Identifying Financial Crises with Machine Learning on Textual Data

Mary Chen [†], Matthew Deininger [*], Seung Jung Lee [*], and Martin Sicilian [‡]

[†] Bank for International Settlements
[*] Federal Reserve Board
[‡] Stanford University Law School

2021 RiskLab/BoF/ESRB Conference on Systemic Risk Analytics
October 19, 2020

Disclaimer: The views expressed herein are those of the discussant, and do not necessarily represent the views of the BIS or the Federal Reserve Board or its staff.

Artificial Intelligence

*"I believe there is no deep difference between what can be achieved by a biological brain and what can be achieved by a computer. It, therefore, follows that computers can, in theory, **emulate** human intelligence – and **exceed** it."*

Stephen Hawking (2016)

Outline of Presentation

- Motivation
  - **Crisis Index** as (another) financial stability indicator.

Outline of Presentation

- Motivation
  - **Crisis Index** as (another) financial stability indicator.
- Description of Data and Machine Learning
  - Crisis data from **Romer and Romer (AER, 2017)**; [Laeven and Valencia (IMF, 2013)].
  - Textual data from the **OECD Economic Outlook**, [Thomson Reuters News Archives], and IMF Article IVs.

Outline of Presentation

- Motivation
  - **Crisis Index** as (another) financial stability indicator.
- Description of Data and Machine Learning
  - Crisis data from **Romer and Romer (AER, 2017)**; [Laeven and Valencia (IMF, 2013)].
  - Textual data from the **OECD Economic Outlook**, [Thomson Reuters News Archives], and IMF Article IVs.
- Analysis
  - Out-of-sample identification (Nowcasting) and combining models.
  - Local projections and forecasting.
  - COVID-19 Results: U.S. was in a financial crisis?

Outline of Presentation

- Motivation
  - **Crisis Index** as (another) financial stability indicator.
- Description of Data and Machine Learning
  - Crisis data from **Romer and Romer (AER, 2017)**; [Laeven and Valencia (IMF, 2013)].
  - Textual data from the **OECD Economic Outlook**, [Thomson Reuters News Archives], and IMF Article IVs.
- Analysis
  - Out-of-sample identification (Nowcasting) and combining models.
  - Local projections and forecasting.
  - COVID-19 Results: U.S. was in a financial crisis?
- Conclusion
  - **Text and machine learning help**, especially in identifying more severe crisis periods.

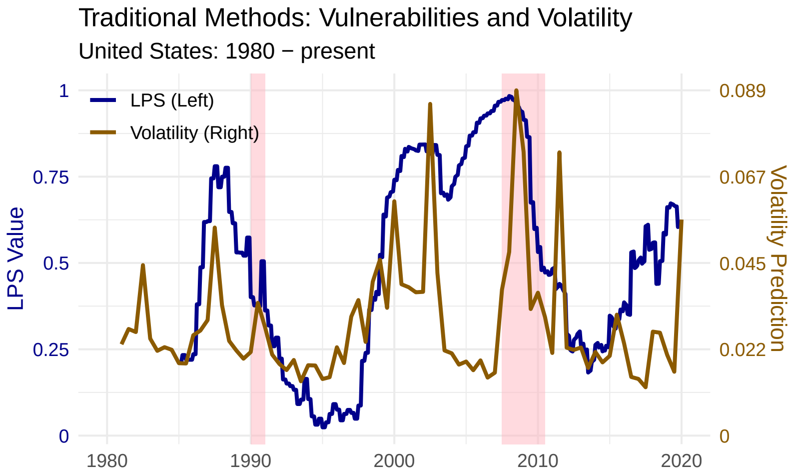
Financial Conditions/Stress and Vulnerability Indexes

- Financial conditions indexes / financial stress indexes
  - Provide useful measures of tight conditions or financial stress in the financial system.
  - Chicago National Financial Conditions Index (NFCI) (Brave and Butters (2012)).
  - Systemic Financial Stress Index (Duprey, Klaus, and Peltonen (2017)).

Financial Conditions/Stress and Vulnerability Indexes

- Financial conditions indexes / financial stress indexes
  - Provide useful measures of tight conditions or financial stress in the financial system.
  - Chicago National Financial Conditions Index (NFCI) (Brave and Butters (2012)).
  - Systemic Financial Stress Index (Duprey, Klaus, and Peltonen (2017)).
- Financial vulnerability indexes
  - Provide indicators of buildup of vulnerabilities that can potentially lead to crises.
  - Useful for figuring out when to activate counter-cyclical capital buffers (CCyB).
  - Credit-to-GDP Gap (Drehmann and Juselius (2014)).
  - AKLPW and LPS Indicators (Aikman et al. (2017), Lee et al. (2020)).

Volatility, LPS Index, and Financial Crises in the U.S.



Shaded areas show Romer Crises of severity 5 or higher

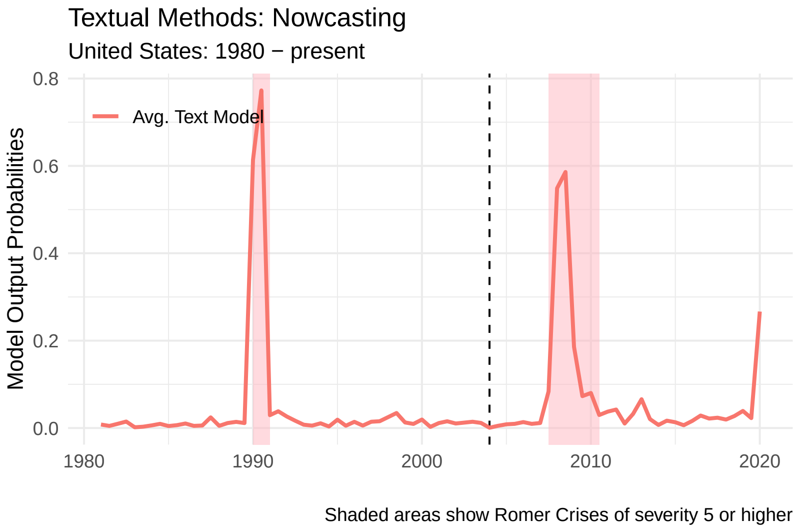
Crisis Index

- Provides information on whether a country is in a crisis state (or not), and does this in a consistent/objective manner.

