_j^1}), \dots, (Y_i^{s_i^k}, Y_j^{s_j^l})\} = \{s_i^1 - s_j^1, \dots, s_i^k - s_j^l\}. \quad (6)$$

We then aggregate $\Delta_d\{X_i, X_j\}$ across all clusters and estimate the relative lag of $\{X_i, X_j\}$ by considering the mode or median of the resulting set

$$\gamma\{X_i, X_j\} = \begin{cases} \text{Mode}(\bigcup_{d=1}^K \Delta_d\{X_i, X_j\}) & \text{Mode estimation} \\ \text{Median}(\bigcup_{d=1}^K \Delta_d\{X_i, X_j\}) & \text{Median estimation} \end{cases} \quad (7)$$

Table 2 shows the example of two time series X_1 and X_2 , for which the ground truth lag value is 3. We extract eleven STS from each time series by a sliding window, and then perform KM clustering on these STS, setting the value of K to 11. In each cluster, we calculate the relative lags of STS $\Delta_d\{X_1, X_2\}$ ($d = 1, \dots, 11$). After that, we aggregate lags across all clusters together, as $\bigcup_{d=1}^{11} \Delta_d\{X_1, X_2\} = \{-7, 3, 3, 3, 3, 3, 3, 3, 3, -10, -9\}$. Figure 3 displays the histogram of the relative lags of STS from X_1 and X_2 .

Table 2: Example of calculating the relative lags of STS in each cluster from two time series.

Cluster	Subsequence	Lag
ϕ_1	(Y_1^9, Y_2^2)	-7
ϕ_2	(Y_1^7, Y_2^{10})	3
ϕ_3	(Y_1^1, Y_2^4)	3
ϕ_4	(Y_1^2, Y_2^5)	3
ϕ_5	(Y_1^0, Y_2^3)	3
ϕ_6	(Y_1^6, Y_2^9)	3
ϕ_7	(Y_1^3, Y_2^6)	3
ϕ_8	(Y_1^4, Y_2^7)	3
ϕ_9	(Y_1^5, Y_2^8)	3
ϕ_{10}	$(Y_1^10, Y_2^0), (Y_1^10, Y_2^1)$	-10, -9
ϕ_{11}	Y_1^8	NaN

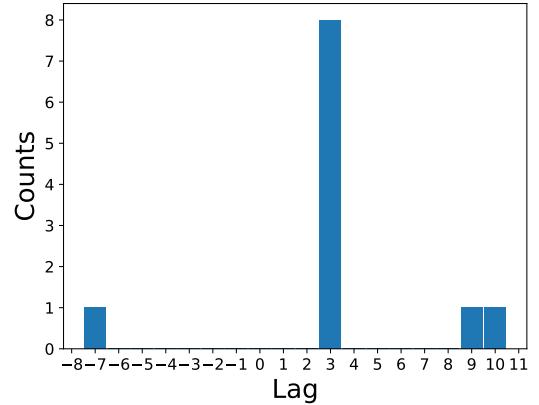


Figure 3: Histogram of the relative lags of STS from two time series.

By considering the mode or median of $\bigcup_{d=1}^{11} \Delta_d\{X_1, X_2\}$, we arrive at $\gamma\{X_1, X_2\} = 3$. From Table 2 and Figure 3, we observe that even though there are outliers $\{-7, -10, -9\}$ in $\bigcup_{d=1}^{11} \Delta_d\{X_1, X_2\}$, we are still able to correctly recover the ground truth value of 3.

A refinement of the algorithm is to calculate a *voting matrix* $V_{n \times n}$ by counting the sum of the number of lags between STS from X_i and X_j in the same clusters. We can then set a voting threshold denoted by θ to filter out a small number of counts. This is motivated by the reasoning that unless STS corresponding to a pair of time series are consistently clustered together across different clusters, the resulting lead-lag estimate is unlikely to be accurate. This altogether amounts to

$$V_{ij} = \begin{cases} |\bigcup_{d=1}^K \Delta_d\{X_i, X_j\}| & \text{if } |\bigcup_{d=1}^K \Delta_d\{X_i, X_j\}| \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where $|\cdot|$ counts the number of elements in the $\bigcup_{d=1}^K \Delta_d\{X_i, X_j\}$.

Finally, the lead-lag matrix $\Gamma_{n \times n}$ is built by

$$\Gamma_{ij} = \begin{cases} \gamma\{X_i, X_j\} & \text{if } V_{ij} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

We summarize the above procedures in Algorithm 3, which is our main algorithm.

Algorithm 3 : Lead-lag Relationship Detection Algorithm

Input: Time series matrix $X_{n \times T}$.

Output: Lead-lag matrix $\Gamma_{n \times n}$.

- 1: STS Y_i^p are extracted from each time series X_i by a sliding window.
 - 2: An enlarged universe matrix $U_{N \times q}$ is created in Step 2.
 - 3: Apply KNN to create the sparse matrix $P_{N \times N}$, and the similarity matrix $G_{N \times N}$ is then only computed using the Gaussian kernel between neighbours from $P_{N \times N}$.
 - 4: Clusters are extracted by performing KM clustering on $U_{N \times q}$ or SP clustering on $G_{N \times N}$. For each cluster, record lags between every pair of time series $\{X_i, X_j\}$.
 - 5: For each pair of time series $\{X_i, X_j\}$, we calculate the voting matrix V_{ij} by counting lags between STS from X_i and X_j across all clusters. We use a voting threshold θ to filter small counts.
 - 6: Calculate the lead-lag matrix $\Gamma_{n \times n}$ by considering *mode* or *median* of lags based on the voting matrix $V_{n \times n}$.
-

5 Synthetic data experiments

The purpose of our synthetic data experiments is to assume that our data is generated by a multi-factor model with known ground truth lead-lag matrix L , and then validate the performance of our proposed algorithms under different scenarios.

5.1 Setup

As noted earlier, our focus is the single membership setting. We generate synthetic data from the lagged multi-factor model (4) with $k = \{1, 2, 3\}$ factors. We let $M = 5$, $T = 100$ and $n = 6$. The factors f and errors ϵ are assumed to be i.i.d. $\mathcal{N}(0, 1)$. We define B and L as follows:

Table 3: Top row: Loading matrix B . Bottom row: Lag matrix L .

	Homogeneous Setting	Heterogeneous Setting
B	$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$
L	$\begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 4 & 0 \\ 0 & 0 \\ 0 & 2 \\ 0 & 4 \end{bmatrix}$
	$k = 1$	$k = 2$
		$k = 3$

When estimating the lead-lag matrix, we use a sliding window of length $q = 90$ and a shift of $s = 1$. After estimating the lead-lag matrix, we calculate the error matrix E to evaluate the performance of the method, which we denote as

$$E_{n \times n} = \Gamma_{n \times n} - \Psi_{n \times n}, \quad (10)$$

where $\Gamma_{n \times n}$ is the estimated lead-lag matrix, and $\Psi_{n \times n}$ is the ground truth lead-lag matrix, which can be obtained from $L_{n \times k}$.

5.2 Results

We first explore the data by computing the Pearson and distance correlations between STS in the universe matrix $U_{N \times q}$, for scenarios with different k . Figure 4 shows the correlation between STS in the two settings, homogeneous and heterogeneous. Each heatmap represents similarities between the STS, with a darker shade of red corresponding to a higher similarity.

For the homogeneous setting ($k = 1$), we note that all 6×6 blocks, with each block corresponding to a different time series, have groups of red pixels stretching diagonally from the top left to the bottom right. This reflects a high similarity between pairs of STS starting at different initial time points. Thus, for example, time series 1 and time series 2 have highest correlation between STS at a lag of 1, which reflects the ground truth. For the heterogeneous setting ($k \in \{2, 3\}$), we note that only the diagonal blocks show high similarities between STS, which reflects the structure of B . Again, we note a high similarity between STS at the ground truth lag value. Our proposed method clusters these similar STS together to estimate the lag between the corresponding time series.

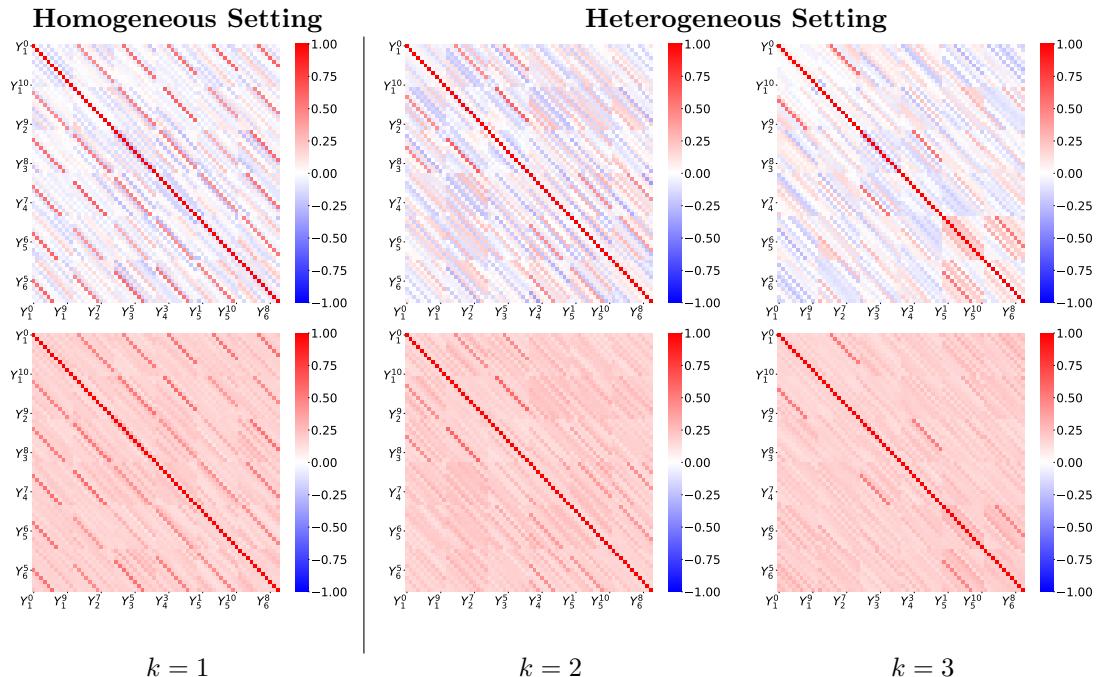


Figure 4: Top panel: Heatmap of enlarged universe $U_{N \times q}$ by Pearson correlation. Bottom panel: Heatmap of enlarged universe $U_{N \times q}$ by distance correlation.

We next follow the methodology described in Section 4 to estimate the lead-lag matrix. In what follows, we use KM clustering to cluster STS. We set the number of clusters for a model with k factors to $11 \cdot k$. A similar result, utilizing SP clustering, is shown in Appendix A.1 Figures [16, 17]. We first consider not setting a voting threshold ($\theta = 1$). Figure 5 contains the number of votes associated with each time series pair. The top plots denote the voting matrix, the middle plots represent the error matrix for the mode estimation, and the bottom plots correspond to the

median estimation. For the homogeneous case, we can observe that the voting matrix consists of all 0s across the diagonal and is non-zero for all other entries. The error matrices based on mode and median estimation consist of all 0s, which indicates our algorithm fully recovers the lead-lag relationships between these six time series. However, for the heterogeneous setting, we can observe that the error matrices have some non-zero values.

