

# Accounting 7

## Data Analytics and Financial Narratives

### Spring 2022

Wanli Zhao  
wanli.zhao@unibocconi.it

Christoph J. Sextroh  
c.j.sextroh@tilburguniversity.edu

### Course Description

This course introduces students to recent advances in textual analysis and big data analytics in accounting and finance. We will frame the discussion around fundamental questions related to the dynamics behind financial communication and their implications for financial markets. The goal is to give students both a solid understanding about some of the core questions in financial communication as well as how to measure fundamental concepts from textual or otherwise unstructured data. After the course, students should be able to develop new research questions in the area of corporate disclosure and financial communication as well as apply different methodologies to capture information from financial narratives.

### Course Contents and Format

There will be six sessions, each organized around a specific topic or type of financial narrative:

	Date	Topic	Instructor
1	March 30	The Complexity of Corporate Filings	Sextroh
2	April 1	Language Markers of Spoken Narratives	Sextroh
3	April 4	Extracting the Content of Financial Narratives	Sextroh
4	April 8	Measuring Constructs from Textual Similarity	Zhao
5	April 14	Using Machine Learning to Detect Misreporting	Zhao
6	April 29	Measuring Constructs from Specific Textual Features	Zhao

Based on seminal as well as recent journal articles, we will discuss different research questions as well as different methodologies to explore these questions. Each session will consist of paper presentations by students and discussions of these papers as well as applied introductions into measuring specific concepts using Python.

### Preparation/Material

Each session includes several mandatory readings. **Absolutely make sure to do the readings before class!** Being prepared for class is important for active participation and discussion. Please check the detailed reading list below for more information.

## Attendance Policy

Class participation is necessary to understand the presented methods and to use them properly. Participants are required to attend at least 75% of the course hours to receive a grade. All assignments are mandatory.

## Assessment

Grades for the course will be determined on the basis of the following components:

- Class participation (20 points)
- Paper presentations (20 points)
- Homework assignments (20 points)
- Research proposal (40 points)

The total score will be rescaled to the following scheme to determine the grade:

Grade	Points
F	< 18
C	18 - 19
C+	20 - 21
B-	22 - 23
B	24 - 25
B+	26 - 27
A-	28 - 29
A	30
A+	> 30

## Class Participation

Participants are expected to read and prepare the assigned material before class and to contribute actively to class discussions. To prepare for the sessions, read the papers and think about the following questions:

- (1) What is the basic research question? Why is the research question important?
- (2) What are the most relevant prior or contemporaneous papers to the study?
- (3) What is the conceptual background of the research question?
- (4) What are the research design and empirical methods used to test the hypotheses?
- (5) What are the results and key takeaways?
- (6) What are limitations and potential extensions of the study?

## Paper Presentations

Each student will be assigned multiple paper presentations throughout the course. Presentations should be concise but make sure to incorporate on the six questions mentioned above. Given the purpose of this course, in your discussion of (4) above, make sure to touch upon the following aspects:

- Why is the method appropriate for the research question?
- What are the advantages of the method that help to get new insights on the issue?
- How has the method been implemented?
- What are potential difficulties, challenges, or problems in the application of the method?

There is no restriction to the number of slides. However, please make sure to limit your presentations to 20-25 minutes.

## Homework Assignments

There will be two homework assignments (counting each 10% towards the final grade). These homework assignments require students to implement and apply specific concepts and methods discussed in class using Python. More details will be announced in due time.

## Research Proposal

Coming up with interesting and researchable questions is a crucial skill for a career in academia. Although we focus on specific methods and concepts, students should also digest the ideas of the literature and should be able to critically think about new ideas based on the literature, methods and tools discussed in class. This assignment will practice this important skill of developing interesting and relevant research questions and thus will also provide valuable legwork for students' own PhD thesis.

The research idea should focus on a question that requires the analysis of financial narratives (e.g., corporate filings, regulatory documentations, conference calls, social media, message boards, etc.). The proposal should clearly describe and motivate the research question and its incremental contribution. In addition, the proposal should outline a suggested research design and discuss the choice of analysis methods to tackle the research question. Also, the proposal should be transparent about potential methodological problems and critically discuss limitations of inference.

Some additional remarks/suggestions regarding the research proposal:

### (1) Selecting a topic and finding a research question

- Interesting questions not only follow from reading the academic literature, but often emerge from following recent developments in the field (e.g., by reading the financial press)
- Do a careful search on prior studies (including existing working papers) to ensure that your topic is original. Clearly explain your intended incremental contribution.
- Be as specific as you can be in outlining and motivating your research question

### (2) Literature review

- Carefully read prior literature following the questions outlined above.
- A literature review is not just an overview/summary of prior literature. Instead use the literature...
  - ...to give an overview about relevant (!) prior findings
  - ...to clearly outline your contribution
  - ...to develop a clear a concise argument towards your hypothesis

### (3) Research design and methodology

- How will you address the question empirically? What methodology and specifications will you use?
- How do you measure each variable? Where would you obtain the data?
- What are the potential challenges in your proposal (e.g., data availability, programming/processing time, endogeneity). And how do you plan to deal with it? Be specific.

