

Prof. Dr. Alexander Hillert

Textual Analysis

Chapter 1: Introduction

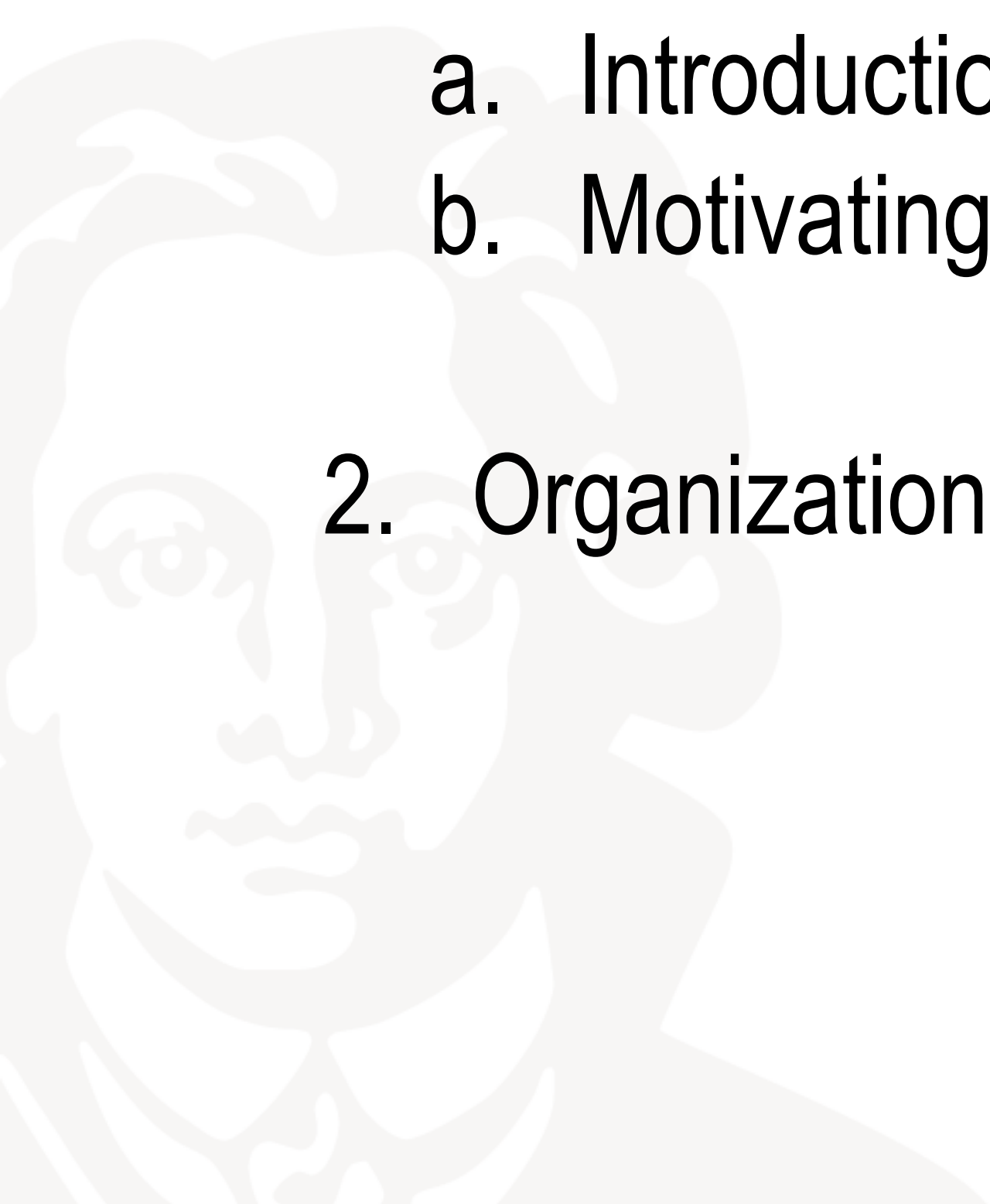
July 18, 20, and 22, 2022

WHU – Otto Beisheim School of Management

This chapter's agenda:

1. Introduction and motivation
 - a. Introduction of the lecturer
 - b. Motivating example for textual analysis in finance/economics

2. Organizational issues



Personal information - Prof. Dr. Alexander Hillert

- 2021/12 – today: Professor for Finance and Data Science at Goethe University and Leibniz Institute for Financial Research SAFE
- 2016/12 – 2021/11: House of Finance-Professor of Sustainable Asset Management at Goethe University
- 2016/08 – 2016/11: assistant professor at the Research Center SAFE
- 2015 – 2016: postdoc at the University of Mannheim
- 2009 – 2015: Ph.D. student at the CDSB of the University of Mannheim
- 2005 – 2009: studies in business economics at the University of Mannheim and HEC Paris