K-means++ clustering

Homogeneous Setting							Heterogeneous Setting													
TS 1	0	15	13	11	11	11	TS 1	0	9	7	4	4	3	TS 1	0	8	1	0	0	0
TS 2	15	0	13	11	11	12	TS 2	9	0	9	2	2	2	TS 2	8	0	3	3	2	1
TS 3	13	13	0	11	11	12	TS 3	7	9	0	2	1	1	TS 3	1	3	0	9	1	0
TS 4	11	11	11	0	11	11	TS 4	4	2	2	0	11	9	TS 4	0	3	9	0	3	1
TS 5	11	11	11	11	0	13	TS 5	4	2	1	11	0	9	TS 5	0	2	1	3	0	10
TS 6	11	12	12	11	13	0	TS 6	3	2	1	9	9	0	TS 6	0	1	0	1	10	0
TS 1	0	0	0	0	0	0	TS 1	0	0	0	7	5	3	TS 1	0	0	-2	0	0	0
TS 2	0	0	0	0	0	0	TS 2	0	0	0	9	7	5	TS 2	0	0	1	-2	-8	-8
TS 3	0	0	0	0	0	0	TS 3	0	0	0	-10	-7	-9	TS 3	2	-1	0	0	-6	0
TS 4	0	0	0	0	0	0	TS 4	-7	-9	10	0	0	0	TS 4	0	2	0	0	-3	-6
TS 5	0	0	0	0	0	0	TS 5	-5	-7	7	0	0	0	TS 5	0	8	6	3	0	0
TS 6	0	0	0	0	0	0	TS 6	-3	-5	9	0	0	0	TS 6	0	8	0	6	0	0
TS 1	0	0	0	0	0	0	TS 1	0	0	0	7	5	3	TS 1	0	0	-2	0	0	0
TS 2	0	0	0	0	0	0	TS 2	0	0	0	9	7	5	TS 2	0	0	1	-2	-6	-8
TS 3	0	0	0	0	0	0	TS 3	0	0	0	-7	-7	-9	TS 3	2	-1	0	0	-6	0
TS 4	0	0	0	0	0	0	TS 4	-7	-9	7	0	0	0	TS 4	0	2	0	0	-3	-6
TS 5	0	0	0	0	0	0	TS 5	-5	-7	7	0	0	0	TS 5	0	6	6	3	0	0
TS 6	0	0	0	0	0	0	TS 6	-3	-5	9	0	0	0	TS 6	0	8	0	6	0	0

$k = 1$							$k = 2$							$k = 3$						
TS 1	0	0	0	0	0	0	TS 1	0	0	0	7	5	3	TS 1	0	0	-2	0	0	0
TS 2	0	0	0	0	0	0	TS 2	0	0	0	9	7	5	TS 2	0	0	1	-2	-6	-8
TS 3	0	0	0	0	0	0	TS 3	0	0	0	-7	-7	-9	TS 3	2	-1	0	0	-6	0
TS 4	0	0	0	0	0	0	TS 4	-7	-9	7	0	0	0	TS 4	0	2	0	0	-3	-6
TS 5	0	0	0	0	0	0	TS 5	-5	-7	7	0	0	0	TS 5	0	6	6	3	0	0
TS 6	0	0	0	0	0	0	TS 6	-3	-5	9	0	0	0	TS 6	0	8	0	6	0	0

Figure 5: Top panel: Voting matrix without voting threshold ($\theta = 1$). Middle panel: Error matrix based on mode estimation without the voting threshold ($\theta = 1$). Bottom panel: Error matrix based on median estimation without the voting threshold ($\theta = 1$).

We extend our experiment by setting the voting threshold $\theta = 6$, for which results are shown in Figure 6. Thus, any value lower than 6 will be replaced by 0 in the voting matrix, and subsequently in the lead-lag matrix. We note that, for both the homogeneous and heterogeneous settings, the voting matrix consists of non-zeros across the diagonals for each block, and the other blocks consist only of 0s. We conclude that our proposed method with a voting threshold has good performance on synthetic data, and fully recovers the lead-lag relationship in both the homogeneous and heterogeneous settings.

K-means++ clustering

Homogeneous Setting

TS 1	0	15	13	11	11	11
TS 2	15	0	13	11	11	12
TS 3	13	13	0	11	11	12
TS 4	11	11	11	0	11	11
TS 5	11	11	11	11	0	13
TS 6	11	12	12	11	13	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

$k = 1$

Heterogeneous Setting

TS 1	0	9	7	0	0	0
TS 2	9	0	9	0	0	0
TS 3	7	9	0	0	0	0
TS 4	0	0	0	0	11	9
TS 5	0	0	0	11	0	9
TS 6	0	0	0	9	9	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

TS 1	0	0	0	0	0	0
TS 2	8	0	0	0	0	0
TS 3	0	0	0	9	0	0
TS 4	0	0	9	0	0	0
TS 5	0	0	0	0	0	10
TS 6	0	0	0	0	10	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0
TS 1	TS 2	TS 3	TS 4	TS 5	TS 6	

$k = 2$

$k = 3$

Figure 6: Top panel: Voting matrix with voting threshold ($\theta = 6$). Middle panel: Error matrix based on mode estimation with the voting threshold ($\theta = 6$). Bottom panel: Error matrix based on median estimation with the voting threshold ($\theta = 6$).

5.3 Simulation

In the next set of synthetic experiments, we increase n to 60. The B and L matrices follow an analogous pattern to the $n = 6$ case. The Adjusted Rand Index (ARI) calculates a similarity metric between two clusterings by examining every sample pair and counting the number of pairs assigned to the same or different clusters in both the predicted and actual clusterings. The ARI is calculated by adjusting the Rand Index for chance agreement, and it ranges from -1 to 1, where a score of 1 indicates perfect agreement between the two clusterings, a score of 0 indicates random agreement, and a score less than 0 indicates disagreement between the two clusterings. As observed in Figure 7, it is evident that for both the homogeneous and heterogeneous settings, when the σ ranges from 0 to 1.5, the KM clustering achieves a stable ARI around 0.8, whereas the SP clustering fluctuates slightly in the earlier σ values. However, when the σ value exceeds 1.5, the SP clustering gains a slightly higher performance than the KM clustering, and the ARI for both methods drops significantly.

Figure 8 shows the average and confidence interval for the MSE when estimating the true lag with different σ levels based on 100 simulations without voting threshold ($\theta = 1$, top panel) and with voting threshold ($\theta = 6$, bottom panel).

In the homogeneous setting ($k = 1$), both plots follow the same trend. Specifically, for σ ranging from 0 to 1.5, the MSE of the K-means++ mode (KM_Mod) and spectral mode (SP_Mod) is near 0. In comparison, the MSE of K-means++ median (KM_Med) and spectral median (SP_Med) is

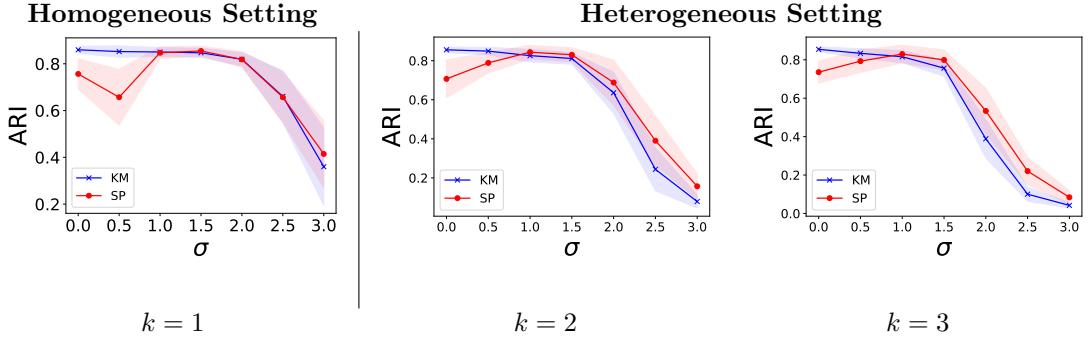


Figure 7: Average and confidence interval for the ARI with different levels of σ based on 60 time series and 100 simulations for every iteration.

higher. When σ ranges from 1.5 to 3.0, all four methods' MSE increases significantly due to the higher noise level.

In the heterogeneous setting ($k \in \{2, 3\}$), we note that for methods not using the voting mechanism, the MSE is significantly higher, having a value of around 12. However, once the voting mechanism is added, the MSE becomes significantly lower. When the σ values range between 0 to 1.5, the MSE of KM clustering goes down to near 0, whereas SP clustering have a slightly higher MSE. After the value of σ reaches 1.5 and beyond, the MSE increases dramatically for all four methods.

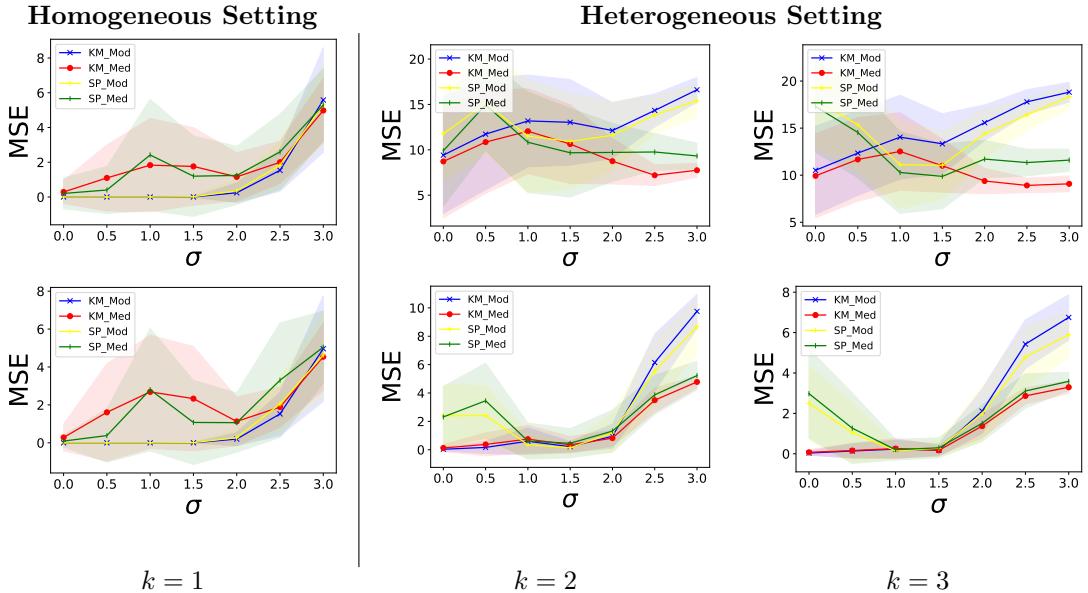


Figure 8: Top panel: Average and confidence interval for the MSE with different levels of σ based on 100 simulations for every iteration without voting threshold ($\theta = 1$). Bottom panel: Average and confidence interval for the MSE with different σ levels based on 100 simulations for every iteration with voting threshold ($\theta = 6$).

We next study the sensitivity of our proposed method to the value of the voting threshold θ . In Figure 9, we first observe that in the homogeneous setting ($k = 1$), the four methods maintain a constant MSE as the voting threshold increases. This aligns with our expectations. Moreover, we also notice that mode estimation performs better than median estimation. This is evident throughout the plots in Figure 7 and 8. Thus, the MSE for median estimation lies between 2 and 3, whereas the MSE for the mode estimation drops to near 0. For the heterogeneous setting

($k \in \{2, 3\}$), we note that as the voting threshold increases, the MSE decreases to near 0 for all methods and remains low thereafter. This altogether suggests that the proposed method is robust to the choice of the voting threshold θ .

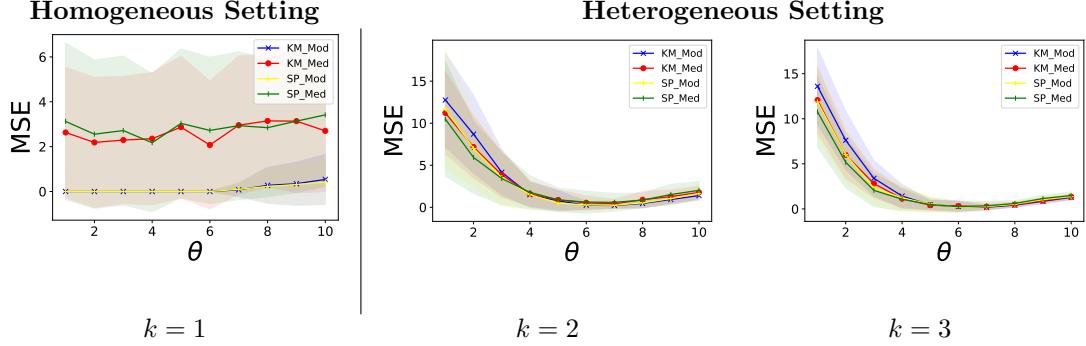
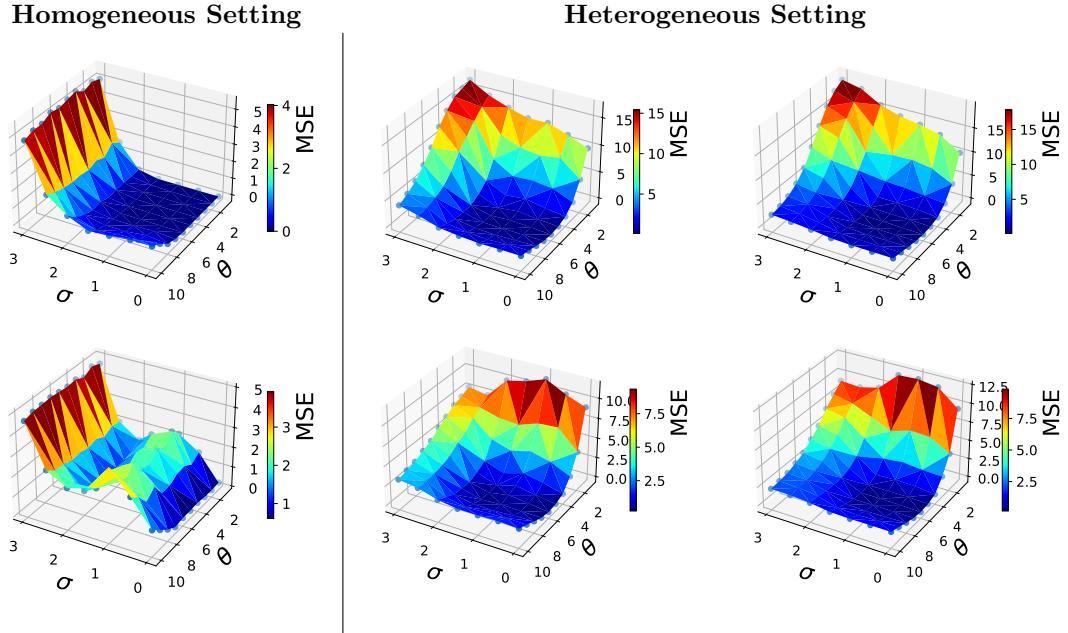


Figure 9: Average and confidence interval for the MSE with different levels of θ based on 100 simulations when $\sigma = 1$.