There is no set page limit, although a proposal that goes beyond ten pages (single spaced, font size 12, times new roman) is probably too long. It is up to you to organize the proposal and allocate space to different sections (but it is hard to imagine that motivating the research

question would take 8 pages). Try to be specific and concise.

**Due date for the written research proposal:**

Please email your research proposal to both instructors by 11:59 PM of May 29, 2022.

## Required Software

Besides conceptual discussions around research topics and measurement concepts, students will also be introduced to approached on how to implement the measurement of different concepts in practice. We will use the Python in class and also for homework assignments. If you did not yet follow a basic course in programming with Python, please contact us immediately. We can then recommend specific online tutorials as an introduction.

We will use Anaconda, a scientific Python distribution that is very convenient to install and to operate. It automates installing Python for scientific work and gives you simple tools to keep everything up to date (which is not necessarily simple if you would do that manually, especially if you are not on Linux). It makes working with Python very easy, even for people without a programming background.

If you have not done so, please install the latest Anaconda distribution before the first class (<https://www.continuum.io/downloads>). Make sure you install the latest Anaconda Python 3.x distribution and not the old 2.7 distribution.

Once you have Anaconda installed, open your start menu, look for the anaconda group and open the anaconda command prompt in administrator mode (If you are on a MAC, find any type of terminal. It should be tucked away in your "Utilities" folder). Once you have a terminal open, type: "*conda update conda*" and once you updated conda type "*conda update --all*". Conda is a small program that keeps all the different parts and modules and packages in Python up to date. You want to make sure you updated everything to the latest version before coming to the first class. Finally install the spaCy and gensim packages using the commands

```
conda install spacy  
conda install gensim
```

Afterwards, you have to run

```
python -m spacy download en
```

To download, the necessary word models that spaCy uses. Please refer to the spaCy documentation for additional installation help.

# Detailed Reading List

## Session 1

### The Complexity of Corporate Filings

In this session we will introduce the perspective of “text as data”. From the perspective of financial communication as a signalling game for expectation management, we will discuss about qualitative information in regulatory filings with emphasis on the “complexity”, “readability”, or “vagueness” of narrative information.

#### Assigned readings:

- Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 221-247.
- Lawrence, A., 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56, 130-147.
- Loughran, T., and B. McDonald, 2014. Measuring Readability in Financial Disclosures, *Journal of Finance* LXIX, 1643-1671.
- Guay, W., D. Samuels, D. Taylor. 2016. Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62, 234-269.
- Bonsall, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp, 2017. A plain English measure of financial reporting readability. *Journal of Accounting and Economics* 63(2), 329-357.

#### Additional readings:

- Dyer, T., M. Lang, and L. Stice-Lawrence, 2016. Do managers really guide through the fog? On the challenges in assessing the causes of voluntary disclosure. *Journal of Accounting and Economics*, 62, 270–276.
- De Franco, G., O.-K. Hope, D. Vyas, and Y. Zhou, 2015. Analyst Report Readability. *Contemporary Accounting Research* 32, 76-104.

#### Background readings:

Realizing that text is data also means that we have to deal with unique complexities in processing this often complex and unstructured data. Gentzkow, Kelly, and Taddy (2019) provide an excellent discussion of viewing text as data for economic research:

- Gentzkow, M., B. T. Kelly, and M. Taddy, 2019. Text as data. *Journal of Economic Literature* 57(3), 535-574.

## Session 2

### Language Markers of Spoken Narratives

In this session, we will discuss about language markers in spoken narratives and their usefulness as signals about the economic fundamentals of the firm. We will also introduce word dictionaries as a way to measure certain constructs from text.

#### Assigned readings:

- Davis, A. K., J. M. Piger, and L. M. Sedor (2012): “Beyond the numbers: Measuring the information content of earnings press release language,” *Contemporary Accounting Research*, 29, 845–868.

- Loughran, Tim, and Bill McDonald, 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* LXVI, 35-65.
- Larcker, D. F., and A. A. Zakolyukina, 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50, 495-540.
- Bochkay, K., J. Hales., and S. Chava, 2020. Hyperpole or Reality? Investor Response to Extreme Language in Earnings Conference Calls. *The Accounting Review* 95, 31-60.

Additional readings:

- Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008. More than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance* LXIII, 1437-1467.
- Henry, E., and A. J. Leone, 2016. Measuring Quality Information in Capital Markets Research: Comparison of Alternative Methodologies to Measure Disclosure Tone. *The Accounting Review* 91, 153-178.
- Sprenger, T. O., P. G. Sandner, A. Tumasjan, and I. M. Welp, 2014. News or noise? Using Twitter to identify and understand company-specific news flow. *Journal of Business Finance & Accounting*, 41, 791–830.

