Syllabus

Innovative Data Sources in Economic Research

M.WIWI-VWL.XXXX

Summer semester 2025

(as of March 10, 2025)

Dr. Carlo Birkholz

Phone: +49 (0)551 39-28314

Email: cbirkhlz@gmail.com

Website: https://carlo-birkholz.github.io/

Room: Oeconomicum, Room 2.206

Office hour: By appointment

1. General information

1.1 Course content

In recent years, economic research has undergone a profound transformation driven by the rapid expansion of innovative data sources and the technical capabilities to harness them. Advances in satellite technology, web infrastructure, and machine learning have made it possible to collect and analyze vast amounts of data that were previously inaccessible to economists. These new data sources allow us to measure economic activity at unprecedented spatial and temporal granularity, filling critical gaps left by traditional data sources. Satellite imagery, for instance, provides insights into economic development, infrastructure, and environmental change even in data-scarce regions. Online data, whether scraped from websites or accessed through APIs, can capture real-time economic behavior such as consumer sentiment, firm dynamics, or price changes across borders. Text data — from scanned documents to social media posts — can be processed using modern natural language processing tools, including large language models, to extract economic signals from unstructured information. This seminar explores how these innovative data sources are transforming economic research, covering both the technical methods required to process them and their applications to real-world economic questions.

1.2 Course goals

Students learn what different types of innovative data sources exist, how they are being applied in frontier economic research currently, and how to make them accessible for their own analysis. The students then learn how to implement a basic data collection and analysis utilizing data from an innovative source discussed during the seminar, and how to present their work to an academic audience.

1.3 Prerequisites

Basic command of R and Python are recommended, as coding will be conducted in these languages. As a refresher, the Chair offers a self-study module for R.

1.4 Credit points

6 ECTS-LP.

1.5 Registration

Please register via FlexNow by Wednesday, April 23, 2025. Attendance at the introductory meeting is compulsory for registration. The number of participants is restricted to 12.

2. Course overview

2.1 Description of the teaching and learning methods

Students must demonstrate an overview of innovative data sources and their application in recent economic research. By accessing and analyzing data from an innovative source, students demonstrate their ability to apply the programming concepts of the seminar and econometric methods using state-of-the-art statistical software packages, as well as their ability to identify which areas of research might benefit from the use of innovative data sources, which they critically reflect on in a term paper. By presenting the term paper, they show their ability to concisely present complex empirical concepts. Moreover, during oral discussions, students demonstrate their ability to defend, as well as critically reflect upon arguments from the empirical literature. **The course language is English.**

2.2 Meetings

Introductory meetings:

- Course outline - Wednesday, April 16, 2025, 14:00–15:00 s.t. (seminar room: see eCampus/EXA)
- Satellite data - Thursday, May 8, 2025, 14:00–18:30 s.t. (computer room: see eCampus/EXA)
- Internet data - Thursday, May 15, 2025, 14:00–18:30 s.t. (computer room: see eCampus/EXA)
- Text as data - Thursday, May 22, 2025, 14:00–18:30 s.t. (computer room: see eCampus/EXA)

Submission deadline for exposé (Friday, May 30, 2025, 23:55)

Please upload on Stud.IP an electronic version of your exposé (in PDF format).

Mid-term meetings: Thursday, June 5 & June 19, 2025, 14:00–18:00 s.t. (computer room: see eCampus/EXA)

- Coding Lab and Q&A

Final meeting: Thursday, July 10, 2025, 13:00–18:30 s.t. (seminar room: see eCampus/EXA)

- Presentation of term paper

Submission deadline for term paper (Friday, July 18, 2025, 23:55)

Please upload on Stud.IP an electronic version of your term paper (in PDF format) as well as replication files.

Communication:

Instead of emails, we will use the two communication fora on Stud.IP (one for content-related and one for organizational questions).

2.3 Examination and grading of the module

The grading consists of three components:

- Term paper [70%]
- Presentation of term paper [30%]
- Participation in class [at the margin]

Furthermore, in order to pass the course, participants must

- be present and actively participate during all seminar sessions

- submit all written course requirements on time
- achieve a weighted average grade exceeding or equal to 4.0

2.4 Course materials

- The preliminary list of required readings is announced below.
- Supplementary course materials will be announced at the beginning of each meeting and/or published on Stud.IP.