Figure 10 shows the average MSE as a function of σ and θ . These are based on 100 simulations for every setting. The plots for each panel represent the methods, which are KM.Mod, KM.Med, SP.Mod, and SP.Med, from top to bottom, respectively. For the homogeneous setting ($k = 1$), we can observe that for all four methods, the MSE is heavily influenced by the increase in σ , which is consistent with Figure 8. We also note that when the voting threshold increases, there are no significant changes in the MSE which is consistent with Figure 9. For the heterogeneous case, we note that, in general, the optimal MSE is achieved when the voting threshold is approximately 6 across a wide range of values of σ . This is consistent with Figures 8 and 9, and suggests that the choice of the voting threshold is robust to model noise.



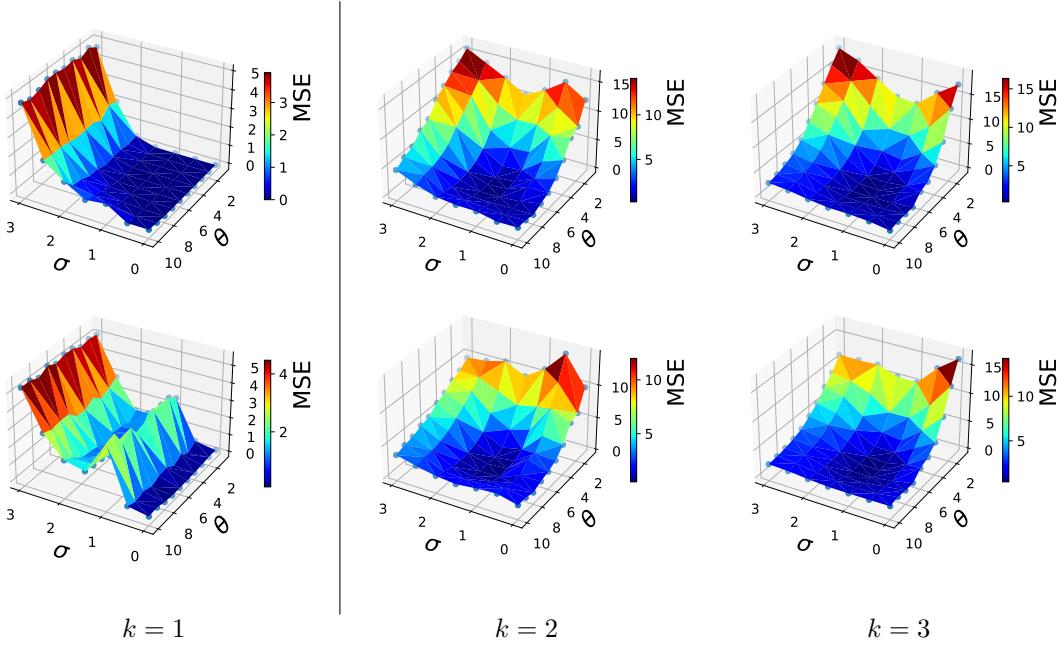


Figure 10: Average MSE as a function of σ and θ based on 100 simulations. From top panel to bottom panel, KM_Mod, KM_Med, SP_Mod, and SP_Med.

6 Financial data experiments

6.1 Data description

In this section, we apply our methods to a large-scale experiment using financial data. As mentioned earlier, this is a context where lead-lag relationships naturally occur. For the financial data experiments, we consider three data sets, which vary in terms of the number and type of assets and the number of days.

- For the first data set, we look at US equities from [Wharton's CRSP](#) data set. There are a total of 679 equities within this data set, over a range of 5211 trading days from 2000/01/03 to 2020/12/31.
- The second data set is derived from the same [Wharton's CRSP](#) data set., but looks at Exchange Traded Funds (ETFs). It consists of 14 ETFs with 3324 trading days from 2006/04/12 to 2019/07/01.
- The third data set is the [Pinnacle Data Corp CLC](#), which includes 52 futures contracts with 5166 trading days from 2000/01/05 to 2020/10/16. All futures contracts are adjusted for rolling effects. This data set spans multiple asset classes, which include commodities, fixed income, and currency futures.

All data sets are considered at daily frequency. We summarize all of the above in Table 4, with additional details about Pinnacle Data Corp CLC available in Appendix A.3 Tables [17, 18, 19, 20, 21, 22, 23, 24].

Table 4: Summary of the three financial data sets considered in the numerical experiments.

Data source	Type	Freq	# of assets	Start date	End date	# of days
Wharton's CRSP	Equity	Daily	679	2000/01/03	2020/12/31	5211
Wharton's CRSP	ETF	Daily	14	2006/04/12	2019/07/01	3324
Pinnacle Data Corp	Futures	Daily	52	2000/01/05	2020/10/16	5166

6.2 Data pre-processing

With regard to the US equity and ETF data sets, we download the close-to-close adjusted daily returns from Wharton's CRSP. Due to the large number of NaNs in the equity data set, we drop the days for which more than 10% of the equities have zero returns as well as the equities for which more than 50% of days have zero returns. Instead of working with raw returns, we consider the market excess returns, a standard measure of how well each equity performed relative to the broader market. For both of these data sets, the return of the S&P Composite Index is selected to compute the market excess returns by subtracting it from the return of each asset (i.e., for simplicity, we assume each asset has $\beta = 1$ exposure to the market). Also, we winsorize the extreme value of excess returns for which any value is larger than 0.15 or smaller than -0.15.

For the futures data set, we download the close-to-close price series from the Pinnacle Data Corp CLC data set, and discard the days for which more than 10% of the futures have zero prices in the respective dates, and drop the futures for which more than 160 days have zero prices. Afterwards, we first use forward-fill, then backward-fill to fill out the zero prices. Lastly, we compute the log-return from the close-to-close price. The remainder of the data pre-processing is the same as above.

6.3 Benchmark

In order to evaluate our proposed methodology, we also introduce a benchmark to detect lead-lag relationships without the use of clustering. It is very common to compute a sample cross-correlation function (CCF) between two time series. A CCF between time series X_i and X_j evaluated at lag m is given by

$$\text{CCF}^{ij}(m) = \text{CORR}(\{X_i^{t-m}\}, \{X_j^t\}), \quad (11)$$

where $\text{CORR}()$ denotes a choice of the CCF. The corresponding lead-lag matrix $\Gamma_{n \times n}$ is estimated by computing the signed normalized area under the curve of CCF, given by

$$\Gamma_{ij} = \frac{\text{MAX}(I(i,j), I(j,i)) \cdot \text{SIGN}(I(i,j) - I(j,i))}{I(i,j) + I(j,i)}, \quad (12)$$

where $I(i,j) = \sum_{m=1}^M |\text{CCF}^{ij}(m)|$ for a user-specified maximum lag M .

We summarize the benchmark procedures in Algorithm 4.

Algorithm 4 : CCF Algorithm

Input: Time series matrix $X_{n \times T}$.

Output: Lead-lag matrix $\Gamma_{n \times n}$.

- 1: Calculate CCF for every pair of time series $\{X_i, X_j\}$.
 - 2: Calculate the lead-lag matrix $\Gamma_{n \times n}$ by computing the signed normalized area under the curve of CCF.
-

6.4 Trading strategies

Given n time series each of length T , we first extract data by a sliding window with length $l = 21$. Next, we begin by extracting the STS with length $q = 10$ via a sliding window shifted by $s = 1$ from the data, and cluster the STS to calculate the estimated lead-lag matrix $\Gamma_{n \times n}$ after applying the voting threshold $\theta = 6$, which is validated in the synthetic data experiment. We then utilize the lead-lag matrix to rank the time series from the most leading to the most lagging using the *RowSum* ranking [Huber (1963), Gleich and L.-h. Lim (2011)], in order to then group the time series into leaders and laggars, with the former used to predict the latter.

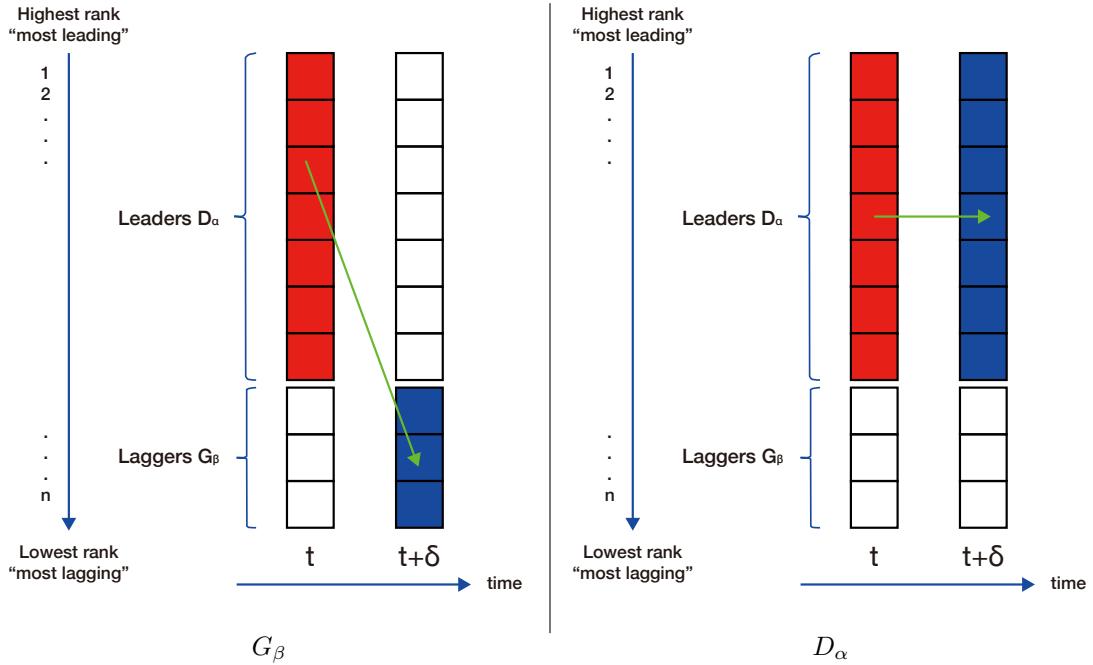


Figure 11: G_β strategy: Use D_α predict G_β (left). D_α strategy: Use D_α predict D_α (right).

Momentum is a widely studied phenomenon in the finance literature [Jegadeesh, Luo, et al. (2022), Jegadeesh and Titman (2001), Tan, Roberts, and Zohren (2023), Poh, Roberts, and Zohren (2022), Wood, Giegerich, et al. (2021), Wood, Roberts, and Zohren (2021), B. Lim, Zohren, and Roberts (2019)], which refers to the tendency of assets that have performed well in the recent past to continue to perform well in the near future, and vice versa. We denote the top $\alpha = \{0.7, 0.75, 0.8, 0.85\}$ fraction of the time series as Leaders D_α , and the bottom $\beta = 1 - \alpha$ as Laggers G_β . We use the exponentially weighted moving average (EWMA) signal on the past $p = \{1, 3, 5, 7\}$ days of the average winsorized time series excess returns of the D_α to predict the average future $\delta = \{1, 3, 5, 7\}$ days of the excess return of the G_β and D_α . We assume the G_β can catch up with the D_α , and the D_α provides the necessary momentum to maintain the trend

in δ days, respectively. This is depicted in Figure 11. Afterwards, we shift the sliding window by $s = 1$, and re-apply the method to calculate the lead-lag matrix $\Gamma_{n \times n}$ and rank the time series until the end of the time series. For clarity, Figure 12 reflects our trading pipeline at time t , and we summarize the trading strategy in Algorithm 5.

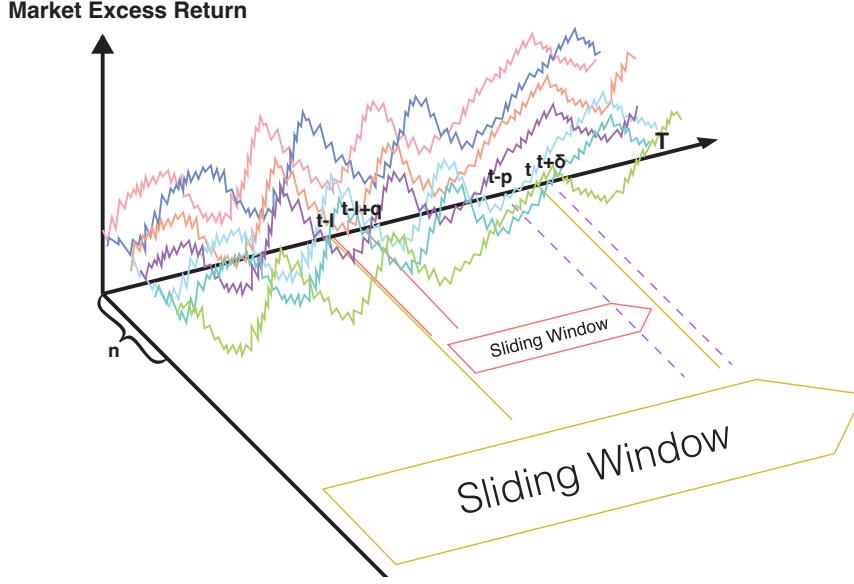


Figure 12: Trading pipeline.