Research interests:

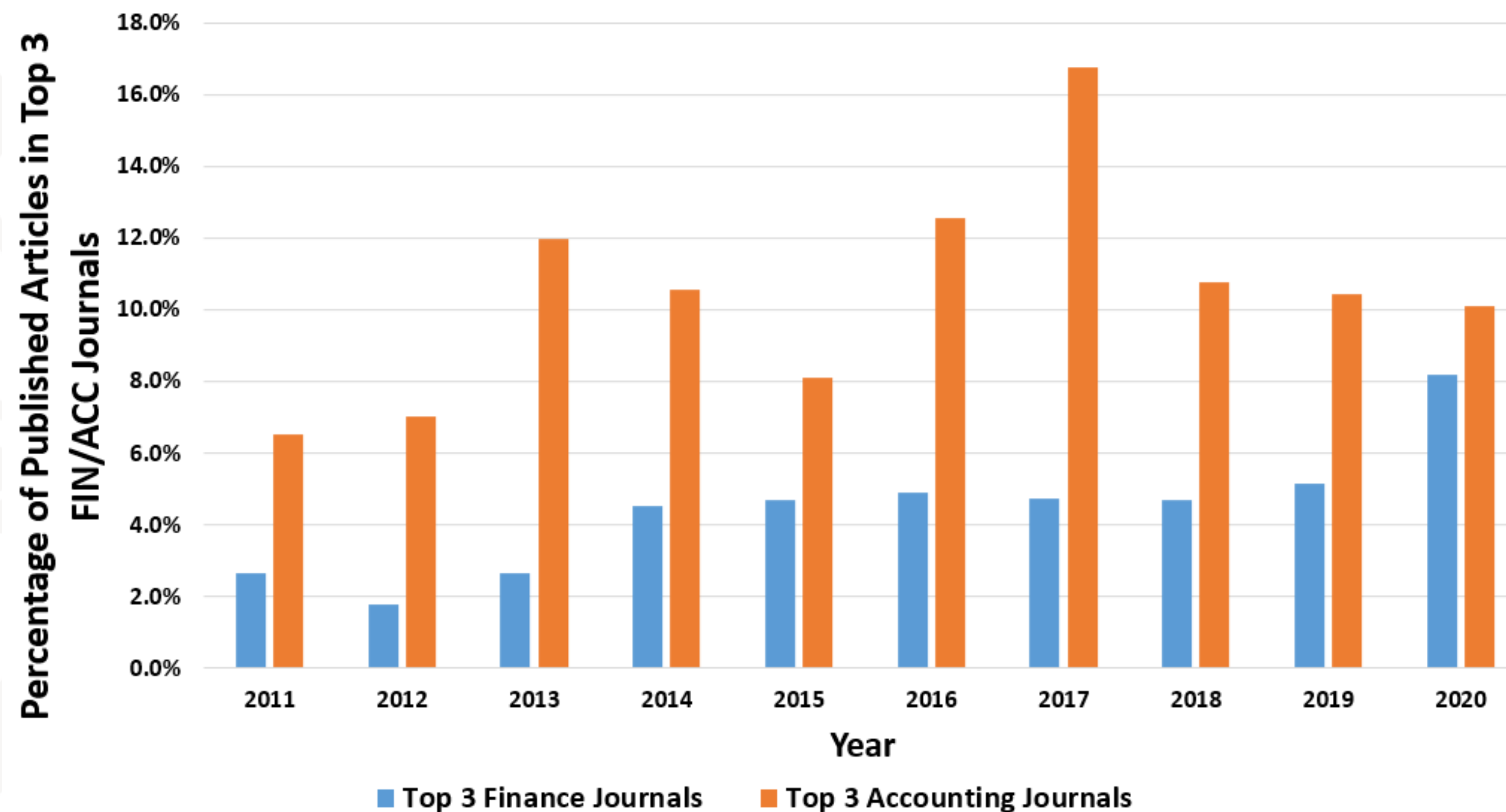
- Behavioral finance, textual analysis, empirical asset pricing, empirical corporate finance.
- Overarching question: how do market participants collect and process information?