Crisis Index

- Provides information on whether a country is in a crisis state (or not), and does this in a consistent/objective manner.
- Useful for input into various policies.
  - Crisis management.
  - Macroprudential (whether to set/maintain zero or low CCyB).
  - Monetary policy.
  - Fiscal policy.

CDLS Crisis Index and Financial Crises



Problems with Expert-Identified Crisis Data

- Slow to update
  - Romer and Romer (AER, 2017) still only goes to 2012 (and no sign of an update ever coming).
  - Laeven and Valencia (2020) goes to 2017, updates more frequently.

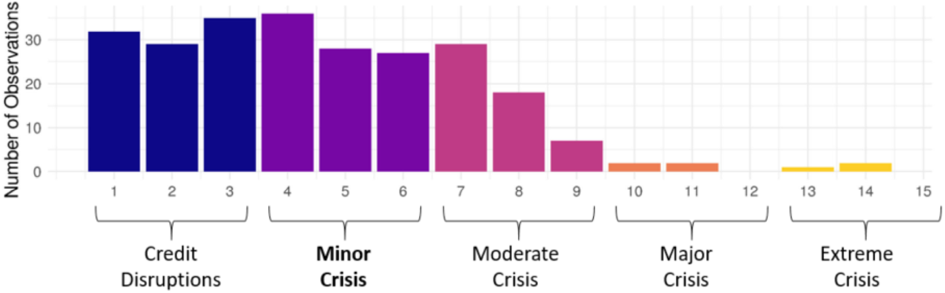
Problems with Expert-Identified Crisis Data

- Slow to update
  - Romer and Romer (AER, 2017) still only goes to 2012 (and no sign of an update ever coming).
  - Laeven and Valencia (2020) goes to 2017, updates more frequently.
- Vague determinations
  - Some crisis definitions have a list of criteria, some are simply "determined by country experts."
  - Start dates are sometimes vague - monthly vs. annual
  - End dates rarely have any explanation.

Problems with Expert-Identified Crisis Data

- Slow to update
  - Romer and Romer (AER, 2017) still only goes to 2012 (and no sign of an update ever coming).
  - Laeven and Valencia (2020) goes to 2017, updates more frequently.
- Vague determinations
  - Some crisis definitions have a list of criteria, some are simply "determined by country experts."
  - Start dates are sometimes vague - monthly vs. annual
  - End dates rarely have any explanation.
- Look-behind bias

Romer and Romer Crises - Narrative Determination



Credit Disruptions: Strains on financial markets, funding problems, increased cost of credit intermediation. Significant enough to be mentioned, but not expected to cause any significant macroeconomic effects.

Minor Crisis: Severe or substantial problems in the financial sector that affect the credit supply without leading to large macroeconomic effects or a worsening economic outlook for the country.

Moderate Crisis: Widespread and severe problems in the financial sector, with a macroeconomic effect, but not described as the financial system seizing up entirely.
This level of severity most coincides with Reinhart & Rogoff, Laeven & Valencia, etc.

Text Useful for Understanding Macro-financial Series or Crises?

- Updates quickly.

Text Useful for Understanding Macro-financial Series or Crises?

- Updates quickly.
- Easy to see which words or phrases are driving determinations of crises.

Text Useful for Understanding Macro-financial Series or Crises?

- Updates quickly.
- Easy to see which words or phrases are driving determinations of crises.
- Use contemporary, publicly available sources available for many countries.

Text Useful for Understanding Macro-financial Series or Crises?

- Updates quickly.
- Easy to see which words or phrases are driving determinations of crises.
- Use contemporary, publicly available sources available for many countries.
- Other research, Angelico et al. (2019), Kalamara et al. (2019), and Cerchiello et al. (2017), find it useful too!

Text Sources

- OECD Economic Outlook
  - Begins in 1967 with more countries in 1980; semi-annual releases.
  - Main text source used in Romer and Romer (AER, 2017).

Text Sources

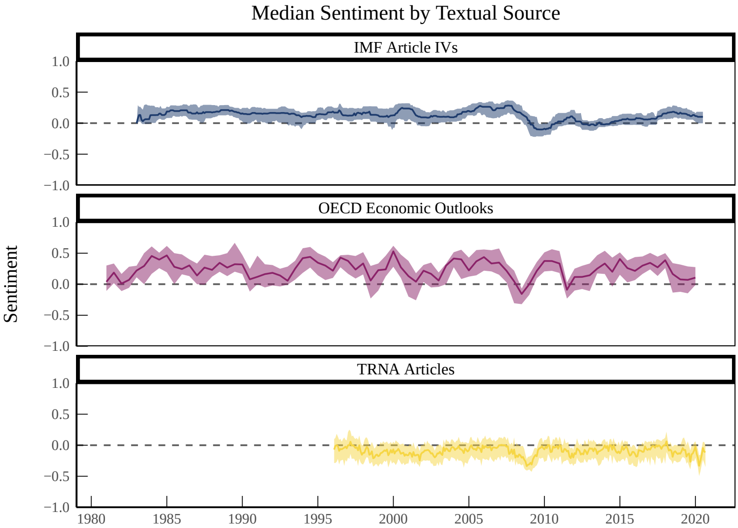
- OECD Economic Outlook
  - Begins in 1967 with more countries in 1980; semi-annual releases.
  - Main text source used in Romer and Romer (AER, 2017).
- Thomson Reuters News Archives
  - Begins in 1996; to-the-minute releases (we aggregate monthly).
  - Wide range of countries covered (64); about 10 million articles.