Algorithm 5 : Trading strategy

Input: Time series matrix $X_{n \times T}$.

- 1: Build $X_{n \times T}$ by applying a sliding window of length l to obtain $X_{n \times l}$.
 - 2: Use the **Lead-lag Relationship Detection Algorithm** on $X_{n \times l}$ to obtain the lead-lag matrix $\Gamma_{n \times n}$.
 - 3: Based on $\Gamma_{n \times n}$, rank the time series from the most leading to the most lagging using the *RowSum* ranking.
 - 4: Pick the top α fraction of the time series as Leaders D_α , and the bottom $\beta = 1 - \alpha$ as Laggards G_β .
 - 5: Apply the EWMA on the past p days of the average winsorized time series excess returns of the D_α to predict the average future δ days of the excess return of the G_β and D_α .
 - 6: Shift the sliding window by s , and re-apply Steps 1 - 5 until the end of the time series.
-

6.5 Performance evaluation

When assessing the effectiveness of various trading strategies, we rely on the following metrics to evaluate their performance:

We compute the Profit and Loss (PnL) of G_β on a given day $t + \delta$ as

$$\text{PnL}_{G_\beta}^{t+\delta} = \text{sign}(\text{EWMA}(\text{ret}_{D_\alpha}^{t-p} : \text{ret}_{D_\alpha}^t)) \cdot \overline{\text{ret}_{G_\beta}^{t+\delta}}, t = l, \dots, T - \delta, \quad (13)$$

since the strategy earns money whenever the sign of the forecast agrees with the sign of the future return. Correspondingly, the PnL of D_α on a given day $t + \delta$ is given by:

$$\text{PnL}_{D_\alpha}^{t+\delta} = \text{sign}(\text{EWMA}(\text{ret}_{D_\alpha}^{t-p} : \text{ret}_{D_\alpha}^t)) \cdot \overline{\text{ret}_{D_\alpha}^{t+\delta}}, t = l, \dots, T - \delta, \quad (14)$$

where $ret_{D_\alpha}^{t-p}$ and $ret_{D_\alpha}^t$ are the excess return of D_α at $t-p$ and t , respectively, while $\text{EWMA}(ret_{D_\alpha}^{t-p} : ret_{D_\alpha}^t)$ denotes the exponentially weighted moving average from the excess return of D_α from $t-p$ to t . Furthermore, $\overline{ret}_{G_\beta}^{t+\delta}$ depicts the mean of the excess return of G_β at $t+\delta$, and $\overline{ret}_{D_\alpha}^{t+\delta}$ is the mean of the excess return of D_α at $t+\delta$. We rescale the PnL by their volatility to target equal risk assignment, and set our annualized volatility target σ_{tgt} to be 0.15.

$$\text{PnL}_{\text{rescaled}} = \frac{\sigma_{\text{target}}}{\text{STD}(\text{PnL}) \cdot \sqrt{252}} \cdot \text{PnL}. \quad (15)$$

The cumulative PnL sums the daily PnL across all trading days

$$\text{Cumulative PnL} = \sum \text{PnL}_{\text{rescaled}}. \quad (16)$$

The annualized expected excess return ($E[\text{Returns}]$) is to measure the excess return earned by an investment over a benchmark index during a one-year period, and can be computed by

$$E[\text{Returns}] = \text{AVG}(\text{PnL}_{\text{rescaled}}) \cdot 252. \quad (17)$$

The annualized volatility is a measurement of the amount of risk associated with an investment over a one-year period

$$\text{Volatility} = \text{STD}(\text{PnL}_{\text{rescaled}}) \cdot \sqrt{252}. \quad (18)$$

Furthermore, we calculate the downside deviation and maximum drawdown to measure downside risk, and then the Sortino ratio is often used by investors who are more concerned with downside risk than with overall risk or volatility, which is derived by

$$\text{Sortino ratio} = \frac{E[\text{Returns}]}{\text{downside deviation}}. \quad (19)$$

The Calmar ratio is often used by investors who are more concerned with long-term risk and downside protection. A higher Calmar ratio indicates that the strategy has generated higher returns relative to its maximum drawdown, while a lower Calmar ratio suggests that the strategy has underperformed given the level of risk it has taken. It is calculated by

$$\text{Calmar ratio} = \frac{E[\text{Returns}]}{\text{maximum drawdown}}. \quad (20)$$

The hit rate measures the percentage of successful trades made by the strategy. It is also known as the win rate or success rate, and is defined as

$$\text{Hit rate} = \frac{|\text{PnL}_{\text{rescaled}}^+|}{|\text{PnL}_{\text{rescaled}}|}, \quad (21)$$

where $|\text{PnL}_{\text{rescaled}}^+|$ is the number of profitable trades, and $|\text{PnL}_{\text{rescaled}}|$ is the total number of trades.

The average profit / average loss (avg. profit / avg. loss) ratio measures the average size of profits relative to the average size of losses generated by the strategy.

$$\text{Avg. profit / avg. loss} = \frac{\text{AVG}(\text{PnL}_{\text{rescaled}}^+)}{\text{AVG}(\text{PnL}_{\text{rescaled}}^-)}, \quad (22)$$

where $\text{AVG}(\text{PnL}_{\text{rescaled}}^+)$ is the average profit per trade, and $\text{AVG}(\text{PnL}_{\text{rescaled}}^-)$ is the average loss per trade.

The PnL per trade illustrates the amount earned by the strategy, in basis points for each basket of G_β or D_α traded in the markets (excluding transaction costs), and is given by

$$\text{PnL per trade} = \text{AVG}(\text{PnL}_{\text{rescaled}}) \cdot 10^4, \quad (23)$$

where we assume that the strategy trades the same amount of notional every day (i.e., a constant unit bet size is used every trading day).

We also compute the annualized Sharpe ratio to quantify the profit gained per unit of risk taken

$$\text{Sharpe ratio} = \frac{\text{AVG}(\text{PnL}_{\text{rescaled}})}{\text{STD}(\text{PnL}_{\text{rescaled}})} \cdot \sqrt{252}. \quad (24)$$

It is important to assess the statistical significance of Sharpe ratio when back-testing a sample of hypothetical strategies [Bailey and De Prado (2014), Ledoit and Wolf (2008), Michael, Cucuringu, and Howison (2022)]. We use a test with the null hypothesis $H_0 : \text{Sharpe ratio} = 0$, and implement the method proposed by [Bailey and De Prado (2014)] to compute the test statistic

$$\frac{(\text{Sharpe ratio}) \cdot \sqrt{T - 1}}{\sqrt{1 - \gamma_1 \cdot (\text{Sharpe ratio}) + (\gamma_2 - 1) \cdot (\text{Sharpe ratio})^2 / 4}}, \quad (25)$$

where the Sharpe ratio is what we are testing, T is the length of the sample, and γ_1 and γ_2 are the skewness and kurtosis of the returns distribution for the selected strategy, respectively. This test statistic is assumed to be standard normal under the null hypothesis.

To evaluate the predictive performance of our method, we construct a simple trading strategy. If the strategy is profitable with a statistically significant Sharpe ratio, this indicates that we are able to leverage the discovered lead-lag relationships for the prediction task.

6.6 Results

For the equity data set, full results with different tuning settings are reported in [supplemental material](#) [Zhang et al. (2023)] Table S1. We find that using the EWMA on the past seven days of the average winsorized time series excess returns of the D_α to predict the average future seven days of the excess return of the G_β and D_α with $\alpha = 0.75$ performs consistently well across all methods. In Tables [5, 6], we report the performance of the CCF and four methods based on the various metrics (rescaled to target volatility). Across the four methods we proposed, we note that the Sharpe ratio values are relatively high. In addition to the Sharpe ratio, the P-value associated with the Sharpe ratio also supports our hypothesis that the Sharpe ratio is statistically significant. It is evident that most P-values are 0 while others are significantly lower than 0.05, which provides us with more confidence when analyzing the results of the experiment. In particular, according to the G_β strategy, it has been observed that KM_Med outperforms other methods in terms of

most metrics. Conversely, based on the D_α strategy, KM_Mod performs better than other methods across most metrics. It is noteworthy that KM clustering produces similar results for both the G_β and D_α strategies, and they are also much faster than SP clustering in terms of running time. In Figure 13, we depict the cumulative PnL (rescaled to target volatility) across trading days for the scenarios in Tables [5, 6]. The left and right plots present G_β , and D_α strategies, respectively. We also present the benchmark performance for comparison.

Table 5: Equity data set: performance metrics for G_β strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
		CCF	KM_Mod	KM_Med	SP_Mod
E[Returns]	0.089	0.126	0.118	0.127*	0.104
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.105	0.103	0.101*	0.103	0.105
Maximum drawdown	-0.313	-0.26	-0.215	-0.214*	-0.287
Sortino ratio	0.85	1.222	1.166	1.237*	0.988
Calmar ratio	0.285	0.484	0.548	0.594*	0.362
Hit rate	0.499	0.521*	0.51	0.516	0.519
Avg. profit / avg. loss	1.117*	1.068	1.107	1.091	1.051
PnL per trade	3.542	4.996	4.672	5.041*	4.119
Sharpe ratio	0.595	0.839	0.785	0.847*	0.692
P-value	0.009	0*	0*	0*	0.002

Table 6: Equity data set: performance metrics for D_α strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
		CCF	KM_Mod	KM_Med	SP_Mod
E[Returns]	0.101	0.122	0.095	0.129	0.134*
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.105*	0.107	0.108	0.106	0.105*
Maximum drawdown	-0.288	-0.251	-0.21*	-0.283	-0.227
Sortino ratio	0.964	1.148	0.876	1.222	1.269*
Calmar ratio	0.352	0.488	0.452	0.456	0.59*
Hit rate	0.518	0.521	0.513	0.522	0.525*
Avg. profit / avg. loss	1.05	1.069	1.066	1.072*	1.065
PnL per trade	4.018	4.856	3.765	5.121	5.312*
Sharpe ratio	0.675	0.816	0.632	0.86	0.892*
P-value	0.003	0*	0.005	0*	0*

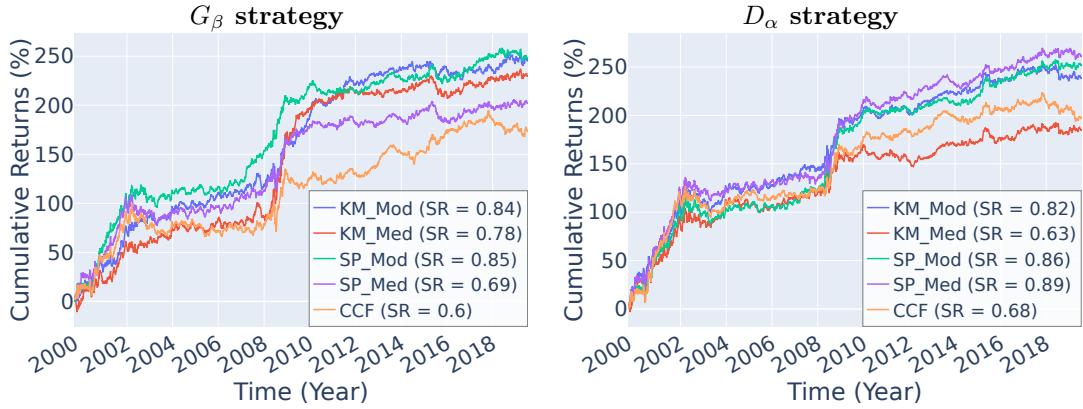


Figure 13: Equity data set: cumulative PnL for G_β strategy (left) and D_α strategy (right) - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

Tables [7, 8] and Figure 14 present results for the ETF data set using the same settings as for the equities data. Compared to equities data, we find no evidence of being able to consistently detect lead-lag relationships leading to a profitable outcome. We note, though, the performance is strong in some cases, namely the D_α strategy with the exception of KM_Med. Full results across all tuning settings are reported in [supplemental material](#) [Zhang et al. (2023)] Table S2. Results for futures data, again with the same settings as the equities data, are reported in Tables [9, 10] and Figure 15. For this data set, we do not see an ability to consistently detect profitable lead-lag relationships for any of the strategies. Full results across all tuning settings are reported in [supplemental material](#) [Zhang et al. (2023)] Table S3.