Selected papers using textual analysis

- Media Makes Momentum, 2014, *Review of Financial Studies* 27, pp. 3467-3501.
We use more than 2 million firm-specific newspapers articles to analyze how newspaper coverage and article tone are related to the momentum effect (one of the most prominent return anomalies).
- Stock Repurchases and Liquidity, 2016, *Journal of Financial Economics* 119, pp. 186-209.
Not textual analysis per se, but similar methods:
We extract data on actual share repurchases from the firms' 10-Q and 10-K filings (quarterly and annual financial reports) to analyze the two-way causal relation between share repurchases and stock liquidity.

Unstructured data become increasingly important for top research

- % of articles published in top 3 finance and accounting journals using some data from the Securities and Exchange Commission's EDGAR database.



293 SEC EDGAR-based publications since 2011
ACC: 158
FIN: 135

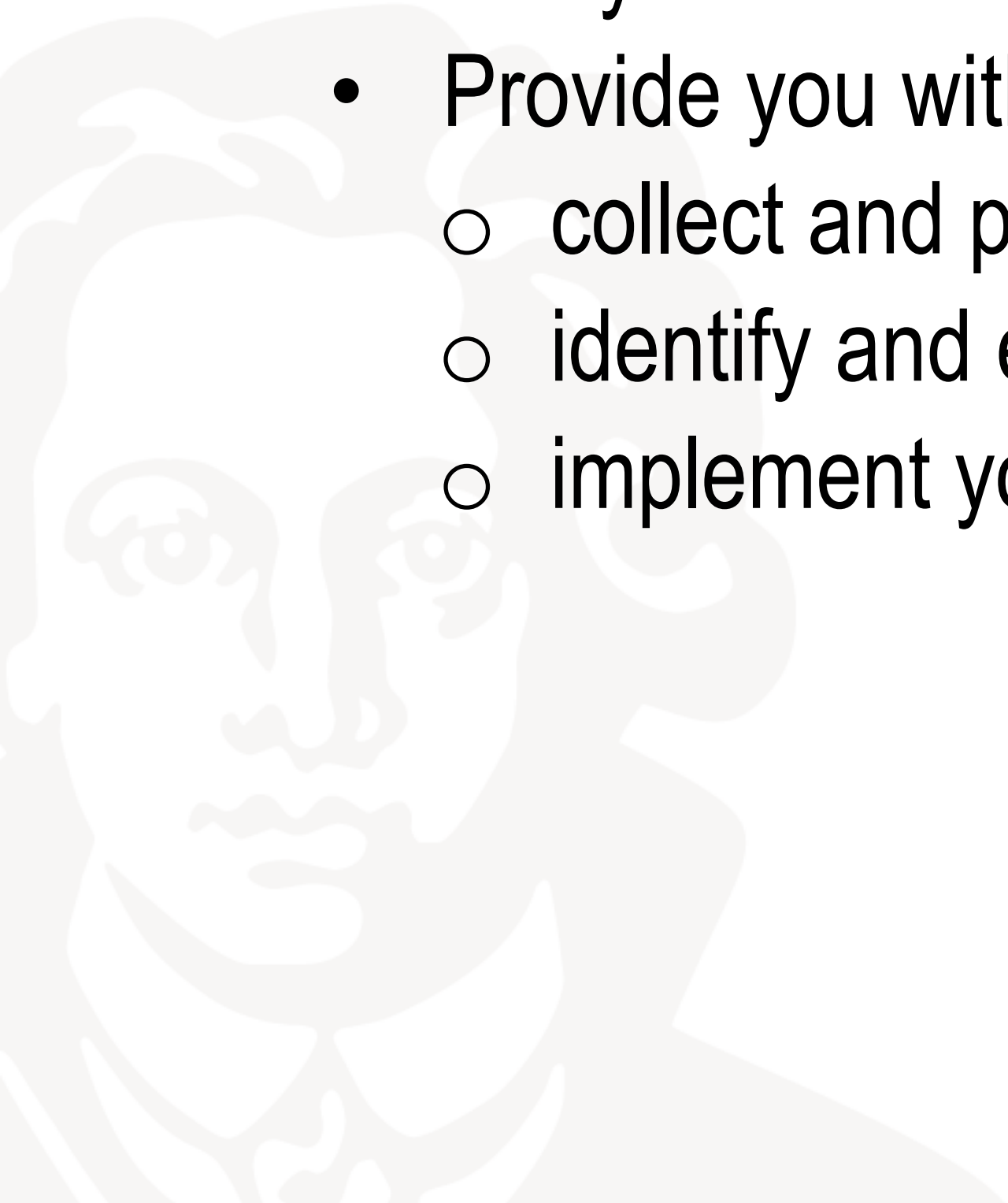
Source: own computations

→ Textual analysis is a rapidly growing area/tool in accounting and finance.

