Application of Invariant Feature-Based Multi-Reference Alignment on Lead-Lag Detection

Danni Shi
danni.shi@eng.ox.ac.uk
University of Oxford
Oxford, UK

Mihai Cucuringu
mihai.cucuringu@stats.ox.ac.uk
Department of Statistics, University of Oxford
Oxford, UK
The Alan Turing Institute
London, UK

Jan-Peter Calliess
jan-peter.calliess@oxford-man.ox.ac.uk
University of Oxford
Oxford, UK

ABSTRACT

In time series analysis, numerous methods have been developed to detect, measure and apprehend lead-lag relationships between variables, which refers to the time-delay of patterns in a time series relative to the other. Often, the extent of lead-lag between two time series is directly measured with a similarity metric. This approach suffers inconsistency and inaccuracy especially under high-noise regimes.

This work represents several frameworks of Multireference alignment (MRA), the estimation of a unified, de-noised latent signal from a population of time series with linearly-shifted and noisy patterns, and their application on lead-lag detection. The recovered latent signal is treated as the reference vector whose lead-lag metric is calculated against other time series. The relative lead-lag between two actual time series is then obtained as the difference between their lags against the reference vector. Results show that models with MRA-induced intermediates outperform the simplistic models relying on pairwise similarity in predicting the relative shifts between time series when the signal-to-noise ratio (SNR) is low. The effect of clustering the time series advancing the recovery of the latent signals is also investigated

The results of lead-lag detection can be used on trading stock returns. We developed a trading framework that utilises the relative shift estimations to construct financial signals from the leading time series and trade on multiple portfolios of the lagging time series.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

High-dimensional time series, lead-lag, clustering, financial markets

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Danni Shi, Mihai Cucuringu, and Jan-Peter Calliess. 2018. Application of Invariant Feature-Based Multi-Reference Alignment on Lead-Lag Detection. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Time series forecasting is a major topic in data science and has been extensively researched from multiple perspectives. Forecasting financial time series is especially challenging due to low SNR and the deficiency of auto-correlation. Yet, in the financial market, it has been observed that the prices of certain groups of assets respond to events and factors quicker than others. In multivariate time series models, market factors can create similar patterns of change in the financial time series of different assets, each with a different time delay. This is what we call a lead-lag structure amongst the time series variables. We aim to construct a measure that estimates the extent and length of lead-lag between two stocks. In this paper, the extent of lead-lag is represented as the maximum cross-correlation between two time series up to a shift. The corresponding length of lead-lag is the number of timescale shifts (as we primarily works with discrete-time data in finance) to achieve the maximum cross-correlation.

Current methods of lead-lag detection often rely on computing a lead-lag score between a pair of time series (some refs here), which are direct and effective in some use cases. However, several problems arise with this approach:

- Non-robust: At low SNR, pairwise measurement is significantly interfered by the noise in the data and hence become inaccurate
- Inconsistent: Pairwise lag predictions need to be synchronized for reliable applications. However, the lag between a pair of time series may not agree with the others. For example, estimations may show that variable X and Y each leads variable Z by 1 day, but that X lags Y by 1 day too.
- Computationally expensive: The computation of the pairwise metric of N time series has a complexity $O(N)$, making it non-scalable to large datasets.

This result can be subsequently used for price forecasts and devising trading strategies.

We model the financial time series as the superposition of multiple shifted factor series and noise. We aim to recover the factor series with heterogeneous multireference alignment (MRA)[1], a method initially proposed to solve problems in cryo-electron microscopy. The original method aims to recover the original signal from cyclically shifted copies of noisy observations while the financial data we work with contains non-cyclic shifts. We make an educated assumption that the shifts are small compared to the length of the time series so that a large proportion of the time series agrees with the cyclically shifted version. We hope that the violation of the rules based on cyclical shifts is negligible and that the method is still effective. If we successfully recover the factor series, we can apply cross-correlation analysis[2] of the financial time series against the factor series and find the optimal lags as the time lag of the maximum correlation. We obtain a summarised lead-lag relationship by ranking the optimal lags for each factor series.

2 LEAD-LAG DETECTION BASED ON CROSS-CORRELATION

2.1 Pairwise Lag Measurement

2.2 Relative Lag Measurement with a Reference Signal

3 MULTIREFERENCE ALIGNMENT

3.1 Homogeneous MRA

3.2 Heterogeneous MRA

3.3 MRA Based on Correlation and Synchronization

4 EXPERIMENTS AND RESULTS

4.1 Synthetic Data

4.2 Financial Data

5 DISCUSSIONS

6 CONCLUSIONS

7 REFERENCES

The use of BibTEX for the preparation and formatting of one's references is strongly recommended. Authors' names should be complete — use full first names ("Donald E. Knuth") not initials ("D. E. Knuth") — and the salient identifying features of a reference should be included: title, year, volume, number, pages, article DOI, etc.

The bibliography is included in your source document with these two commands, placed just before the `\end{document}` command:

```
\bibliographystyle{ACM-Reference-Format}
\bibliography{bibfile}
```

where "`bibfile`" is the name, without the "`.bib`" suffix, of the BibTEX file.