Table 7: ETF data set: performance metrics for G_β strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
		CCF	KM_Mod	KM_Med	SP_Mod
E[Returns]	-0.019	0.02	0.022	-0.005	0.013
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.116	0.115	0.116	0.113	0.108
Maximum drawdown	-0.668	-0.525	-0.465	-0.579	-0.369
Sortino ratio	-0.165	0.176	0.188	-0.043	0.122
Calmar ratio	-0.029	0.038	0.047	-0.008	0.036
Hit rate	0.492	0.512	0.499	0.503	0.5
Avg. profit / avg. loss	1.006	0.978	1.034	0.983	1.016
PnL per trade	-0.76	0.8	0.87	-0.191	0.527
Sharpe ratio	-0.128	0.134	0.146	-0.032	0.088
P-value	0.644	0.628	0.598	0.908	0.749

Table 8: ETF data set: performance metrics for D_α strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
	CCF	KM_Mod	KM_Med	SP_Mod	SP_Med
E[Returns]	0.056	0.065	0.022	0.073	0.081
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.097	0.097	0.105	0.1	0.105
Maximum drawdown	-0.362	-0.267	-0.385	-0.272	-0.382
Sortino ratio	0.581	0.676	0.213	0.728	0.771
Calmar ratio	0.156	0.245	0.058	0.269	0.211
Hit rate	0.504	0.502	0.5	0.51	0.504
Avg. profit / avg. loss	1.056	1.077	1.027	1.05	1.087
PnL per trade	2.234	2.591	0.889	2.901	3.2
Sharpe ratio	0.375	0.435	0.149	0.487	0.538
P-value	0.169	0.105	0.588	0.074	0.051

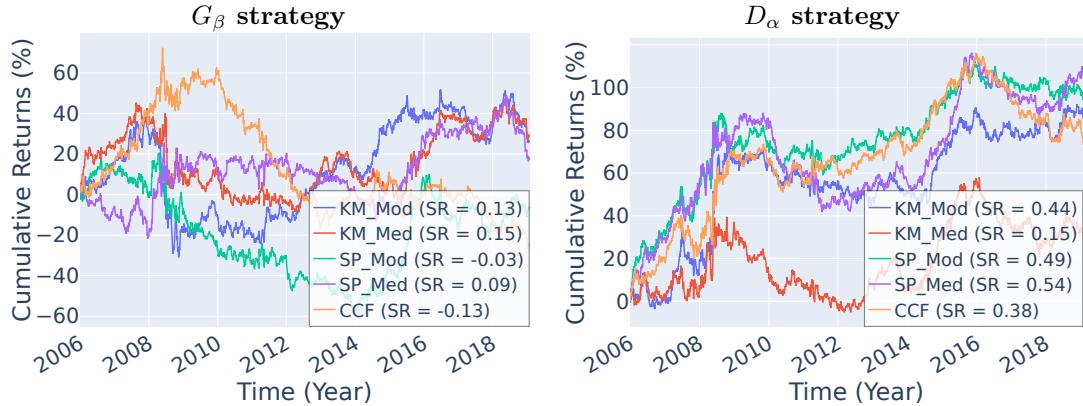


Figure 14: ETF data set: cumulative PnL for G_β strategy (left) and D_α strategy (right) - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

Table 9: Futures data set: performance metrics for G_β strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
	CCF	KM_Mod	KM_Med	SP_Mod	SP_Med
E[Returns]	0.013	-0.001	0.005	-0.01	0.022
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.104	0.102	0.103	0.102	0.105
Maximum drawdown	-0.535	-0.587	-0.534	-0.508	-0.457
Sortino ratio	0.121	-0.007	0.045	-0.097	0.209
Calmar ratio	0.024	-0.001	0.009	-0.019	0.048
Hit rate	0.502	0.498	0.495	0.494	0.508
Avg. profit / avg. loss	1.007	1.009	1.028	1.013	0.994
PnL per trade	0.499	-0.028	0.185	-0.392	0.869
Sharpe ratio	0.084	-0.005	0.031	-0.066	0.146
P-value	0.705	0.983	0.888	0.766	0.51

Table 10: Futures data set: performance metrics for D_α strategy - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

G_β strategy	Benchmark	Proposed			
		CCF	KM_Mod	KM_Med	SP_Mod
E[Returns]	0.061	0.036	0.036	0.027	0.031
Volatility	0.15	0.15	0.15	0.15	0.15
Downside deviation	0.107	0.108	0.108	0.108	0.107
Maximum drawdown	-0.393	-0.411	-0.399	-0.474	-0.462
Sortino ratio	0.574	0.336	0.335	0.253	0.29
Calmar ratio	0.156	0.088	0.091	0.057	0.067
Hit rate	0.511	0.502	0.499	0.498	0.501
Avg. profit / avg. loss	1.032	1.038	1.05	1.043	1.037
PnL per trade	2.439	1.441	1.443	1.079	1.23
Sharpe ratio	0.41	0.242	0.242	0.181	0.207
P-value	0.064	0.274	0.273	0.412	0.35

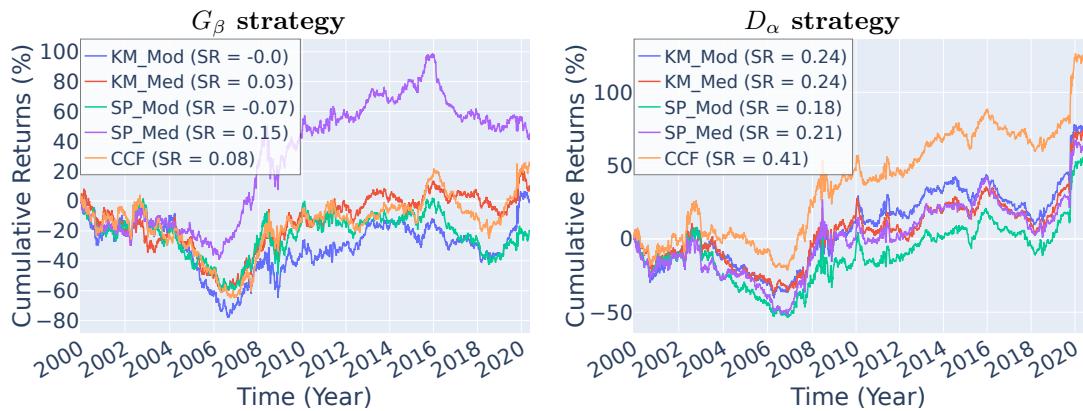


Figure 15: Futures data set: cumulative PnL for G_β strategy (left) and D_α strategy (right) - rescaled to target volatility. The experiment has been set with the values $p = 7$, $\delta = 7$, and $\alpha = 0.25$.

7 Robustness analysis

In this section, we test the robustness of the benchmark and our proposed methods by conducting experiments with different levels of α . For the equity data set, we consider α values ranging from 0.7 to 0.85 with an increment of 0.05. In Table 11, we note that the performance of all methods does not change significantly while maintaining a high Sharpe ratio. We notice that the P-values are almost all lower than 0.05, suggesting that all results are significant in our experiments. We also note that for the SP_Med, the G_β strategy has slightly lower performance.

In Table 12, we report on the performance of the four methods tested on the ETF data set. Here, we only consider the α values of 0.75 and 0.85 due to the smaller cross-section for this data set. The four methods achieve a fairly good performance for high alpha for the D_α strategy, except for KM_Med.

Finally, in Table 13, we provide the performance of the four methods on the futures data set, with α ranging from 0.7 to 0.85 in increments of 0.05. We note that for this data set, the Sharpe ratio for the G_β strategy tends to be more sensitive to change in α and it is hard to achieve profitability.

Table 11: Equity data set: robustness analysis for α - rescaled to target volatility. The experiment has been set with the values $p = 7$, and $\delta = 7$.

CCF		G_β strategy				D_α strategy				
α		0.7	0.75	0.8	0.85		0.7	0.75	0.8	0.85
E[Returns]		0.082	0.089	0.089	0.079		0.088	0.101	0.108	0.106
Volatility		0.15	0.15	0.15	0.15		0.15	0.15	0.15	0.15
Downside deviation		0.106	0.105	0.105	0.105		0.106	0.105	0.105	0.105
Maximum drawdown		-0.313	-0.313	-0.29	-0.259		-0.267	-0.288	-0.254	-0.254
Sortino ratio		0.776	0.85	0.849	0.757		0.824	0.964	1.027	1.009
Calmar ratio		0.263	0.285	0.306	0.306		0.328	0.352	0.425	0.418
Hit rate		0.504	0.499	0.505	0.505		0.517	0.518	0.521	0.52
Avg. profit / avg. loss		1.086	1.117	1.091	1.08		1.037	1.05	1.047	1.048
PnL per trade		3.262	3.542	3.526	3.143		3.473	4.018	4.288	4.209
Sharpe ratio		0.548	0.595	0.592	0.528		0.583	0.675	0.72	0.707
P-value		0.015	0.009	0.009	0.02		0.01	0.003	0.001	0.002
KM_Mod		G_β strategy				D_α strategy				
α		0.7	0.75	0.8	0.85		0.7	0.75	0.8	0.85
E[Returns]		0.117	0.126	0.117	0.143		0.116	0.122	0.117	0.123
Volatility		0.15	0.15	0.15	0.15		0.15	0.15	0.15	0.15
Downside deviation		0.104	0.103	0.103	0.101		0.105	0.107	0.106	0.106
Maximum drawdown		-0.245	-0.26	-0.236	-0.213		-0.207	-0.251	-0.289	-0.288
Sortino ratio		1.127	1.222	1.137	1.42		1.099	1.148	1.102	1.157
Calmar ratio		0.476	0.484	0.494	0.672		0.559	0.488	0.405	0.427
Hit rate		0.521	0.521	0.516	0.522		0.51	0.521	0.52	0.522
Avg. profit / avg. loss		1.059	1.068	1.078	1.085		1.107	1.069	1.067	1.065
PnL per trade		4.628	4.996	4.63	5.676		4.596	4.856	4.648	4.881
Sharpe ratio		0.778	0.839	0.778	0.954		0.772	0.816	0.781	0.82
P-value		0.001	0	0.001	0		0.001	0	0.001	0

KM_Med		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.115	0.118	0.121	0.138	0.094	0.095	0.086	0.1
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.103	0.101	0.099	0.097	0.109	0.108	0.109	0.108
Maximum drawdown		-0.235	-0.215	-0.22	-0.177	-0.198	-0.21	-0.286	-0.259
Sortino ratio		1.115	1.166	1.225	1.428	0.866	0.876	0.788	0.925
Calmar ratio		0.487	0.548	0.549	0.781	0.476	0.452	0.299	0.386
Hit rate		0.517	0.51	0.509	0.517	0.515	0.513	0.512	0.514
Avg. profit / avg. loss		1.074	1.107	1.113	1.101	1.058	1.066	1.058	1.068
PnL per trade		4.545	4.672	4.796	5.486	3.741	3.765	3.397	3.97
Sharpe ratio		0.764	0.785	0.806	0.922	0.629	0.632	0.571	0.667
P-value		0.001	0	0	0	0.005	0.005	0.012	0.003
SP_Mod		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.137	0.127	0.129	0.129	0.139	0.129	0.129	0.123
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.102	0.103	0.103	0.102	0.105	0.106	0.105	0.105
Maximum drawdown		-0.199	-0.214	-0.206	-0.162	-0.276	-0.283	-0.28	-0.256
Sortino ratio		1.346	1.237	1.258	1.267	1.333	1.222	1.226	1.18
Calmar ratio		0.691	0.594	0.628	0.798	0.505	0.456	0.46	0.482
Hit rate		0.523	0.516	0.522	0.522	0.524	0.522	0.52	0.521
Avg. profit / avg. loss		1.076	1.091	1.069	1.07	1.074	1.072	1.079	1.068
PnL per trade		5.456	5.041	5.135	5.131	5.532	5.121	5.115	4.895
Sharpe ratio		0.917	0.847	0.863	0.862	0.929	0.86	0.859	0.822
P-value		0	0	0	0	0	0	0	0
SP_Med		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.115	0.104	0.098	0.084	0.142	0.134	0.134	0.13
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.105	0.105	0.106	0.105	0.104	0.105	0.105	0.104
Maximum drawdown		-0.279	-0.287	-0.271	-0.256	-0.221	-0.227	-0.217	-0.215
Sortino ratio		1.099	0.988	0.927	0.794	1.367	1.269	1.273	1.253
Calmar ratio		0.414	0.362	0.363	0.327	0.644	0.59	0.617	0.604
Hit rate		0.52	0.519	0.518	0.513	0.527	0.525	0.526	0.525
Avg. profit / avg. loss		1.059	1.051	1.046	1.051	1.067	1.065	1.06	1.059
PnL per trade		4.579	4.119	3.904	3.321	5.648	5.312	5.316	5.15
Sharpe ratio		0.769	0.692	0.656	0.558	0.949	0.892	0.893	0.865
P-value		0.001	0.002	0.004	0.014	0	0	0	0

Table 12: ETF data set: robustness analysis for α - rescaled to target volatility. The experiment has been set with the values $p = 7$, and $\delta = 7$.