Introduction (4)

My motivation and goals for this course

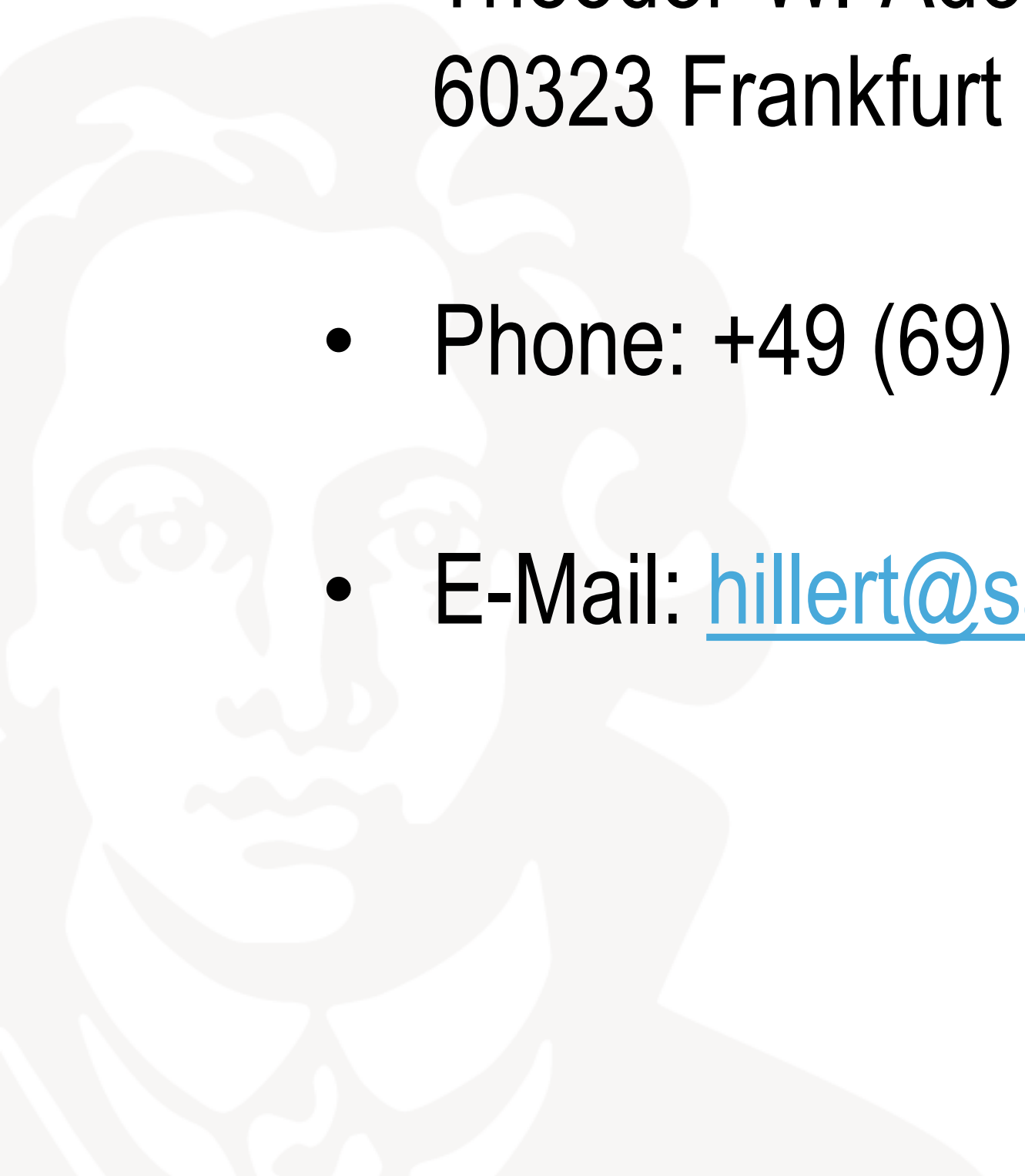
- Textual analysis allows to quantify information/mood/thoughts that could not be analyzed before.
- Get you excited about textual analysis.
- Provide you with the knowledge and skills to:
 - collect and process large (textual) datasets,
 - identify and extract information from texts,
 - implement your own textual analysis project.