Text Sources

- OECD Economic Outlook
  - Begins in 1967 with more countries in 1980; semi-annual releases.
  - Main text source used in Romer and Romer (AER, 2017).
- Thomson Reuters News Archives
  - Begins in 1996; to-the-minute releases (we aggregate monthly).
  - Wide range of countries covered (64); about 10 million articles.
- IMF Article IVs
  - Begins in 1984; release times vary.
  - Alternate source used in Romer and Romer (AER, 2017).

Sentiment Scores

- Based on Financial Stability Dictionary, Correa et al. (2017).
- Calculate (Positive – Negative Words)/(Positive + Negative Words).



Showing the Median and 25th and 75th Percentiles across available countries for each text source.

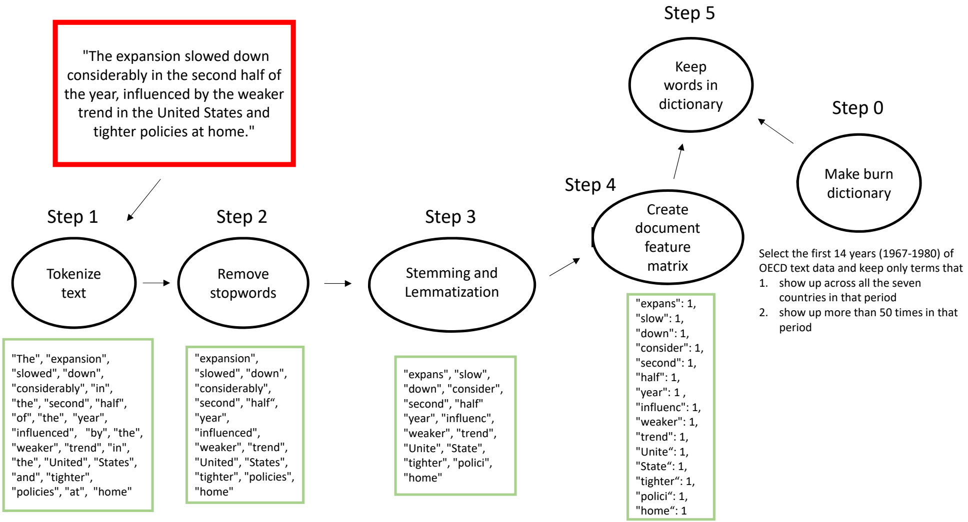
Burning a Dictionary

- Objective
  - Try to maximize out-of-sample predictive power.
  - Do not want particular crisis-specific words like “Asian crisis” or “mortgage backed securities” to drive results.

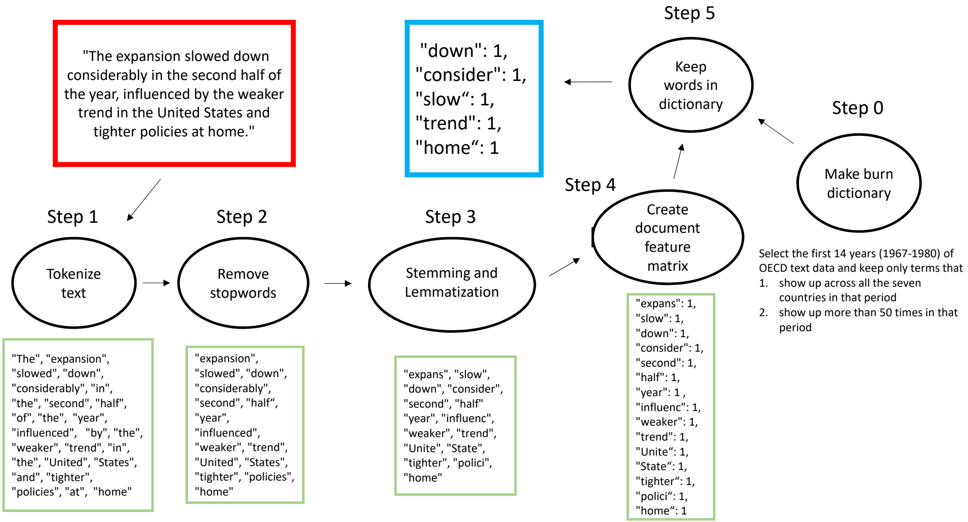
Burning a Dictionary

- Objective
  - Try to maximize out-of-sample predictive power.
  - Do not want particular crisis-specific words like "Asian crisis" or "mortgage backed securities" to drive results.
- "Burn" Dictionary from the main OECD text source
  - Select the first 14 years (1967-1980) of text data.
  - Eliminate most/least frequently used words.
  - What remains is our OECD-based dictionary of 881 terms (unigrams and bigrams).

Data Set-up


"The expansion slowed down considerably in the second half of the year, influenced by the weaker trend in the United States and tighter policies at home."
Step 1
Tokenize text
"The", "expansion", "slowed", "down", "considerably", "in", "the", "second", "half", "of", "the", "year", "influenced", "by", "the", "weaker", "trend", "in", "the", "United", "States", "and", "tighter", "policies", "at", "home"
Step 2
Remove stopwords
"expansion", "slowed", "down", "considerably", "second", "half", "year", "influenced", "weaker", "trend", "United", "States", "tighter", "policies", "home"
Step 3
Stemming and Lemmatization
"expans", "slow", "down", "consider", "second", "half" "year", "influenc", "weaker", "trend", "Unite", "State", "tighter", "polici", "home"
Step 4
Create document feature matrix
"expans": 1, "slow": 1, "down": 1, "consider": 1, "second": 1, "half": 1, "year": 1 , "influenc": 1, "weaker": 1, "trend": 1, "Unite": 1, "State": 1, "tighter": 1, "polici": 1, "home": 1
Step 5
Keep words in dictionary
Step 0
Make burn dictionary
Select the first 14 years (1967-1980) of OECD text data and keep only terms that
1. show up across all the seven countries in that period
2. show up more than 50 times in that period