CCF	G_β strategy		D_α strategy		
	α	0.75	0.85	0.75	0.85
E>Returns]		-0.019	-0.038	0.056	0.041
Volatility		0.15	0.15	0.15	0.15
Downside deviation		0.116	0.116	0.097	0.097
Maximum drawdown		-0.668	-0.672	-0.362	-0.382
Sortino ratio		-0.165	-0.324	0.581	0.422
Calmar ratio		-0.029	-0.056	0.156	0.107
Hit rate		0.492	0.501	0.504	0.494
Avg. profit / avg. loss		1.006	0.949	1.056	1.077
PnL per trade		-0.76	-1.49	2.234	1.617
Sharpe ratio		-0.128	-0.25	0.375	0.272
P-value		0.644	0.365	0.169	0.321
KM_Mod	G_β strategy		D_α strategy		
	α	0.75	0.85	0.75	0.85
E>Returns]		0.02	-0.06	0.065	0.083
Volatility		0.15	0.15	0.15	0.15
Downside deviation		0.115	0.122	0.097	0.096
Maximum drawdown		-0.525	-0.707	-0.267	-0.292
Sortino ratio		0.176	-0.487	0.676	0.86
Calmar ratio		0.038	-0.084	0.245	0.283
Hit rate		0.512	0.505	0.502	0.505
Avg. profit / avg. loss		0.978	0.911	1.077	1.086
PnL per trade		0.8	-2.363	2.591	3.28
Sharpe ratio		0.134	-0.397	0.435	0.551
P-value		0.628	0.147	0.105	0.039
KM_Med	G_β strategy		D_α strategy		
	α	0.75	0.85	0.75	0.85
E>Returns]		0.022	0.004	0.022	0.052
Volatility		0.15	0.15	0.15	0.15
Downside deviation		0.116	0.113	0.105	0.104
Maximum drawdown		-0.465	-0.436	-0.385	-0.336
Sortino ratio		0.188	0.031	0.213	0.504
Calmar ratio		0.047	0.008	0.058	0.155
Hit rate		0.499	0.505	0.5	0.502
Avg. profit / avg. loss		1.034	0.984	1.027	1.06
PnL per trade		0.87	0.14	0.889	2.071
Sharpe ratio		0.146	0.024	0.149	0.348
P-value		0.598	0.932	0.588	0.205
SP_Mod	G_β strategy		D_α strategy		
	α	0.75	0.85	0.75	0.85
E>Returns]		-0.005	-0.076	0.073	0.12
Volatility		0.15	0.15	0.15	0.15
Downside deviation		0.113	0.118	0.1	0.095
Maximum drawdown		-0.579	-0.726	-0.272	-0.258
Sortino ratio		-0.043	-0.648	0.728	1.259
Calmar ratio		-0.008	-0.105	0.269	0.464
Hit rate		0.503	0.497	0.51	0.511
Avg. profit / avg. loss		0.983	0.918	1.05	1.107
PnL per trade		-0.191	-3.035	2.901	4.747
Sharpe ratio		-0.032	-0.51	0.487	0.797
P-value		0.908	0.064	0.074	0.003

SP_Med		G_β strategy		D_α strategy	
α		0.75	0.85	0.75	0.85
E[Returns]		0.013	-0.025	0.081	0.08
Volatility		0.15	0.15	0.15	0.15
Downside deviation		0.108	0.11	0.105	0.103
Maximum drawdown		-0.369	-0.452	-0.382	-0.348
Sortino ratio		0.122	-0.232	0.771	0.779
Calmar ratio		0.036	-0.056	0.211	0.23
Hit rate		0.5	0.495	0.504	0.51
Avg. profit / avg. loss		1.016	0.987	1.087	1.061
PnL per trade		0.527	-1.01	3.2	3.18
Sharpe ratio		0.088	-0.17	0.538	0.534
P-value		0.749	0.54	0.051	0.051

Table 13: Futures data set: robustness analysis for α - rescaled to target volatility. The experiment has been set with the values $p = 7$, and $\delta = 7$.

CCF	G_β strategy				D_α strategy				
	α	0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.011	0.013	0.006	-0.025	0.053	0.061	0.064	0.045
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.104	0.104	0.105	0.107	0.107	0.107	0.106	0.107
Maximum drawdown		-0.594	-0.535	-0.545	-0.667	-0.432	-0.393	-0.38	-0.426
Sortino ratio		0.102	0.121	0.06	-0.236	0.496	0.574	0.603	0.422
Calmar ratio		0.018	0.024	0.012	-0.038	0.123	0.156	0.169	0.106
Hit rate		0.501	0.502	0.504	0.5	0.509	0.511	0.503	0.504
Avg. profit / avg. loss		1.008	1.007	0.993	0.971	1.031	1.032	1.072	1.042
PnL per trade		0.421	0.499	0.249	-1.001	2.116	2.439	2.548	1.793
Sharpe ratio		0.071	0.084	0.042	-0.168	0.355	0.41	0.428	0.301
P-value		0.749	0.705	0.85	0.447	0.108	0.064	0.053	0.173

KM_Mod	G_β strategy				D_α strategy				
	α	0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.006	-0.001	0.006	0.006	0.048	0.036	0.038	0.025
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.103	0.102	0.1	0.099	0.108	0.108	0.108	0.108
Maximum drawdown		-0.484	-0.587	-0.517	-0.505	-0.326	-0.411	-0.35	-0.452
Sortino ratio		0.06	-0.007	0.064	0.059	0.446	0.336	0.352	0.234
Calmar ratio		0.013	-0.001	0.012	0.011	0.147	0.088	0.109	0.056
Hit rate		0.503	0.498	0.501	0.496	0.501	0.502	0.502	0.496
Avg. profit / avg. loss		0.996	1.009	1.004	1.025	1.059	1.038	1.039	1.05
PnL per trade		0.246	-0.028	0.254	0.23	1.901	1.441	1.514	1.004
Sharpe ratio		0.041	-0.005	0.043	0.039	0.319	0.242	0.254	0.169
P-value		0.852	0.983	0.847	0.862	0.148	0.274	0.25	0.446

KM_Med		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.004	0.005	0.009	0.017	0.031	0.036	0.037	0.03
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.103	0.103	0.102	0.102	0.108	0.108	0.108	0.108
Maximum drawdown		-0.551	-0.534	-0.479	-0.454	-0.414	-0.399	-0.434	-0.459
Sortino ratio		0.042	0.045	0.09	0.163	0.284	0.335	0.343	0.28
Calmar ratio		0.008	0.009	0.019	0.037	0.074	0.091	0.085	0.066
Hit rate		0.496	0.495	0.495	0.498	0.501	0.499	0.5	0.498
Avg. profit / avg. loss		1.023	1.028	1.03	1.028	1.036	1.05	1.047	1.046
PnL per trade		0.172	0.185	0.363	0.661	1.217	1.443	1.472	1.202
Sharpe ratio		0.029	0.031	0.061	0.111	0.204	0.242	0.247	0.202
P-value		0.896	0.888	0.783	0.616	0.356	0.273	0.264	0.362
SP_Mod		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		-0.011	-0.01	0	-0.007	0.033	0.027	0.029	0.031
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.103	0.102	0.102	0.102	0.108	0.108	0.108	0.108
Maximum drawdown		-0.528	-0.508	-0.43	-0.504	-0.465	-0.474	-0.467	-0.502
Sortino ratio		-0.106	-0.097	0.002	-0.073	0.309	0.253	0.271	0.283
Calmar ratio		-0.021	-0.019	0	-0.015	0.072	0.057	0.062	0.061
Hit rate		0.496	0.494	0.495	0.502	0.5	0.498	0.5	0.503
Avg. profit / avg. loss		1.004	1.013	1.021	0.984	1.044	1.043	1.037	1.027
PnL per trade		-0.435	-0.392	0.007	-0.295	1.322	1.079	1.158	1.213
Sharpe ratio		-0.073	-0.066	0.001	-0.05	0.222	0.181	0.195	0.204
P-value		0.742	0.766	0.996	0.823	0.315	0.412	0.379	0.357
SP_Med		G_β strategy				D_α strategy			
α		0.7	0.75	0.8	0.85	0.7	0.75	0.8	0.85
E[Returns]		0.007	0.022	0.02	0.004	0.03	0.031	0.035	0.036
Volatility		0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
Downside deviation		0.105	0.105	0.103	0.104	0.107	0.107	0.107	0.107
Maximum drawdown		-0.467	-0.457	-0.455	-0.44	-0.433	-0.462	-0.504	-0.44
Sortino ratio		0.071	0.209	0.191	0.039	0.282	0.29	0.323	0.333
Calmar ratio		0.016	0.048	0.044	0.009	0.069	0.067	0.069	0.081
Hit rate		0.503	0.508	0.497	0.492	0.503	0.501	0.501	0.501
Avg. profit / avg. loss		0.999	0.994	1.036	1.037	1.024	1.037	1.04	1.04
PnL per trade		0.295	0.869	0.786	0.162	1.192	1.23	1.371	1.417
Sharpe ratio		0.05	0.146	0.132	0.027	0.2	0.207	0.23	0.238
P-value		0.823	0.51	0.551	0.902	0.365	0.35	0.298	0.282

8 CO₂ emissions data

8.1 Data description

We acquire CO₂ emissions data (metric tons per capita) for the period 1990–2019 from Climate Watch. This data is depicted in Figure 18 in the appendix, and Table 14 depicts the 31 countries' names and corresponding codes.

Table 14: 31 Countries and their corresponding codes.

Country	Code	Country	Code	Country	Code	Country	Code
Austria	AUT	Belgium	BEL	Bulgaria	BGR	Croatia	HRV
Cyprus	CYP	Czech Republic	CZE	Denmark	DNK	Estonia	EST
Finland	FIN	France	FRA	Germany	DEU	Greece	GRC
Hungary	HUN	Iceland	ISL	Ireland	IRL	Italy	ITA
Latvia	LVA	Lithuania	LTU	Luxemburg	LUX	Malta	MLT
Netherlands	NLD	Norway	NOR	Poland	POL	Portugal	PRT
Romania	ROU	Slovakia	SVK	Slovenia	SVN	Spain	ESP
Sweden	SWE	Switzerland	CHE	UK	GBR		

8.2 Pipeline

Given that $n = 31$ for the CO₂ time series with a period of $T = 30$ years, we extract STS from the data with length $q = 16$ via a sliding window that is shifted by $s = 1$. Next, we apply SP_Med to cluster the STS, and calculate the lead-lag matrix $\Gamma_{n \times n}$ using a voting threshold ($\theta = 3$). From this, we rank the time series from the most leading to the most lagging by using RowSum ranking.

8.3 Results

We find that the first among the leading countries is Poland, while there are five countries that are tied for second place. These countries are Sweden, Belgium, Slovakia, Denmark and Germany. Moreover, we find that the last two lagging countries are Bulgaria and Estonia. We summarize the top ten leading countries, and the last ten lagging countries in Tables [15, 16].

By investigating the top-ranked countries Poland and Sweden, we can attempt to explain the trend in these countries' CO₂ emission rates. We note that these countries belong to the same political group, the European Union. Thus, one possibility is that they follow similar protocols in terms of regulations and restrictions. These restrictions and regulations will help alleviate CO₂ emissions.

The combustion of fossil fuels such as gasoline, natural gas and coal represents one of the main contributors to carbon dioxide pollution. Piotr Żuk, Paweł Żuk, and Pluciński (2021) suggested that indigenous coal is the primary source of mixed energy in Poland, and since the country joined the European Union in 2004, it has become clear that Poland will do much to protect its domestic coal sector, thus, will reject demands for aggressive harmonized decarbonization efforts, despite being identified as a leader with rapid emissions reduction taking place in the 1990s and early 2000s. At the same time, Kolasa-Wiecek (2015) proposed that the Polish energy sector is actively

participating in efforts to reduce emissions of greenhouse gases (GHG) into the environment by lowering the quantity of coal in the fuel mix and increasing the usage of renewable energy sources. Also, Rybak et al. (2022) posited that both Poland and Sweden are the leading countries that effectively reduce GHG and apply environmental taxes.

We note that since joining the EU in 1995, Sweden experienced a rapid drop in CO₂ emissions. In Sweden, De Luca and Pizzolante (2021) proffered that many incentives have been designed to encourage consumers to purchase cleaner automobiles that produce fewer CO₂ emissions. Also, De Luca and Pizzolante (2021) further argued that the Swedish government introduced a new bonus-penalty system for the purpose of administering incentives and taxes on light vehicles. De Luca and Pizzolante (2021) suggested this is based on the strong leadership of Germany, because it is one of the first countries in the world to formulate an approach that proved favourable to the environment, particularly in the field of transportation. For example, the German government first implemented an Eco-tax in 1999, which led to increases in the price of fuel and parking fees, and a reduction in the amount of available parking space for vehicles. Indeed, Germany is a leading country in reducing CO₂ emissions itself.