Contact details

Contact details - Prof. Dr. Alexander Hillert

- Address
Leibniz Institute for Financial Research SAFE
Theodor-W.-Adorno-Platz 3
60323 Frankfurt am Main
- Phone: +49 (69) 798-33714
- E-Mail: hillert@safe-frankfurt.de



„Mutual Fund Shareholder Letters: Flows, Performance, and Managerial Behavior“ by Alexander Hillert, Alexandra Niessen-Ruenzi, and Stefan Ruenzi

- Motivation of the paper
 - In the US, fund reports contain shareholder letters which may cover different topics like: the fund's performance, the fund's portfolio holdings, discussion of economic and market conditions, etc.
 - These letters reach a large number of investors and are quoted in the business press.
 - The writing style of these letters varies a lot from technical descriptions to poetic statements such as:

“Short-term pain can lead to long-term gains”

→ **Does the writing style of letters have an impact on capital flows?**

Disciplined Small-Cap Opportunity Fund

(Dec. 2008; 6 month return -33.22%)

“Economies across the globe were buffeted by the **severe credit contraction** that **destabilized financial markets and led to bank closures, failures of financial services companies**, and massive government bailouts. [...]

The Dow Jones Industrial Average was down 31.9 percent in 2008, its **worst year since 1931**. The Standard & Poor's 500 index **dropped 22 percent in the fourth quarter alone**, and 37 percent for the full year - its **worst performance since 1937**. The **NASDAQ market had its worst year ever.**”

Negativity (Loughran/McDonald word list): 4.33%

American Century Quantitative Equity Fund

(Dec. 2008; 6 month return -33.84%)

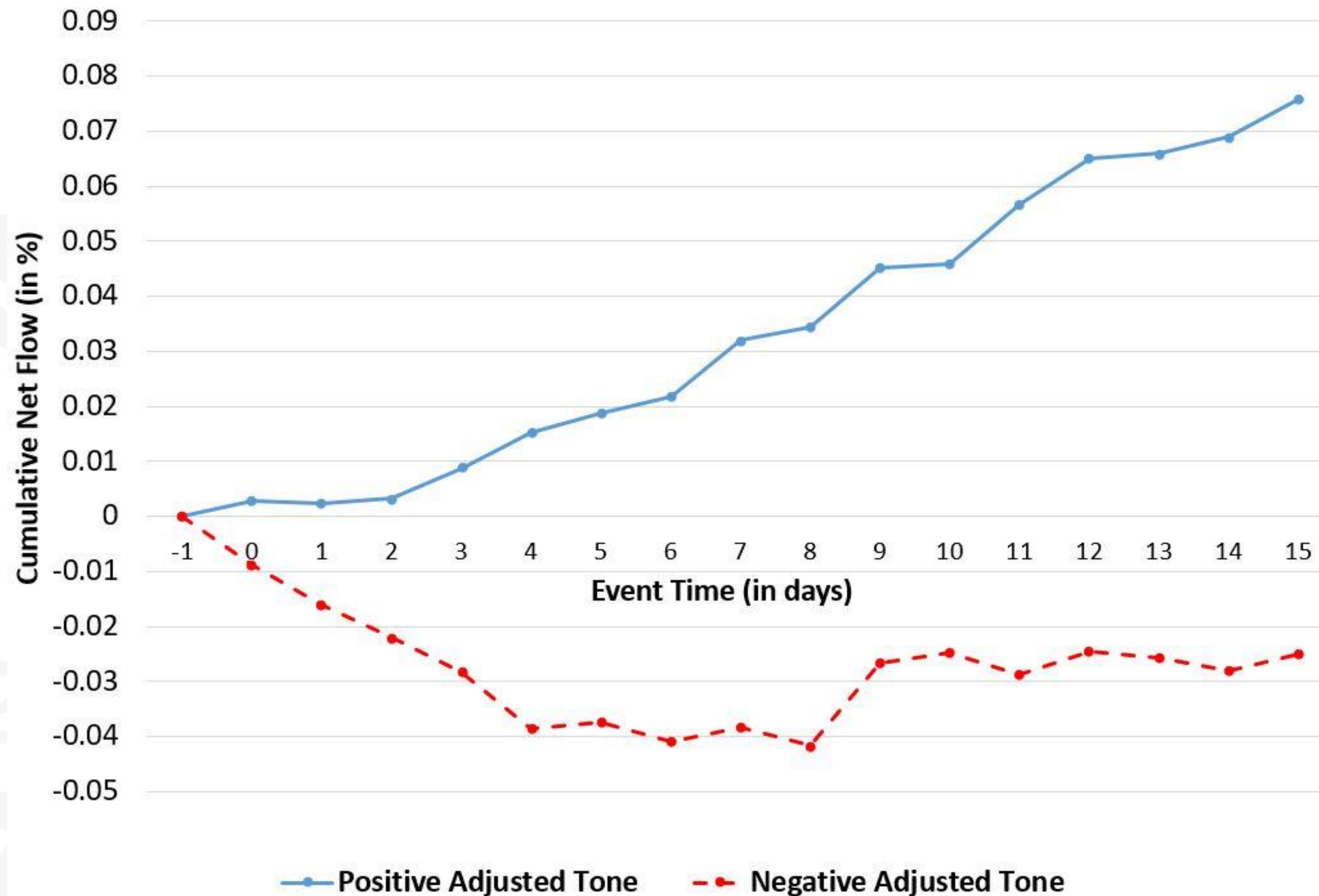
“As a company, American Century Investments® is **well positioned to deal with market turmoil**.[...] In addition, our actively managed, team-based approach allows our portfolio teams to **identify attractive investment opportunities regardless of market conditions**.