### **Session 3**

#### **Extracting the Content of Financial Narratives**

This session focusses on the unique information content of financial narratives. Methodology-wise, we will discuss about anchor phrases, parsing tools, and topic modelling as method to identify and extract specific topics in financial narratives.

Assigned readings:

- Hassan, T.A., S. Hollander, L. van Lent, and A. Tahoun, 2019. Firm-Level Political Risk: Measurement and Effects. *The Quarterly Journal of Economics* 134(4), 2135-2202.
- Bozanic, Z., D. T. Roulstone, and A. Van Buskirk, 2018. Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics* 65, 1–20.
- Li, F., 2010. The information content of forward-looking statements in corporate filings - A naïve Bayesian machine learning approach. *Journal of Accounting Research* 48, 1049–1102.
- Davis, A. K., J. M. Piger, and L. M. Sedor, 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64, 221–245.
- Huang, A. H., R. Lehavy, A. Y. Zang, and R. Zheng, 2017. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science* 64, 2833–2855.

Additional readings:

- Adams, R. B., A. C. Akyol, and P. Verwijmeren, 2018. Director skill sets. *Journal of Financial Economics* 130, 641–662.
- Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang, 2020. Firm-level climate change exposure. Working paper.
- Hassan, T., S. Hollander, L. van Lent, M. Schwedeler, and A. Tahoun, 2021. Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1. Working paper.

## Session 4

### Measuring Constructs from Textual Similarity

In this session, we will discuss about changes in and similarity of financial narratives as signals about firms' underlying economic activities.

#### Assigned readings:

- Brown, S. V. and J. Tucker. 2011. Large-Sample Evidence on Firms' Year-Over-Year MD&A Modifications. *Journal of Accounting Research* 49 (2), 309-346.
- Hoberg, G., and G. Phillips, 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124, 1423-1465.
- Hoberg, Gerard, and Craig Lewis, 2017. Do fraudulent firms produce abnormal disclosure? *Journal of Corporate Finance* 43, 58-85.
- Lang, Mark, and Lorien Stice-Lawrence, 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60, 110-135.
- Cohen, L., C. Malloy, and Q. Nguyen, 2020. Lazy prices. *The Journal of Finance* 75, 1371-1415.

#### Additional readings:

- Hoberg, G., and G. Phillips, 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies* 23(10), 3773-3811.

## Session 5

### Using Machine Learning to Detect Misreporting

In this session, we will discuss about detecting fraudulent behaviour and financial misreporting with an emphasis on machine learning techniques. Background readings provide additional information on the use of machine learning in economic research.

#### Assigned readings:

- Brown, N. C., R. M. Crowley, and W. B. Elliott, 2020. What Are You Saying? Using topic to Detect Financial Misreporting. *Journal of Accounting Research* 58, 237-291.
- Perols, J. L., R. M. Bowen, C. Zimmermann, 2017. Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction. *The Accounting Review* 92, 221-245.
- Yang B., B. Ke, B. Li, Y. J. Yu and J. Zhang, 2020. Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research* 58, 199-235.
- Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan, 2021. Using machine learning to detect misstatements. *Review of Accounting Studies* 26, 468-519.

#### Background readings:

The article by Varian provides an accessible introduction to key concepts in Machine Learning, while the Mullainathan and Spiess (2017) paper provides a deep dive into the difference between estimation and prediction. Belloni, Chernozhukov, and Hansen (2014) provides a good introduction to regularized regression and how it can be used for causal inference. Athey (2018) is a good summary of much of the material that will be presented throughout the session.

- Athey, S., 2018. The Impact of Machine Learning on Economics.

- Varian, H. R., 2014. Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 28(2): 3-2.
- Belloni, A., V. Chernozhukov, and C. Hansen, 2014. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29-50.
- Mullainathan, S., and J. Spiess, 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2): 87-106.

## **Session 6**

### **Measuring Constructs from Specific Textual Features**

In this session, we will discuss about using specific textual features to measure corporate behaviour.

#### Assigned readings:

- Allee, K. D., and M. D. DeAngelis, 2015. The Structure of Voluntary Disclosure Narratives: Evidence from Tone Dispersion. *Journal of Accounting Research* 53, 241-274.
- Lee, J. 2016. Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review* 91, 229-250.
- Allee, K. D., M. D. DeAngelis, and J. R. Moon Jr., 2018. Disclosure "Scriptability". *Journal of Accounting Research*, 56, 363-430.
- Gow, I. D., D. F. Larcker, and A. A. Zakolyukina, 2021. Non-answers During Conference Calls. *Journal of Accounting Research* 59, 1349-1384.

#### Additional readings:

- Andreicovici, I., L. van Lent, V. Nikolaev, and R. Zhang, 2020. Accounting measurement intensity. Working paper.