In contrast, we also perform an analysis of the two most lagging countries. In Estonia, the work of Moora et al. (2017) demonstrated that shale oil is a low-grade carbonaceous fossil fuel, which is utilized locally as their main source of energy. Burning shale oil produces high CO₂ emissions, which sets it apart from other forms of fossil fuel combustion. Moreover, the power sector appears to be the most profound emitter of CO₂ and is responsible for significant amounts of waste. Bulgaria, which has a fleet of old cars ranging from 15 to 20 years, has the extremely difficult challenge of reducing automobile pollution in practice. Moreover, a study by Hristov and Stefanov (2019) demonstrated that upgrading the transportation infrastructure does not always have a positive impact on the decrease in transportation emissions. Furthermore, the only native energy source is lignite, which has a low caloric content. Although the future of fossil fuels is uncertain after 2030, it should be noted that EU Regulation 2018/842 will not impose stringent limitations on GHG emissions in Bulgaria until 2030, as posited by Vitkov (2020).

Table 15: Top ten ranking of leading countries. Table 16: Last ten ranking of lagging countries.

Country	Rank	Country	Rank
Poland	1	Austria	22
Sweden	2	Slovenia	23
Belgium	2	Croatia	23
Slovakia	2	Luxembourg	25
Denmark	2	Romania	26
Germany	2	Malta	27
Netherlands	7	Latvia	28
Hungary	8	Lithuania	28
France	9	Bulgaria	30
UK	10	Estonia	31

9 Conclusion

We develop a clustering-driven methodology for robust detection of lead-lag relationships in high-dimensional multivariate time series, in the setting of lagged multi-factor models. We consider a collection of time series as input, and generate an expanded universe by extracting STS from each time series via a sliding window. Next, we employ several clustering approaches which include KM and SP clustering to cluster the resulting STS. After extracting the clusters, lead-lag estimates from the clusters are merged to help identify the consistent relationships that existed in the original universe.

When applied to financial data sets, our proposed methods attain promising Sharpe ratios, and are statistically significant when compared to the benchmark. In addition to the financial domain, our methods can also be applied to other fields, such as environmental sciences, which is consistent with the results of others' previous work. Thus, our method is generally applicable to a variety of multivariate time series data sets. One possible future work is to fine-tune the weights of the portfolios of lagers. In our current work related to financial data, we buy or sell baskets of lagging assets based on equal weight according to the trading signal, while adjusting the weight of the portfolios of lagers based on the ranking of the lagers. In another direction of future work, we intend to experiment with other clustering methods applicable in our setting [Cucuringu, H. Li, et al. (2020), Cucuringu, Davies, et al. (2019)], since our current work has only experimented with KM and SP clustering. Moreover, we also intend to test various alternative ranking algorithms [Bradley and Terry (1952), Fogel, d'Aspremont, and Vojnovic (2014), Page et al. (1999), Cucuringu (2016), He et al. (2022), d'Aspremont, Cucuringu, and Tyagi (2021)], and consider momentum reversal strategies within the trading pipeline. Lastly, one could further explore the more challenging mixed membership model defined in Section 3.

References

- [1] Jakob Albers et al. “Fragmentation, price formation and cross-impact in bitcoin markets”. In: *Applied Mathematical Finance* 28.5 (2021), pp. 395–448.
- [2] David Arthur and Sergei Vassilvitskii. *k-means++: The advantages of careful seeding*. Tech. rep. Stanford, 2006.
- [3] David H Bailey and Marcos Lopez De Prado. “The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality”. In: *The Journal of Portfolio Management* 40.5 (2014), pp. 94–107.
- [4] Afonso S Bandeira et al. “Multireference alignment using semidefinite programming”. In: *Proceedings of the 5th conference on Innovations in theoretical computer science*. 2014, pp. 459–470.
- [5] Stefanos Bennett, Mihai Cucuringu, and Gesine Reinert. “Lead-lag detection and network clustering for multivariate time series with an application to the US equity market”. In: *Machine Learning* (2022), pp. 1–42.
- [6] Nicolas Boumal et al. “Heterogeneous multireference alignment: A single pass approach”. In: *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*. IEEE. 2018, pp. 1–6.
- [7] Ralph Allan Bradley and Milton E Terry. “Rank analysis of incomplete block designs: I. The method of paired comparisons”. In: *Biometrika* 39.3/4 (1952), pp. 324–345.
- [8] Giuseppe Buccheri, Fulvio Corsi, and Stefano Peluso. “High-frequency lead-lag effects and cross-asset linkages: a multi-asset lagged adjustment model”. In: *Journal of Business & Economic Statistics* 39.3 (2021), pp. 605–621.
- [9] Jay Cao, Jacky Chen, and John Hull. “A neural network approach to understanding implied volatility movements”. In: *Quantitative Finance* 20.9 (2020), pp. 1405–1413.
- [10] Alvaro Cartea, Ryan Donnelly, and Sebastian Jaimungal. “Enhancing trading strategies with order book signals”. In: *Applied Mathematical Finance* 25.1 (2018), pp. 1–35.
- [11] Rama Cont. “Empirical properties of asset returns: stylized facts and statistical issues”. In: *Quantitative finance* 1.2 (2001), p. 223.
- [12] Paul Cotofrei and Kilian Stoffel. “Classification rules+ time= temporal rules”. In: *International Conference on Computational Science*. Springer. 2002, pp. 572–581.
- [13] Mihai Cucuringu. “Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and SDP synchronization”. In: *IEEE Transactions on Network Science and Engineering* 3.1 (2016), pp. 58–79.
- [14] Mihai Cucuringu, Peter Davies, et al. “SPONGE: A generalized eigenproblem for clustering signed networks”. In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 1088–1098.

- [15] Mihai Cucuringu, Huan Li, et al. “Hermitian matrices for clustering directed graphs: insights and applications”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 983–992.
- [16] Yan Cui, Jun Yang, and Zhou Zhou. “State-domain change point detection for nonlinear time series regression”. In: *Journal of Econometrics* (2021).
- [17] Alexandre d’Aspremont, Mihai Cucuringu, and Hemant Tyagi. “Ranking and synchronization from pairwise measurements via SVD”. In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 866–928.
- [18] Giovanni De Luca and Federica Pizzolante. “Detecting Leaders Country from Road Transport Emission Time-Series”. In: *Environments* 8.3 (2021), p. 18.
- [19] Chang Dong et al. “Subsequence Time Series Clustering-Based Unsupervised Approach for Anomaly Detection of Axial Piston Pumps”. In: *IEEE Transactions on Instrumentation and Measurement* 72 (2023), pp. 1–12. DOI: [10.1109/TIM.2023.3264045](https://doi.org/10.1109/TIM.2023.3264045).
- [20] Felix Drinkall, Stefan Zohren, and Janet B Pierrehumbert. “Forecasting COVID-19 Caseloads Using Unsupervised Embedding Clusters of Social Media Posts”. In: *arXiv preprint arXiv:2205.10408* (2022).
- [21] Fajwel Fogel, Alexandre d’Aspremont, and Milan Vojnovic. “Serialrank: Spectral ranking using seriation”. In: *Advances in neural information processing systems* 27 (2014).
- [22] David F Gleich and Lek-heng Lim. “Rank aggregation via nuclear norm minimization”. In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2011, pp. 60–68.
- [23] Yixuan He et al. “GNNRank: Learning global rankings from pairwise comparisons via directed graph neural networks”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 8581–8612.
- [24] R Hristov and S Stefanov. “Study of the potential for reducing CO₂ emissions from road transport in Bulgaria”. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 614. 1. IOP Publishing. 2019, p. 012009.
- [25] Peter J Huber. “Pairwise comparison and ranking: optimum properties of the row sum procedure”. In: *The annals of mathematical statistics* (1963), pp. 511–520.
- [26] Katsuya Ito and Ryuta Sakemoto. “Direct estimation of lead–lag relationships using multinomial dynamic time warping”. In: *Asia-Pacific Financial Markets* 27.3 (2020), pp. 325–342.
- [27] Narasimhan Jegadeesh, Jiang Luo, et al. “Momentum and short-term reversals: theory and evidence”. In: *Nanyang Business School Research Paper* 22-13 (2022).
- [28] Narasimhan Jegadeesh and Sheridan Titman. “Profitability of momentum strategies: An evaluation of alternative explanations”. In: *The Journal of finance* 56.2 (2001), pp. 699–720.

- [29] Eamonn Keogh and Jessica Lin. “Clustering of time-series subsequences is meaningless: implications for previous and future research”. In: *Knowledge and information systems* 8.2 (2005), pp. 154–177.
- [30] Alicja Kolasa-Wiecek. “Stepwise multiple regression method of greenhouse gas emission modeling in the energy sector in Poland”. In: *Journal of Environmental Sciences* 30 (2015), pp. 47–54.
- [31] Oliver Ledoit and Michael Wolf. “Robust performance hypothesis testing with the Sharpe ratio”. In: *Journal of Empirical Finance* 15.5 (2008), pp. 850–859.
- [32] Chung-Sheng Li, Philip S Yu, and Vittorio Castelli. “MALM: A framework for mining sequence database at multiple abstraction levels”. In: *Proceedings of the seventh international conference on Information and knowledge management*. 1998, pp. 267–272.
- [33] Yongli Li et al. “Dynamic patterns of daily lead-lag networks in stock markets”. In: *Quantitative Finance* 21.12 (2021), pp. 2055–2068.
- [34] Bryan Lim, Stefan Zohren, and Stephen Roberts. “Enhancing time-series momentum strategies using deep neural networks”. In: *The Journal of Financial Data Science* 1.4 (2019), pp. 19–38.
- [35] Jessica Lin, Eamonn Keogh, and Wagner Truppel. “Clustering of streaming time series is meaningless”. In: *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. 2003, pp. 56–65.
- [36] Yutong Lu, Gesine Reinert, and Mihai Cucuringu. “Co-trading networks for modeling dynamic interdependency structures and estimating high-dimensional covariances in US equity markets”. In: *arXiv preprint arXiv:2302.09382* (2023).
- [37] Yutong Lu, Gesine Reinert, and Mihai Cucuringu. “Trade co-occurrence, trade flow decomposition, and conditional order imbalance in equity markets”. In: *arXiv preprint arXiv:2209.10334* (2022).
- [38] Navin Madicar et al. “Parameter-free subsequences time series clustering with various-width clusters”. In: *2013 5th International Conference on Knowledge and Smart Technology (KST)*. IEEE. 2013, pp. 150–155.
- [39] Nikolas Michael, Mihai Cucuringu, and Sam Howison. “Option Volume Imbalance as a predictor for equity market returns”. In: *arXiv preprint arXiv:2201.09319* (2022).
- [40] Deborah Miori and Mihai Cucuringu. “Returns-Driven Macro Regimes and Characteristic Lead-Lag Behaviour between Asset Classes”. In: *arXiv preprint arXiv:2209.00268* (2022).
- [41] Harri Moora et al. “Determination of biomass content in combusted municipal waste and associated CO₂ emissions in Estonia”. In: *Energy Procedia* 128 (2017), pp. 222–229.
- [42] Andrew Ng, Michael Jordan, and Yair Weiss. “On spectral clustering: Analysis and an algorithm”. In: *Advances in neural information processing systems* 14 (2001).
- [43] Lawrence Page et al. *The PageRank citation ranking: Bringing order to the web*. Tech. rep. Stanford InfoLab, 1999.

- [44] Amelia Perry et al. “The sample complexity of multireference alignment”. In: *SIAM Journal on Mathematics of Data Science* 1.3 (2019), pp. 497–517.
- [45] Daniel Poh, Stephen Roberts, and Stefan Zohren. “Transfer ranking in finance: applications to cross-sectional momentum with data scarcity”. In: *arXiv preprint arXiv:2208.09968* (2022).
- [46] N Radhakrishnan, James D Wilson, and Philipos C Loizou. “An alternate partitioning technique to quantify the regularity of complex time series”. In: *International Journal of Bifurcation and Chaos* 10.07 (2000), pp. 1773–1779.
- [47] Thanawin Rakthanmanon et al. “Time series epenthesis: Clustering time series streams requires ignoring some data”. In: *2011 IEEE 11th International Conference on Data Mining*. IEEE. 2011, pp. 547–556.
- [48] Sura Rodpongpun, Vit Niennattrakul, and Chotirat Ann Ratanamahatana. “Selective subsequence time series clustering”. In: *Knowledge-Based Systems* 35 (2012), pp. 361–368.
- [49] Jakob Runge et al. “Detecting and quantifying causal associations in large nonlinear time series datasets”. In: *Science advances* 5.11 (2019), eaau4996.
- [50] Aurelia Rybak et al. “The Impact of Environmental Taxes on the Level of Greenhouse Gas Emissions in Poland and Sweden”. In: *Energies* 15.12 (2022), p. 4465.
- [51] Biplob Kumer Sarker et al. “Parallel algorithms for mining association rules in time series data”. In: *International Symposium on Parallel and Distributed Processing and Applications*. Springer. 2003, pp. 273–284.
- [52] Christian Schittenkopf, Peter Tiño, and Georg Dorffner. “The benefit of information reduction for trading strategies”. In: *Applied Economics* 34.7 (2002), pp. 917–930.
- [53] Jianbo Shi and Jitendra Malik. “Normalized cuts and image segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (2000), pp. 888–905.
- [54] Alik Sokolov et al. “Assessing the Impact of Sustainability on Fund Flows: An Excess Information Approach and US-Based Case Study”. In: *The Journal of Impact and ESG Investing* (2022).
- [55] Wee Ling Tan, Stephen Roberts, and Stefan Zohren. “Spatio-Temporal Momentum: Jointly Learning Time-Series and Cross-Sectional Strategies”. In: *arXiv preprint arXiv:2302.10175* (2023).
- [56] Peter Tino, Christian Schittenkopf, and Georg Dorffner. “Temporal pattern recognition in noisy non-stationary time series based on quantization into symbolic streams. Lessons learned from financial volatility trading.” In: (2000).
- [57] Konstantinos Tolikas. “The lead-lag relation between the stock and the bond markets”. In: *The European Journal of Finance* 24.10 (2018), pp. 849–866.
- [58] Kuniaki Uehara and Mitsuomi Shimada. “Extraction of primitive motion and discovery of association rules from human motion data”. In: *Progress in Discovery Science*. Springer, 2002, pp. 338–348.

