TTIC 31220 Final Project: Toward a Reduced-Form Factor Portfolio

Kabir Sawhney, Joseph Denby

ksawhney@uchicago.edu, jgdenby@uchicago.edu

Abstract

This is the layout specification and template definition for the INTERSPEECH 2016 Conference, which will be held in San Francisco, California during September 8-12, 2016. This template has been generated from previous INTERSPEECH templates. The format is essentially the one used for IEEE ICASSP conferences. The maximum number of pages is 5. The 5th page may be used exclusively for references. Index terms should be included as shown below.

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

The returns on different financial assets are influenced by a similar set of underlying risks, including macroeconomic and market risks. In financial theory, the return of any given asset can be represented as a linear combination of returns these risks plus an error term. However, there is no consensus set of risk factors, and a great deal of research has gone into studying what the appropriate set of risk factors are from which asset returns are constructed.

Some researchers have applied linear dimensionality reduction techniques to large factor universes, with a focus on linear PCA, to statistically derive latent factor representations. The goal of our project is to extend this work using non-linear dimensionality reduction, with a focus on the techniques learned in the course.

To study this question, we found a set of factors designed to replicate returns to macroeconomic and financial market risks over time. We applied dimensionality reduction to this data, then used the lower-dimensional representation as an input to a linear regression model with stock returns as the response variable. We found that, when seeking to predict returns on the broad equity market and individual stocks, non-linear dimensionality reduction generally did not perform better than linear PCA or our theoretically-constructed benchmark.

2. Related Work

The inspiration for our project came from Kozak, Nagel, and Santosh (2017) [1], who argued that reduced-form factor models tend to perform well empirically in predicting asset prices. In their paper, the authors use linear PCA to derive a low-dimensional factor representation, from a set of factors defined by Novy-Marx and Velikov (2015) [2]. Then, they show that the principal components derived from their analysis perform about as well as their benchmark, the Fama-French three-factor model [3], in explaining the variation in returns on the portfolios they examined.

Other authors have applied PCA to a variety of different types of financial data. A'it-Sahalia and Xiu (2015) [4] apply PCA to high-frequency data, analyzing the covariance structure of the constituents of the S&P 100 Index on a microsecond-by-microsecond level. Yang (2015) [5] applied linear PCA to the returns of the Australian stock market index and its constituents, with a focus on deriving explanations for which risk factor each principal component represented.

In our review, we found linear PCA being applied to a very wide range of different data sets and research problems, but did not find any analysis which sought to apply any other dimensionality reduction technique. The explanation for this gap seems to be the difficulty of interpretation when using non-linear techniques – principal components derived using the eigenvector method generally have a theoretical analogue, while components derived using other methods don't necessary lend themselves easily to interpretation.

2.1. Basic layout features

- Proceedings will be printed in DIN A4 format. Authors must submit their papers in DIN A4 format.
- Two columns are used except for the title part and possibly for large figures that need a full page width.
- Left and right margin are 20 mm each.
- Column width is 80 mm.
- Spacing between columns is 10 mm.
- Top margin 25 mm (except for the first page which is 30 mm to the title top).
- Bottom margin is 35 mm.
- Text height (without headers and footers) is maximum 235 mm.
- Headers and footers must be left empty. They will be added for printing and the INTERSPEECH 2016 media.
- Check indentations and spacings by comparing to this example file (in PDF).

2.1.1. Headings

Section headings are centered in boldface with the first word capitalized and the rest of the heading in lower case. Subheadings appear like major headings, except they start at the left margin in the column. Sub-sub-headings appear like subheadings, except they are in italics and not boldface. See the examples given in this file. No more than 3 levels of headings should be used.

2.2. Text font

Times or Times Roman font is used for the main text. Font size in the main text must be 9 points, and in the References section 8 points. Other font types may be used if needed for special

purposes. It is VERY IMPORTANT that while making the final PDF file, you embed all used fonts!

LaTeX users: users should use Adobe Type 1 fonts such as Times or Times Roman. These are used automatically by the INTERSPEECH2016.sty style file. Authors must not use Type 3 (bitmap) fonts.

2.3. Figures

All figures must be centered on the column (or page, if the figure spans both columns). Figure captions should follow each figure and have the format given in Figure 1.

Figures should preferably be line drawings. If they contain gray levels or colors, they should be checked to print well on a high-quality non-color laser printer.

Graphics (i. e., illustrations, figures) must not use stipple fill patterns because they will not reproduce properly in Adobe PDF. Please use only SOLID FILL COLORS.

Figures which span 2 columns (i. e., occupy full page width) should be placed at the top or bottom of the page.

2.4. Tables

An example of a table is shown as Table 1. The caption text may be above or below the table.