Citations and references are numbered by default. A small number of ACM publications have citations and references formatted in the "author year" style; for these exceptions, please include this command in the **preamble** (before the command "`\begin{document}`") of your LATEX source:

```
\citestyle{acmauthoryear}
```

Some examples. A paginated journal article [?], an enumerated journal article [?], a reference to an entire issue [?], a monograph (whole book) [?], a monograph/whole book in a series (see 2a in spec. document) [?], a divisible-book such as an anthology or compilation [?] followed by the same example, however we only output the series if the volume number is given [?] (so Editor00a's series should NOT be present since it has no vol. no.), a chapter in a divisible book [?], a chapter in a divisible book in a series [?], a multi-volume work as book [?], a couple of articles in a proceedings (of a conference, symposium, workshop for example) (paginated proceedings article) [? ?], a proceedings article with all possible elements [?], an example of an enumerated proceedings article [?], an informally published work [?], a couple of preprints [? ?], a doctoral dissertation [?], a master's thesis: [?], an online document / world wide web resource [? ? ?], a video game (Case 1) [?] and (Case 2) [?] and [?] and (Case 3) a patent [?], work accepted for publication [?], 'YYYYb'-test for prolific author [?] and [?]. Other cites might contain 'duplicate' DOI and URLs (some SIAM articles) [?]. Boris / Barbara Beeton: multi-volume works as books [?] and [?]. A couple of citations with DOIs: [? ?]. Online citations: [? ? ?]. Artifacts: [?] and [?].

8 ACKNOWLEDGMENTS

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

This section has a special environment:

```
\begin{acks}
...
\end{acks}
```

so that the information contained therein can be more easily collected during the article metadata extraction phase, and to ensure consistency in the spelling of the section heading.

Authors should not prepare this section as a numbered or unnumbered `\section`; please use the "`acks`" environment.

9 APPENDICES

If your work needs an appendix, add it before the "`\end{document}`" command at the conclusion of your source document.

Start the appendix with the "`appendix`" command:

```
\appendix
```

and note that in the appendix, sections are lettered, not numbered. This document has two appendices, demonstrating the section and subsection identification method.

[1] The method recovers the original signals and mixing probabilities of the classes in the data, but does not seem to assign class membership of the observations. We may need to find a way to match the observations with signals. Also, each of our observations can contain a superposition of multiple factors. Not sure how we can split them up into observations of single factors, which seem to be the assumption of the method we are trying to apply.

[2] This method may not work well when the noise level is high.

10 MULTI-LANGUAGE PAPERS

Papers may be written in languages other than English or include titles, subtitles, keywords and abstracts in different languages (as a rule, a paper in a language other than English should include an English title and an English abstract). Use `language=...` for every language used in the paper. The last language indicated is the main language of the paper. For example, a French paper with additional titles and abstracts in English and German may start with the following command

```
\documentclass[sigconf, language=english, language=german,
               language=french]{acmart}
```

The title, subtitle, keywords and abstract will be typeset in the main language of the paper. The commands `\translatedXXX`, `XXX` begin title, subtitle and keywords, can be used to set these elements in the other languages. The environment `translatedabstract` is used to set the translation of the abstract. These commands and environment have a mandatory first argument: the language of the second argument. See `sample-sigconf-i13n.tex` file for examples of their usage.

11 SIGCHI EXTENDED ABSTRACTS

The "`sigchi-a`" template style (available only in LaTeX and not in Word) produces a landscape-orientation formatted article, with a wide left margin. Three environments are available for use with the "`sigchi-a`" template style, and produce formatted output in the margin:

`sidebar`: Place formatted text in the margin.
`marginfigure`: Place a figure in the margin.
`margintable`: Place a table in the margin.

ACKNOWLEDGMENTS

To Robert, for the bagels and explaining CMYK and color spaces.

A RESEARCH METHODS

A.1 Part One

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi malesuada, quam in pulvinar varius, metus nunc fermentum urna, id sollicitudin purus odio sit amet enim. Aliquam ullamcorper eu ipsum vel mollis. Curabitur quis dictum nisl. Phasellus vel semper risus, et lacinia dolor. Integer ultricies commodo sem nec semper.

A.2 Part Two

Etiam commodo feugiat nisl pulvinar pellentesque. Etiam auctor sodales ligula, non varius nibh pulvinar semper. Suspendisse nec lectus non ipsum convallis congue hendrerit vitae sapien. Donec at laoreet eros. Vivamus non purus placerat, scelerisque diam eu, cursus ante. Etiam aliquam tortor auctor efficitur mattis.

B ONLINE RESOURCES

Nam id fermentum dui. Suspendisse sagittis tortor a nulla mollis, in pulvinar ex pretium. Sed interdum orci quis metus euismod, et sagittis enim maximus. Vestibulum gravida massa ut felis suscipit congue. Quisque mattis elit a risus ultrices commodo venenatis eget dui. Etiam sagittis eleifend elementum.

Nam interdum magna at lectus dignissim, ac dignissim lorem rhoncus. Maecenas eu arcu ac neque placerat aliquam. Nunc pulvinar massa et mattis lacinia.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009