Turning Text into Economic Measurement

From Lexicons to Large Language Models

Thomas Renault

Université Paris-Saclay

Text-as-Data in Economics

- Many key economics variables are not directly observable or measurable
- Beliefs, attention, emotions, narratives, policy views
- Text and digital traces allow near real-time measurement
- Computational tools help map unstructured data to economic objects
- Goal is not prediction alone, goal is measurement plus interpretation

Unifying view of my research

- **Core objects**
  - Text and digital traces as economic data
  - Platforms and AI systems as information intermediaries
  - Measurement of sentiment, emotions, attention, uncertainty, polarization, and policy preferences
- **Shared methodological backbone**
  - Construct new measures at scale
  - Validate with ground truth (when possible)
  - Embed measures into forecasting or causal / quasi-causal designs (when possible)

Data types

- Social media content: Twitter (X), StockTwits
- Media text: news articles
- Institutional text: central bank communication, political manifestos, earnings calls

- Access to text data has become more difficult than a few years ago: APIs have been restricted or closed (notably X), and web scraping is harder due to increased protections and the growing recognition that text data is valuable

Pipeline overview

- **Data acquisition**
  - APIs, scraping, archives
- **Cleaning and representation**
  - Tokenization, embeddings, de-duplication, language detection, bot detection
- **Measurement**
  - Sentiment, emotions, topics, veracity ...
  - Methods: lexicons, supervised learning, unsupervised learning, LLMs ...
- **Validation**
  - Human labels, external benchmarks, surveys
- **Inference**
  - Event studies, difference-in-differences, IV, panel models ...

Causality: from descriptive measurement to credible inference

- Computational data increases power and resolution
- But does not remove confounding
- Standard identification strategies remain essential
- Key is to align the design with a transparent measurement pipeline

Intraday investor sentiment from social media

- Does online investor sentiment predict intraday return patterns?
- **Data** : millions of messages published on StockTwits
- **Computational contribution**
  - Large set of investor messages tagged as bullish or bearish $\rightarrow$ compute word weights from differential usage in positive vs negative posts $\rightarrow$ create a finance-specific vocabulary $\rightarrow$ compare with machine learning methods
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. Journal of Banking & Finance, 84, 25-40.

Intraday investor sentiment from social media

- Generic sentiment tools often fail out-of-domain
- Building custom lexicons is working well when you have a large training dataset to learn from
- Large datasets are needed to train a machine learning classifier
- Simple ML algorithms tend to perform well on short text

Twitter-based economic uncertainty indices

- Can we measure economic uncertainty in real time from social media?
- **Data** : tweets containing the keyword "uncertainty"
- **Computational contribution**
  - Uncertainty-related content $\rightarrow$ filter to relevant economic context $\rightarrow$ aggregate frequency and intensity over time $\rightarrow$ compare with news-based indices and other benchmarks
- Altig, D et al. (2020). Economic uncertainty before and during the COVID-19 pandemic. Journal of public economics, 191, 104274.

Twitter-based economic uncertainty indices

- Simple approach often works
- Reproducibility is key
- User location has to be taken into account to build national level indices
- Social media indicators highly correlate with traditional media indicators

Emotions and Policy Views

- How do emotions shape policy views, and how has emotional content evolved?
- **Data** : tweets about policy (citizen and politicians), floor speeches, campaign speeches
- **Computational contribution**
  - Emotion and topic classification at scale with Large Language Models and machine learning
- Algan, Y. et al. (2025). Emotions and policy views. Working paper.

Emotions and Policy Views

- Use LLMs to identify the main emotion and the main topic at the sentence level, on a random sample of 200K sentences
- Validate the classification by comparison with human annotations from three annotators
- Use the resulting labeled sample as a training dataset to train a supervised machine learning classifier (with embeddings)
- Analyze the evolution of emotions over time across tweets, floor speeches, and various types of documents
- Account for composition effects using user fixed effects, and match X data with voter registers to ensure representativeness

Emotions and Policy Views

- Classification with LLMs works well
- Building a human validation dataset is difficult, with high disagreement across annotators at the sentence level
- Using LLMs on a random sample, then scaling with supervised machine learning, is an effective way to avoid labeling the full corpus, which can be costly
- Combining measurement with experiments supports causal interpretation

Offender origin disclosure and immigration attitudes

- Does systematically disclosing offender origin affect natives' attitudes toward immigration

- **Data** : news articles from Factiva

- **Computational contribution**
  - Identify the nationality of criminals in newspaper articles $\rightarrow$ link the nationality to the perpetrator or the victim $\rightarrow$ construct measures at the newspaper level and exploit the media change as a source of variation

- Keita, S. et al. (2024). The usual suspects: Offender origin, media reporting and natives' attitudes towards immigration. The Economic Journal, 134(657), 322-362.

Offender origin disclosure and immigration attitudes

- Simple methods often work well
- Textual analysis serves as an input to the analysis, while the key contribution comes from identification and a research design exploiting a unilateral change in reporting by one newspaper.
- If I were to rewrite the paper today: use LLMs as a robustness check or as a complementary approach.

AI and fact-checking

- How do users interact with LLMs on X for fact-checking, and how does LLM-based fact-checking answer ?

- **Data** : Tweets mentioning Grok and other LLMs, fact-checking requests, and corresponding responses over time

- **Computational contribution**
  - Topic modeling using BERTopic $\rightarrow$ veracity scoring using another LLM $\rightarrow$ compare different LLM answers and ask fact-checkers for ground truth

- Renault, T. et al. (2026) @Grok Is This True? LLM-Powered Fact-Checking on Social Media. Working Paper.

AI and fact-checking

- GPT 4o-mini with batches is cheap
- Prompting should be transparent
- LLM models themselves show some polarization
- Trust in LLM models affects the way they impact beliefs
- Computational measurement is central, but interpretation requires careful validation and design

Toolbox overview

- Starter kit (Python): pandas, nltk, scikit-learn
- Advanced kit (Python): transformers, sentence-transformers, openai, openrouter, torch
- Econometrics: use R for the inference stage (DiD, panel methods, causal designs). R can be embedded within Python using rpy2

Common failure modes

- **Measurement overfitting**: Risk of data dredging and LLM hacking
- **Validation missing**: No benchmark, no human labels, no external comparison
- **Causality claims without design**: Large N does not solve identification
- **Platform drift**: APIs and algorithms change, models update, user composition shifts

Thank you