Data Set-up



"The expansion slowed down considerably in the second half of the year, influenced by the weaker trend in the United States and tighter policies at home."
"down": 1, "consider": 1, "slow": 1, "trend": 1, "home": 1
Step 5
Keep words in dictionary
Step 0
Make burn dictionary
Select the first 14 years (1967-1980) of OECD text data and keep only terms that
1. show up across all the seven countries in that period
2. show up more than 50 times in that period
Step 1
Tokenize text
"The", "expansion", "slowed", "down", "considerably", "in", "the", "second", "half", "of", "the", "year", "influenced", "by", "the", "weaker", "trend", "in", "the", "United", "States", "and", "tighter", "policies", "at", "home"
Step 2
Remove stopwords
"expansion", "slowed", "down", "considerably", "second", "half", "year", "influenced", "weaker", "trend", "United", "States", "tighter", "policies", "home"
Step 3
Stemming and Lemmatization
"expans", "slow", "down", "consider", "second", "half" "year", "influenc", "weaker", "trend", "Unite", "State", "tighter", "polici", "home"
Step 4
Create document feature matrix
"expans": 1, "slow": 1, "down": 1, "consider": 1, "second": 1, "half": 1, "year": 1 , "influenc": 1, "weaker": 1, "trend": 1, "Unite": 1, "State": 1, "tighter": 1, "polici": 1, "home": 1

Supervised Machine Learning

- Avoiding data leakage
  - Given the time-series nature, even validation set respects time (for determining hyper-parameters).
  - We also split training set vs. testing set at 2004 for out-of-sample analysis.

Supervised Machine Learning

- Avoiding data leakage
  - Given the time-series nature, even validation set respects time (for determining hyper-parameters).
  - We also split training set vs. testing set at 2004 for out-of-sample analysis.
- "Chunk" method
  - Up to 2004: Training data (for creating and tuning the models).
  - Post 2004: Testing data (for creating ROC curves and performance metrics).

Supervised Machine Learning

- Avoiding data leakage
  - Given the time-series nature, even validation set respects time (for determining hyper-parameters).
  - We also split training set vs. testing set at 2004 for out-of-sample analysis.
- "Chunk" method
  - Up to 2004: Training data (for creating and tuning the models).
  - Post 2004: Testing data (for creating ROC curves and performance metrics).
- "Expanding" method
  - First train up to 2004.
  - Predict one period ahead, then retrain with that new period, and repeat.

Machine Learning Models

- **Support Vector Machines (SVMs).**
- **Random Forests.**
- GLMNET - Elastic Net/Ridge/Lasso.
- Neural Networks.
- Adaptive Boosted Forests, Extreme Random Forests, Naive Bayes, KNN, etc.

Simple Benchmark Models

- Simple bivariate logistical models

Simple Benchmark Models

- Simple bivariate logistical models
- Realized volatility calculated from daily stock return data.
    - Scaled by country.
    - Incorporated in many financial conditions/stress indexes.
    - Has long histories for many countries.

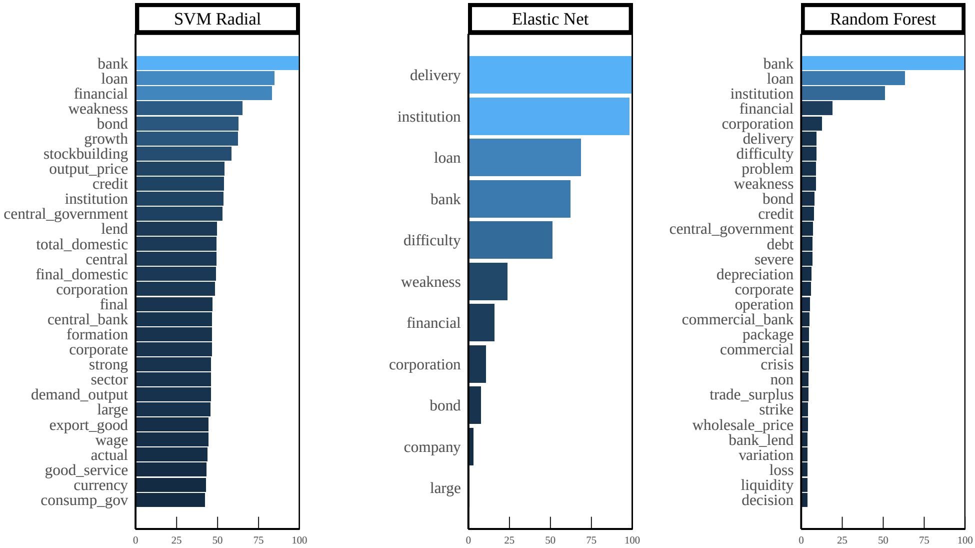
Simple Benchmark Models

- Simple bivariate logistical models
- Realized volatility calculated from daily stock return data.
  - Scaled by country.
  - Incorporated in many financial conditions/stress indexes.
  - Has long histories for many countries.
- Sentiment scores.
  - A simple textual analysis method that often yields promising results.

Out-of-sample Nowcasting Results

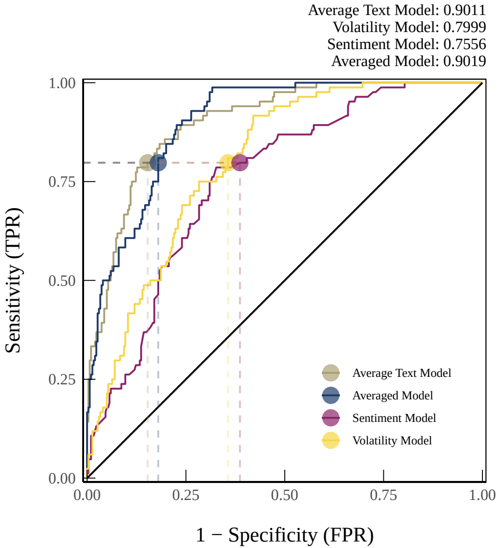
- Variable Importance (in-sample)
- ROC Comparison of Models (out-of-sample)
- Confusion Matrix (out-of-sample)
- LIME (out-of-sample)