Our seasoned investment professionals have **substantial experience** and have **successfully navigated previous market crises.**”

Negativity (Loughran/McDonald word list): 2.78%

Introductory example (3)

Cumulative net flows for up to 15 days after the filing date of the shareholder letter



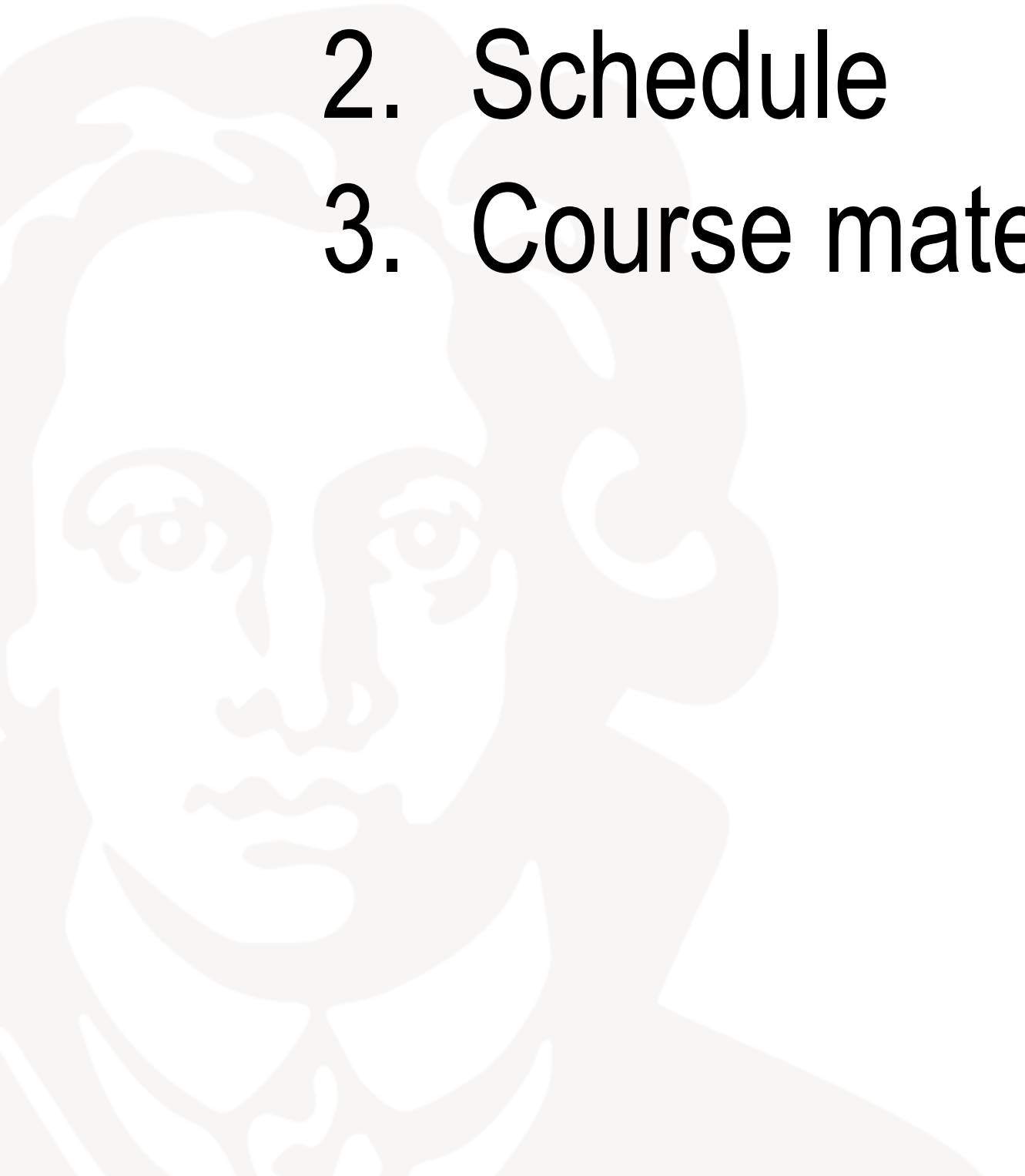
- “Adjusted tone” = tone after controlling for the fund’s actual performance and other fund characteristics.
- Results:
 - Fund investors react quickly to shareholder letters.
 - A more negative tone predicts lower net flows.