3. Term paper requirements

Exposé

The exposé should be no longer than a single page with font size 12 and line spacing 1.5 times, alignment: justification, margin (right, left, top and bottom margin): 2.5 cm each. It should briefly express the data source you want to access, the type of data you look to collect from it, how you will achieve the data collection and what type of economic questions could be answered using this data.

Term paper

In your term paper you describe your data collection project in detail, and provide some descriptive statistics or other summarizing analysis of the data you have collected. You can describe challenges you have encountered during your data collection project, and how you have or could, in a larger scale project, address them. Discuss applications of your data in research and policy making, and what the upsides and downsides are compared to alternative data sources.

The length of the term paper must not exceed 15 pages. The cover sheet, reference list, figures, and tables (if any) are excluded from the page count. Please use the following formatting: Font size: 12 pt, line spacing: 1.5 times, alignment: justification, margin (right, left, top and bottom margin): 2.5 cm each. Please make sure that you are consistent in your referencing and bibliography.

Use a referencing style that is in line with basic conventions in the economics literature. Always quote within a work using the same method. The Chair recommends the in-text author-year citation style (see, for example, information sheet "Citation according to the Harvard system"). All sources that you have cited in the text must appear in the reference list. Everything that you have only read but not cited in the text must NOT appear in the reference list. Term papers need to be written in English. The work must be done independently by the student.

Every idea, statement or fact taken from other sources must be appropriately cited. Quotes need to be indicated as such. Plagiarism is sanctioned with the grade 5 (failed). The general rules of the Faculty of Business and Economics apply. A signed declaration must be attached to each scientific work: "I hereby declare that I wrote this term paper independently, without assistance from external parties, and without use of other resources than those indicated. All information taken from other publications

or sources in text or in meaning are duly acknowledged in the text. I give my consent to have this thesis checked by plagiarism software."

Additionally, please attach a **declaration** if you have used AI-based chatbots such as ChatGPT. Sample annex:

Declaration on the use of ChatGPT and comparable tools: In this paper, I have used ChatGPT or another AI tool as follows.:

[ ] not at all

[ ] during brainstorming

[ ] when creating the outline

[ ] to write individual passages, altogether to the extent of ...% of the entire text

[ ] for the development of source code

[ ] for optimizing or restructuring source code

[ ] for proofreading or optimizing

[ ] further, namely: …

I hereby declare that I have stated all uses completely. Missing or incorrect information will be considered as an attempt to cheat.

Replication files

Please upload all files necessary for replicating your empirical analysis to Stud.IP together with your submission. Your empirical analysis is considered replicable if all results used in your paper (regression tables, figures, etc.) can be generated from the raw data using your code scripts.

Your submission should include the following files:

- all raw data
- the merged dataset
- code to clean and merge the raw data
- code to analyze the data
- a short instruction for the replication, e.g., a README file or Makefile

The raw data should be marked as such and stored in a separate folder. If the datasets are very large, feel free to use file transfer services such as WeTransfer or Dropbox links. The merged dataset should be saved in a common format (e.g., csv, dta, rds). Please make sure that your code can be executed outside your own computer without further adjustments to the code. For that reason, please use relative

file paths only. Add comments to specify which part of the code corresponds to which part of your paper.

Preliminary reading list:

INTRODUCTORY MEETINGS

<u>*Measurement*</u>

- *Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. Journal of Economic Perspectives, 30(4), 171-198.*
- *Lehnert, P., Niederberger, M., Backes-Gellner, U., & Bettinger, E. (2023). Proxying economic activity with daytime satellite imagery: Filling data gaps across time and space. PNAS nexus, 2(4).*
- *https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/*
- *https://www.worldpop.org/methods/population-estimation-for-sustainable-development/*
- *Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535-574.*

<u>*Applications*</u>

- *Hodler, R., & Raschky, P. A. (2014). Regional favoritism. The Quarterly Journal of Economics, 129(2), 995-1033.*
- *Evsyukova, Y., Rusche, F., & Mill, W. (2025). LinkedOut? A field experiment on discrimination in job network formation. The Quarterly Journal of Economics, 140(1), 283-334.*
- *Birkholz, C., & Gomtsyan, D. (2024). The Global Software Production Network. Link*
- *Macaulay, A., & Song, W. (2023, May). News media, inflation, and sentiment. In AEA Papers and Proceedings (Vol. 113, pp. 172-176). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.*

Note: Dates and contents subject to change!