Variable Importance Results - 5 (or more severe) Crises on OECD Text

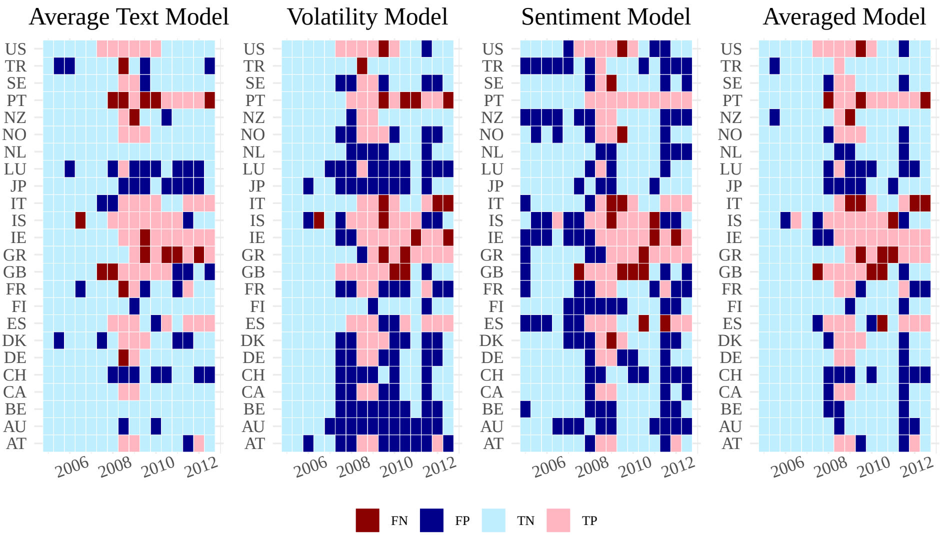


ROC Results - 5 (or more severe) Crises on OECD Text

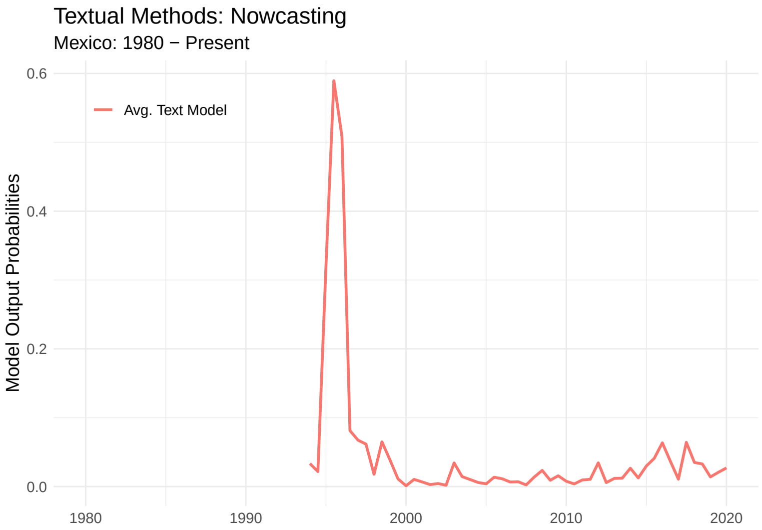
- Text-only model performs pretty well out of sample (2005 to 2012).
- Average model (based on simple averages of model outputs) performs marginally better.



Confusion Matrix - 5 (or more severe) Crises on OECD Text



Out-of-sample Mexico - 5 (or more severe) Crises on OECD Text



LIME Example - 5 (or more severe) Crises on OECD Text

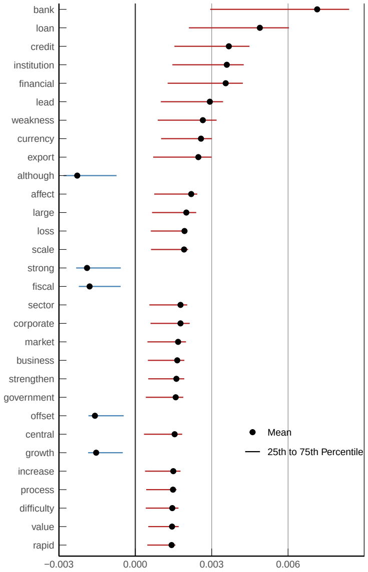
- Local Interpretable Model-agnostic Explanations (LIME) is based on local approximations of feature weights.
- Shapley values allow inference testing - Joseph (2019).
- Provides more intuition than Variable Importance.

US 2008-01-01

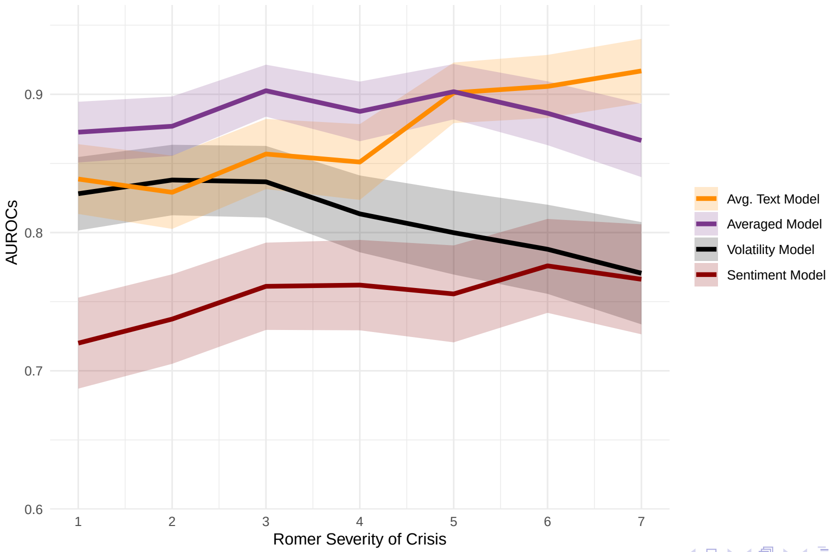
The US economy is at the epicentre of a financial crisis , which is causing considerable disruption to real activity . The trigger for the crisis was a sharp rise in delinquencies on subprime mortgages , which led to large losses on the securities backed by these mortgages . As investors came to realise that mortgage - and asset - based securities were much riskier than supposed , demand for and trading of such products dried up , resulting in further losses on a variety of credit - based securities . Banking institutions linked to these leveraged products incurred large losses , necessitating measures to restore their financial health . This involves a Banks have tightened lending standards [1]
The housing market is tumbling...