Organizational Issues

Part 2

Organizational issues

1. Roadmap of the course
2. Schedule
3. Course materials



Roadmap of the course

We will switch between lecture style and programming. After discussing the main contributions, we will start with a first programming task.

Our roadmap for this course

1. Literature review

- Main contribution on textual analysis in accounting, finance, and economics
 - Naïve Bayes approach: Antweiler and Frank (2004)
 - Introduction to machine learning
 - Dictionary approach: Tetlock (2007) and Loughran and McDonald (2011)
- Recommendations for your textual analysis projects
- Selected topics and papers
 - Readability: Loughran and McDonald (2014, 2020)
 - Textual similarity: Tetlock (2011) and Cohen et al. (2020)

2. Databases

3. Programming in Python: Implementation of textual analysis methods

- data collection
- data editing / regular expressions
- tone measurement
- dimensions other than tone, e.g., readability, document similarity
- introduction to machine learning

Three-day workshop

Date and Place	Time	Topic	
Monday, July 20	10:00 to 13:00 (incl. one break)	Introduction Literature review	
	13:00 to 14:00	Lunch break	Homework for Wednesday: work through the introduction to regular expressions.
	14:00 to 18:00 (incl. two breaks)	Literature review (continued) Programming: introduction to Python	
Wednesday, July 22	10:00 to 13:00 (incl. one break)	Text databases Programming: data collection	
	13:00 to 14:00	Lunch break	
	14:00 to 18:00 (incl. two breaks)	Programming: (1) data editing + application of regular expressions and (2) measuring tone	Homework for Friday: work through the datatype introduction (esp. the part on Counters) and the NLTK introduction.
Friday, July 24	10:00 to 13:00 (incl. one break)	Programming: (1) language complexity/readability and (2) stemming	
	13:00 to 14:00	Lunch break	
	14:00 to 18:00 (incl. two breaks)	Programming: (1) textual similarity and (2) Naïve Bayes/machine learning Q&A session	

Grading (1)