Table 1: *This is an example of a table.*

ratio	decibels
1/10	-20
1/1	0
2/1	≈ 6
3.16/1	10
10/1	20
100/1	40
1000/1	60

2.5. Equations

Equations should be placed on separate lines and numbered. Examples of equations are given below. Particularly,

$$x(t) = s(f_\omega(t)) \tag{1}$$

where $f_\omega(t)$ is a special warping function

$$f_\omega(t) = \frac{1}{2\pi j} \oint_C \frac{\nu^{-1k} \mathrm{d}\nu}{(1-\beta\nu^{-1})(\nu^{-1}-\beta)} \tag{2}$$

A residue theorem states that

$$\oint_C F(z)\,\mathrm{d}z = 2\pi j \sum_k \mathrm{Res}[F(z), p_k] \tag{3}$$

Applying (3) to (1), it is straightforward to see that

$$1 + 1 = \pi \tag{4}$$

Finally we have proven the secret theorem of all speech sciences. No more math is needed to show how useful the result is!

Proper typesetting helps to understand the content faster. The preferred way of printing vectors and matrices is:

$$\boldsymbol{x}_i,\ \boldsymbol{\alpha},\ \boldsymbol{X} \tag{5}$$

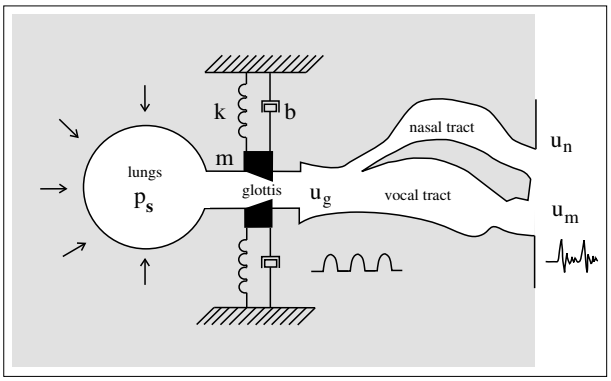

Figure 1: *Schematic diagram of speech production.*

2.6. Hyperlinks

For technical reasons, the proceedings editor will strip all active links from the papers during processing. Hyperlinks can be included in your paper, if written in full, e. g. "http://www.foo.com/index.html". The link text must be all black. Please make sure that they present no problems in printing to paper.

2.7. Multimedia files

The INTERSPEECH 2016 organizing committee offers the possibility to submit multimedia files. These files are meant for audio-visual illustrations that cannot be conveyed in text, tables and graphs. Just like you would when including graphics, make sure that you have sufficient author rights to the multimedia materials that you submit for publication. The proceeding media will NOT contain readers or players, so be sure to use widely accepted file formats, such as MPEG, Windows WAVE PCM (.wav) or Windows Media Video (.wmv) using standard codecs. The files you submit will be accessible from the abstract cards on the media and via a bookmark in the manuscript. From within the manuscript, refer to a multimedia illustration by its filename. Use short file names without blanks.

2.8. Page numbering

Final page numbers will be added later to the document electronically. *Don't make any footers or headers!*

2.9. References

The reference format is the standard IEEE one. References should be numbered in order of appearance, for example [1], [2], [3, pp. 417–422], and [4].

2.10. Abstract

The total length of the abstract is limited to 200 words. The abstract included in your paper and the one you enter during web-based submission must be identical. Avoid non-ASCII characters or symbols as they may not display correctly in the abstract book.

2.11. Author affiliation

Please list country names as part of the affiliation for each country.

2.12. Submitted files

Authors are requested to submit PDF files of their manuscripts. You can use commercially available tools or for instance http://www.pdfforge.org/products/pdfcreator. The PDF file should comply with the following requirements: (a) there must be no PASSWORD protection on the PDF file at all; (b) all fonts must be embedded; and (c) the file must be text searchable (do CTRL-F and try to find a common word such as 'the'). The proceedings editors (Causal Productions) will contact authors of non-complying files to obtain a replacement. In order not to endanger the preparation of the proceedings, papers for which a replacement is not provided timely will be withdrawn.

3. Discussion

This is the discussion. This is the discussion. This is the discussion. Is there any discussion?

This is the next paragraph of the discussion. And the last sentence of it.

4. Conclusions

Authors must proofread their PDF file prior to submission to ensure it is correct. Authors should not rely on proofreading the Word file. Please proofread the PDF file before it is submitted.

5. Acknowledgements

The ISCA Board would like to thank the organizing committees of the past INTERSPEECH conferences for their help and for kindly providing the template files.

6. References

[1] S. B. Davis and P. Mermelstein, "Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.

[2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[3] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. New York: Springer, 2009.

[4] F. Lastname1, F. Lastname2, and F. Lastname3, "Title of your INTERSPEECH 2016 publication," in *INTERSPEECH 2016 – 16th Annual Conference of the International Speech Communication Association, September 8–12, San Francisco, California, USA, Proceedings*, 2016, pp. 100–104.