LIME Results - 5 (or more severe) Crises on OECD Text

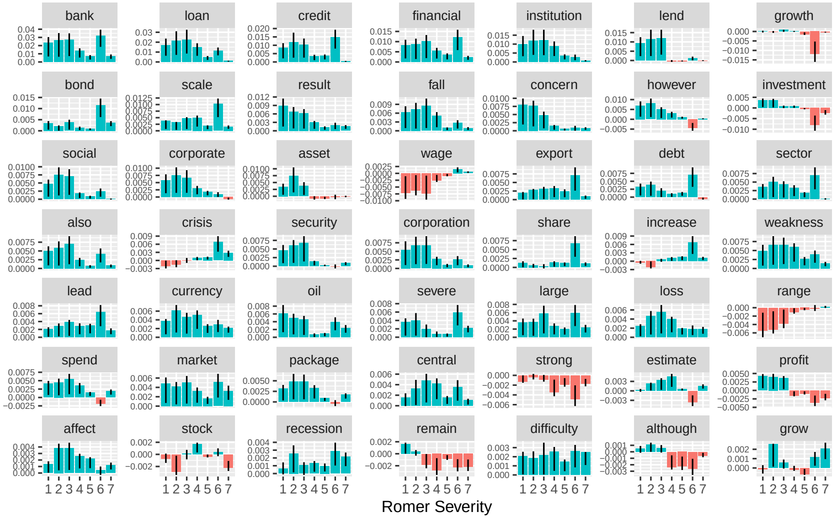
- Top average feature weights suggest that mentions of “bank” and “loan” increases probability of a crisis between 0.5 and 1 percentage point, on average.
- Mentions of “strong” decreases the crisis probability.



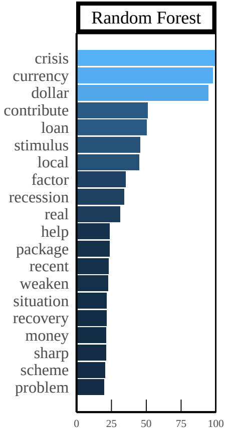
Out-of-sample Results - Different Degrees of Crises on OECD Text

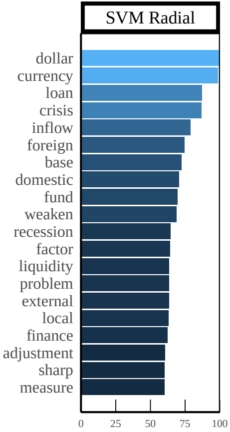


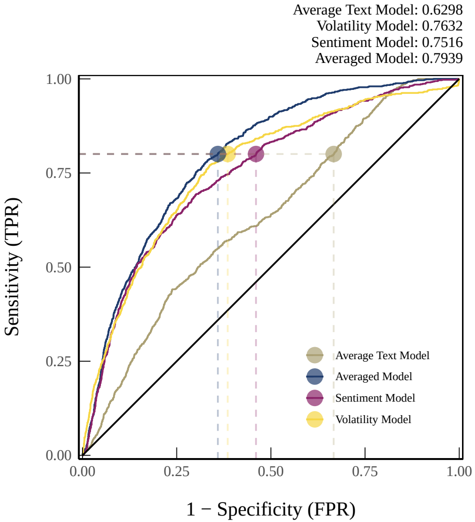
LIME Results - Different Degrees of Crises on OECD Text



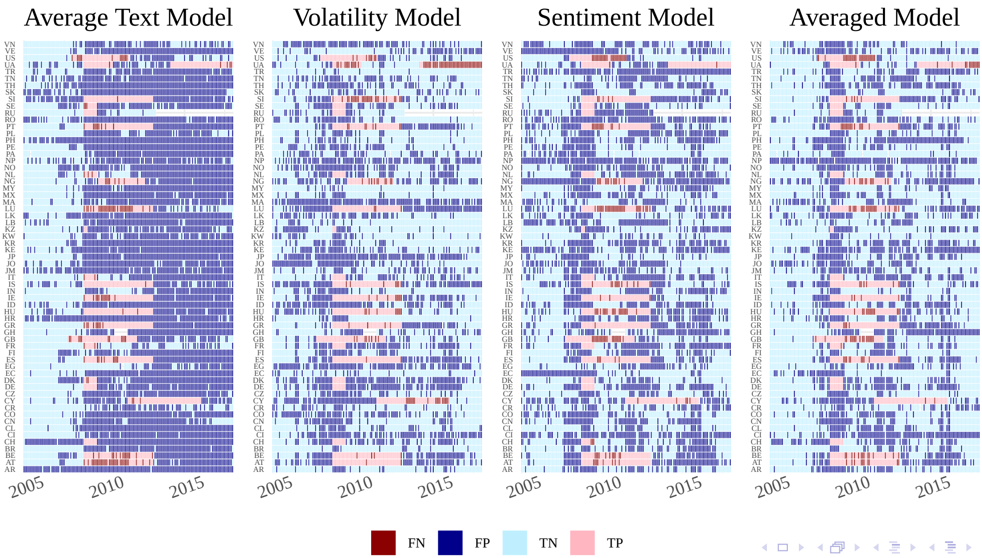
Variable Importance and ROC Curves - LV Crises on TRNA Text