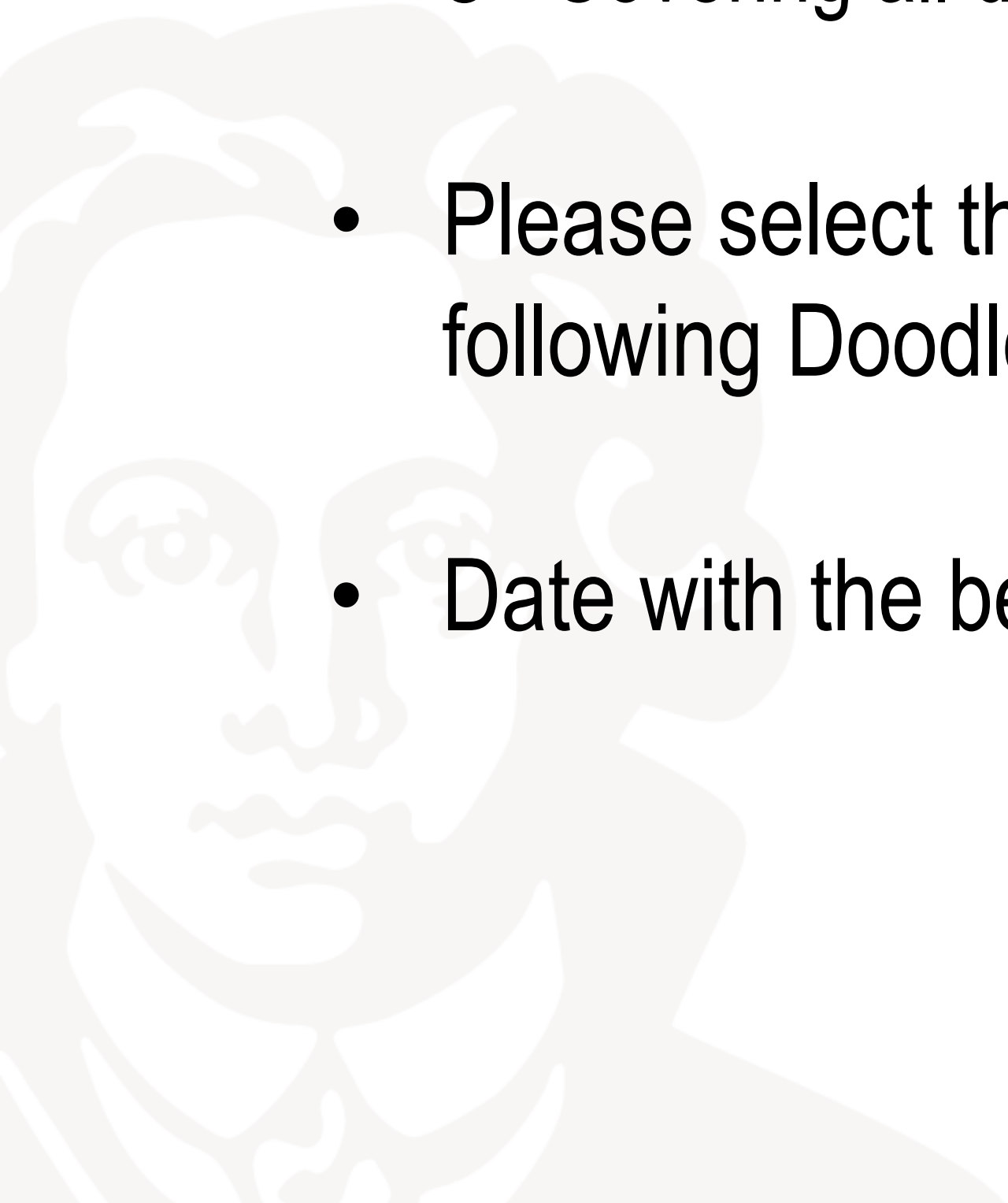
Grading

- Grade is determined by final exam.
- 72-hour take home exam.
The exam is a 48-hour take home exam but to give external doctoral students more time, there is an extra 24 hours.
- Exam consists of two parts:
 - Discussion of textual analysis research (about 1/3).
The questions are not area-specific but are about the dos and don'ts in textual analysis.
 - Programming problems in Python (about 2/3).
 - Similar to the ones we discuss in class. Ranging from data editing to textual similarity.
You will get programming templates similar to the ones that we use in class → you do not need to write a program from scratch but only complete the template.
 - Program codes from class can be used in the exam.

Grading (2)

Exam date

- 7 possible exam dates:
 - Between July 26 and August 4.
 - Covering all days of the week.
- Please select the dates that work for you by Wednesday, July 22, at 10:00 am using the following Doodle: <https://doodle.com/meeting/participate/id/dRgypALe>
- Date with the best availability will be selected.



Course materials - Overview

Course materials


1. Lecture slides
2. Programming problems + solutions
3. Introductions to specific programming topics
4. Papers
5. Python / Anaconda

All course materials (except for papers) will be available on Moodle.

 **Open all**  **Close all**

Instructions: Clicking on the section name will show / hide the section.

How to install Python

 Instructions on how to download Python


Please follow the instructions in the PDF to download your Python software.

Slides

Please find attached the slides for Chapter 2 "Literature review and methodology".

 Chapter 2 - Literature & Methodology

 Chapter 3 - Databases

 Chapter 4 - Programming

Data

 Data

Programming

 Templates

 Introductions

 Solutions

Recommended Surveys

Surveys on textual analysis

- [Loughran, T., and B. McDonald, 2020](#). Textual Analysis in Finance. Annual Review of Financial Economics 12, 357-375.
focuses on measuring complexity/readability
- [Gentzkow, M., Kelly, B., and M. Taddy, 2019](#). Text as data. *Journal of Economic Literature* 57, 535-74.
focusses on machine learning methods
- [Loughran, T., and B. McDonald, 2016](#). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54, 1187-1230.
mostly about tone measurement; includes recommendations for implementation.
- [Tetlock, P., 2014](#). Information transmission in finance. Annual Review of Financial Economics 6, 365-384.
mainly about (1) what research has already been done in finance and (2) what research questions should be addressed in the future.

Do you have questions on organizational issues?



Introduction of course participants

Please briefly introduce yourself

- Research field and research interests
- Programming knowledge (which programming languages)
- Expectations