- [59] Rahul Kumar Vijay and Satyasai Jagannath Nanda. “Earthquake pattern analysis using subsequence time series clustering”. In: *Pattern Analysis and Applications* 26.1 (2023), pp. 19–37.
- [60] Nikolai Vitkov. “Greenhouse Emissions And Prospects For Local Fuel TPPs In Bulgaria”. In: *2020 12th Electrical Engineering Faculty Conference (BulEF)*. IEEE. 2020, pp. 1–5.
- [61] Milena Vuletić, Felix Prenzel, and Mihai Cucuringu. “Fin-GAN: Forecasting and Classifying Financial Time Series via Generative Adversarial Networks”. In: *Available at SSRN 4328302* (2023).
- [62] Kieran Wood, Sven Giegerich, et al. “Trading with the Momentum Transformer: An Intelligent and Interpretable Architecture”. In: *arXiv preprint arXiv:2112.08534* (2021).
- [63] Kieran Wood, Stephen Roberts, and Stefan Zohren. “Slow momentum with fast reversion: A trading strategy using deep learning and changepoint detection”. In: *arXiv preprint arXiv:2105.13727* (2021).
- [64] Di Wu et al. “Detecting leaders from correlated time series”. In: *International Conference on Database Systems for Advanced Applications*. Springer. 2010, pp. 352–367.
- [65] Takehisa Yairi, Yoshiyo Kato, and Koichi Hori. “Fault detection by mining association rules from house-keeping data”. In: *proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space*. Vol. 18. Citeseer. 2001, p. 21.
- [66] Can-Zhong Yao and Hong-Yu Li. “Time-varying lead-lag structure between investor sentiment and stock market”. In: *The North American Journal of Economics and Finance* 52 (2020), p. 101148.
- [67] Yichi Zhang et al. *Supplemental Material: Robust Detection of Lead-Lag Relationships in Lagged Multi-Factor Models*. Mendeley Data. Version V4. 2023. DOI: [10.17632/2djzdzjvn96.2](https://doi.org/10.17632/2djzdzjvn96.2).
- [68] Seyedjamal Zolhavarieh, Saeed Aghabozorgi, and Ying Wah Teh. “A review of subsequence time series clustering”. In: *The Scientific World Journal* 2014 (2014).
- [69] Piotr Żuk, Paweł Żuk, and Przemysław Pluciński. “Coal basin in Upper Silesia and energy transition in Poland in the context of pandemic: The socio-political diversity of preferences in energy and environmental policy”. In: *Resources Policy* 71 (2021), p. 101987.

A Appendix

A.1 Synthetic data experiment: spectral clustering

Figures [16, 17] display the result of applying SP clustering, the voting matrix and the error matrix based on mode and median estimation without and with voting threshold with different $k = \{1, 2, 3\}$.

Figure 16: Top panel: Voting matrix without voting threshold ($\theta = 1$). Middle panel: Error matrix based on mode estimation without the voting threshold ($\theta = 1$). Bottom panel: Error matrix based on median estimation without the voting threshold ($\theta = 1$).

Spectral clustering

Homogeneous Setting							Heterogeneous Setting						
TS 1	0	12	11	11	11	11	TS 1	0	9	9	0	0	0
TS 2	12	0	11	11	10	10	TS 2	9	0	11	0	0	0
TS 3	11	11	0	11	11	11	TS 3	9	11	0	0	0	0
TS 4	11	11	11	0	11	11	TS 4	0	0	0	0	11	10
TS 5	11	10	11	11	0	13	TS 5	0	0	0	11	0	10
TS 6	11	10	11	11	13	0	TS 6	0	0	0	10	10	0
	TS 1	TS 2	TS 3	TS 4	TS 5	TS 6		TS 1	TS 2	TS 3	TS 4	TS 5	TS 6
TS 1	0	0	0	0	0	0	TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0	TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0	TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0	TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0	TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0	TS 6	0	0	0	0	0	0
	TS 1	TS 2	TS 3	TS 4	TS 5	TS 6		TS 1	TS 2	TS 3	TS 4	TS 5	TS 6
TS 1	0	0	0	0	0	0	TS 1	0	0	0	0	0	0
TS 2	0	0	0	0	0	0	TS 2	0	0	0	0	0	0
TS 3	0	0	0	0	0	0	TS 3	0	0	0	0	0	0
TS 4	0	0	0	0	0	0	TS 4	0	0	0	0	0	0
TS 5	0	0	0	0	0	0	TS 5	0	0	0	0	0	0
TS 6	0	0	0	0	0	0	TS 6	0	0	0	0	0	0
	TS 1	TS 2	TS 3	TS 4	TS 5	TS 6		TS 1	TS 2	TS 3	TS 4	TS 5	TS 6

$k = 1$	$k = 2$	$k = 3$

Figure 17: Top panel: Voting matrix with voting threshold ($\theta = 6$). Middle panel: Error matrix based on mode estimation with the voting threshold ($\theta = 6$). Bottom panel: Error matrix based on median estimation with the voting threshold ($\theta = 6$).

A.2 CO2 emissions plots

Figure 18 shows 31 countries' CO2 emissions data (metric tons per capita) for the period 1990–2019 from Climate Watch.

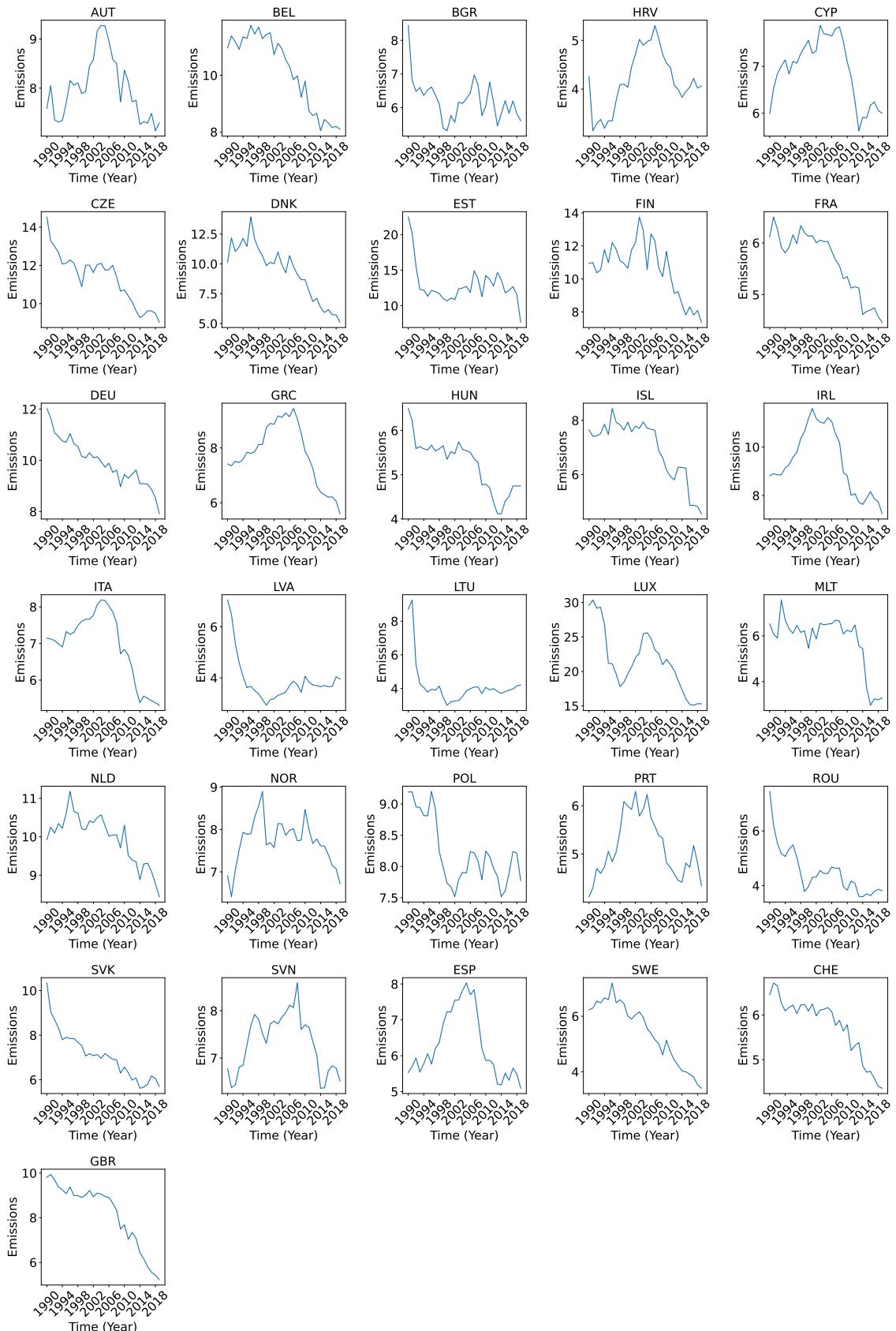


Figure 18: 31 countries' CO2 emissions (metric tons per capita) during the period 1990–2019.

A.3 Futures data set details

Tables [17, 18, 19, 20, 21, 22, 23, 24] show the futures contracts we used and its description from the Pinnacle Data Corp CLC Database. The data is based on ratio-adjusted methods, which removes the contract-to-contract gap, yet it will not go negative as it reduces the size of the price bars if they go lower and increases them if they go higher.

Table 17: Grains

Identifier	Description
KW	WHEAT, KC
MW	WHEAT, MINN
NR	ROUGH RICE
W_	WHEAT, CBOT
ZC	CORN, Electronic
ZL	SOYBEAN OIL, Electronic
ZM	SOYBEAN MEAL, Electronic
ZO	OATS, Electronic
ZR	ROUGH RICE, Electronic
ZS	SOYBEANS, Electronic
ZW	WHEAT, Electronic

Table 18: Meats

Identifier	Description
DA	MILK III, Comp.
ZF	FEEDER CATTLE, Electronic
ZT	LIVE CATTLE, Electronic
ZZ	LEAN HOGS, Electronic

Table 19: Wood fibre

Identifier	Description
LB	LUMBER

Table 20: Metals

Identifier	Description
ZG	GOLD, Electronic
ZI	SILVER, Electronic
ZP	PLATINUM, electronic
ZA	PALLADIUM, electronic
ZK	COPPER, electronic

Table 21: Indexes

Identifier	Description
AX	GERMAN DAX INDEX
CA	CAC40 INDEX
DX	US DOLLAR INDEX
EN	NASDAQ, MINI
ES	S & P 500, MINI
GI	GOLDMAN SAKS C. I.
LX	FTSE 100 INDEX
MD	S & P 400 (Mini electronic)
NK	NIKKEI INDEX
SC	S & P 500, composite

Table 22: Bonds

Identifier	Description
DT	EURO BOND (BUND)
FB	T-NOTE, 5yr composite
GS	GILT, LONG BOND
SS	STERLING, SHORT
TY	T-NOTE, 10yr composite
TU	T-NOTES, 2yr composite
US	T-BONDS, composite
UB	EURO BOBL
UZ	EURO SCHATZ

Table 23: Currency

Identifier	Description
AN	AUSTRALIAN \$\$ composite
BN	BRITISH POUND, composite
CN	CANADIAN \$\$ composite
EC	EURODOLLAR, composite
FN	EURO, composite
JN	JAPANESE YEN, composite
MP	MEXICAN PESO
SN	SWISS FRANC, composite

Table 24: Oils

Identifier	Description
ZB	RBOB, Electronic
ZH	HEATING OIL, electronic
ZN	NATURAL GAS, electronic
ZU	CRUDE OIL, Electronic