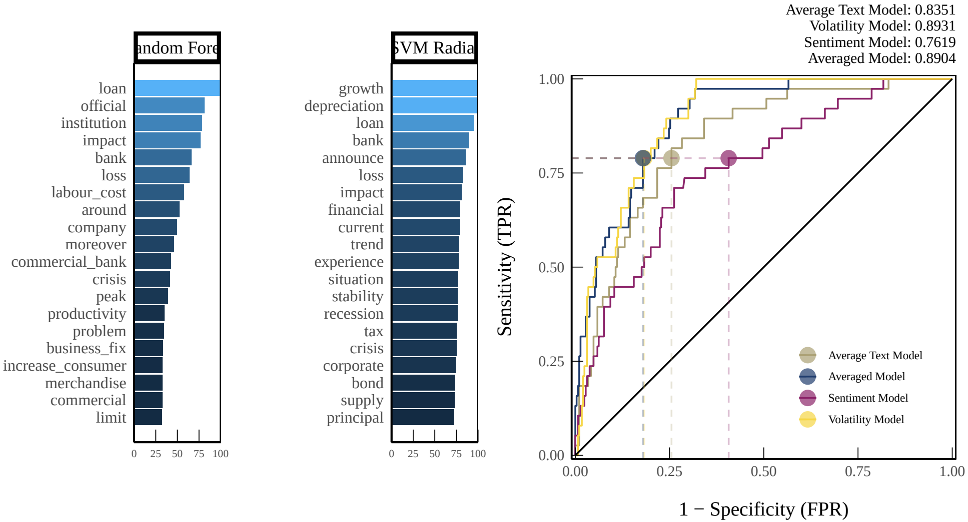




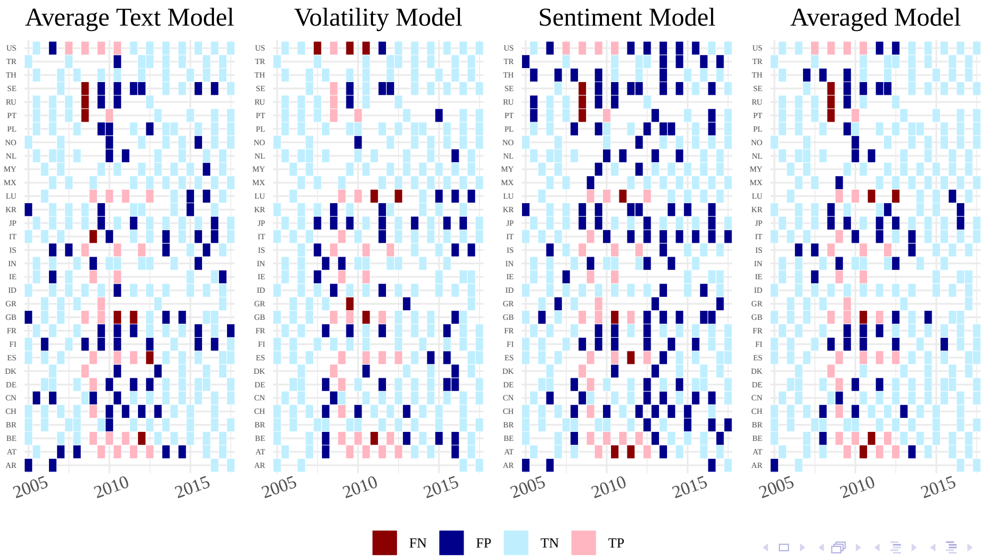
Confusion Matrix - LV Crises on TRNA Text



Variable Importance and ROC Curves - LV Crises on IMF Text



Confusion Matrix - LV Crises on IMF Text



Super Combination Model

OECD Models

1. Text Model SVM-Radial
2. Text Model RF

- Trained on Romer

TRNA Models

3. Average Text Model
4. Volatility Model
5. Sentiment Model

- Trained on LV Banking

IMF Article IVs

6. Average Text Model

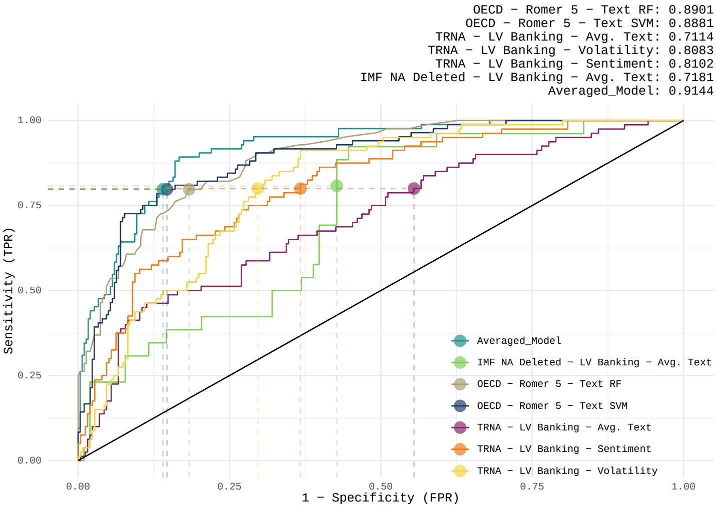
- Trained on LV Banking

Predicting Romer

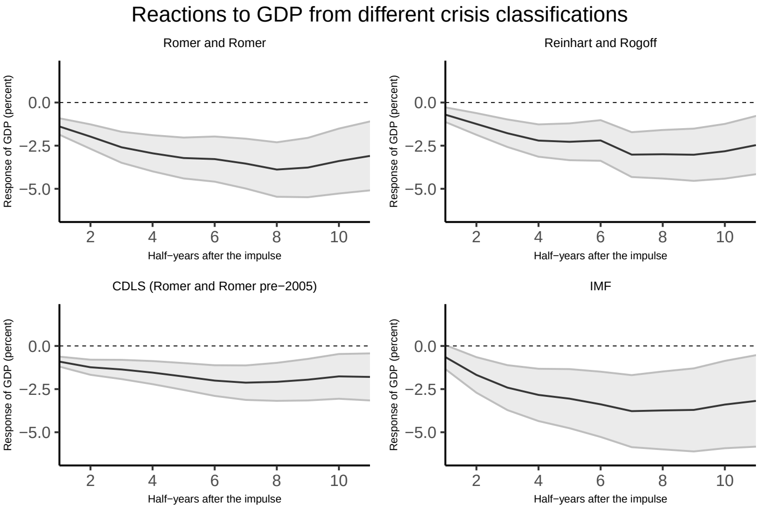
5 (or more severe) Crisis

ROC Results - Super Combination Model

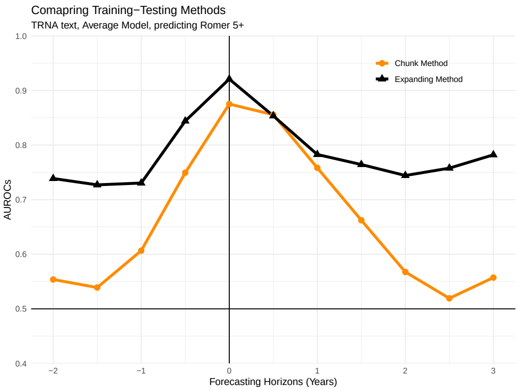
- Combining text can boost AUROC to 0.9144.
- This is a prototype.



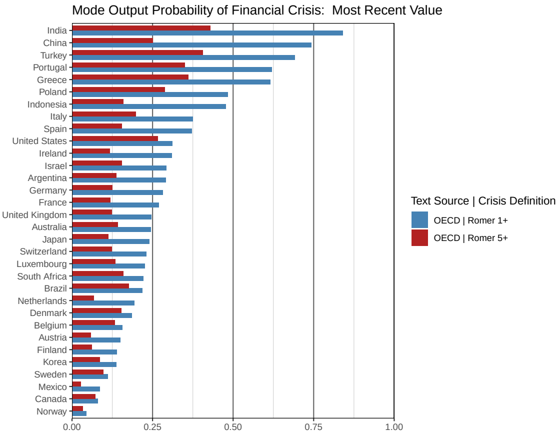
Local Projections from Different Crises Definitions



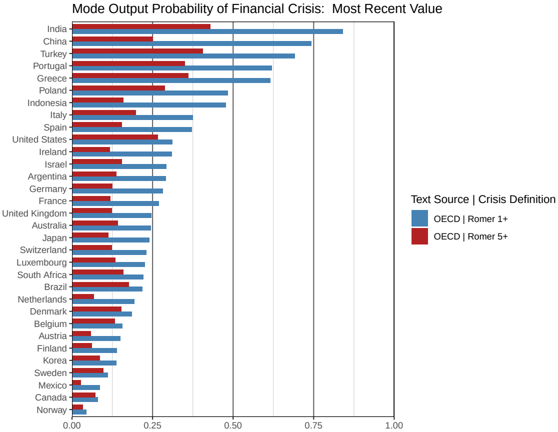
Expanding Window Results - 5 (or more severe) Crises on TRNA

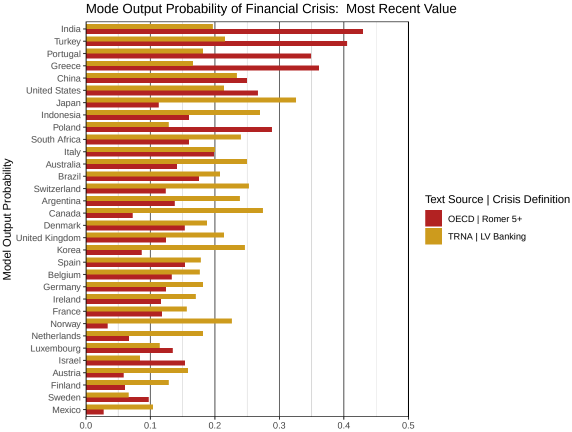


Credit Disruption and Financial Crisis Probabilities during COVID-19



Credit Disruption and Financial Crisis Probabilities during COVID-19





LIME Example, U.S. - 5 (or more severe) Crises on OECD Text

US 2020-01-01
United States
The COVID - 19 outbreak has brought the longest economic expansion on record to a juddering halt . GDP contracted by 5 % in the first quarter at an annualised rate , and the unemployment rate has risen precipitously . If there is another virus outbreak later in the year , GDP is expected to fall by over 8 % in 2020 (the double - hit scenario) . If , on the other hand , the virus outbreak subsides by the summer and further lockdowns are avoided (the single - hit scenario) , the impact on annual growth is estimated to be a percentage point less . The unemployment rate will remain elevated after states lift their shelter - in - place orders , reflecting ongoing difficulties in sectors such as hospitality and transportation , and the sheer scale of job losses . With unemployment remaining high , inflation is projected to stay low , although less so if subsequent lockdowns are avoided .
Massive monetary and fiscal responses have shielded households and businesses , but more will be needed to reduce lingering effects such as large numbers of bankruptcies and labour - market exits . Complementary payments to augment unemployment insurance should continue , while the tax burden of households and businesses should be lowered when they are directly affected by the lockdown . Additional support will be needed to help workers return to work . Some states and local governments will face financial difficulties as their main revenue sources have dried up , and their debt burden will need to be addressed . Importantly , well - designed public financial support for developing a vaccine and treatment of COVID - 19 could help prevent a recurrence of a pandemic again leading to deaths and debilitating the economy .

Other Related Work

- Looked at continuous machine learning models - works great too.
- Looked at predicting the beginning of a crisis - this is harder to do, but still good results.
- Applications to other macro-financial variables are endless.

Conclusion

- Textual data and machine learning provide information not captured by volatility or other data.
- May be particularly useful in detecting and confirming more severe crises.
- Different text sources provide different information.
- In particular, TRNA has more forward-looking components.
- Major caveat is that future crises may look quite different.

Future Romers

Future Romers