Documentation for "Media Sentiment and International Asset Prices"

Samuel P. Fraiberger Do Lee Damien Puy Romain Ranciere

September 12, 2021

1 Contents

This folder contains:

1. News sentiment indicators used in the paper: `./data/`
2. The main results from Figures 1 through 3 of the paper: `./results/`
3. Programs necessary to produce them: `./code/`

2 News articles and sentiment measures

The news sentiment indicators are in `./indicator/news_sentiment.csv`.

- The data file is a daily panel of 25 countries from 1991 to 2019.

In the `.csv` file, we provide three types of news sentiment indicators:

1. `sentiment_country`: Country-specific news sentiment, available from 1991 to 2019
    - The variable was produced by `./code/1_sentiment_country.do`
2. `sentiment_local`: Local news sentiment, available from 1991 to 2015
    - The variable was produced by `./code/2_sentiment_local.do`
3. `sentiment_global`: Global news sentiment, available from 1991 to 2019
    - The variable was produced by `./code/3_sentiment_global.do`

Fraiberger: World Bank & NYU, sfraiberger@worldbank.org; Lee: NYU, dql204@nyu.edu; Puy: IMF, Dpuy@imf.org; Ranciere: USC, NBER & CEPR, Ranciere@usc.edu. The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, the World Bank, their Executive Boards or their management.

2.1 Country-specific news sentiment indicator

- Our dataset of news articles comes from Factiva.com.
- Each article is annotated with topics and geographic tags generated by Factiva using a proprietary algorithm.
- In the main paper, we focused on news articles that were:
  - English articles published by Reuters between 1991 and 2015
  - Tagged with either economic news or financial market news
  - Tagged with one of the 25 countries in our sample (9 AE and 16 EM).
- In this folder, we provide an updated version of the news sentiment index based on Reuters articles published between 1991 and 2019.
- To measure news sentiment, we use a bag-of-words model, which allows us to reduce complex and multi-dimensional text data into a single number.
  - First, we combine existing lists of positive and negative words found in financial texts by Loughran and Mcdonald (2011) and in texts related to economic policy by Young and Soroka (2012).
  - We then expand our lists by including the inflections of each word: for example, the word lose belongs to the negative list, hence we also include the words losing, loser, lost, loss, etc, leading to a final list of 7,217 negative words and 3,250 positive words.
- We define the sentiment of an article as the difference between the share of positive and negative words.
  - In this repository, we provide our baseline estimates where we allow each word to contribute to the sentiment measure proportionally to its frequency of occurrence.
- Finally, we compute a daily sentiment index for each country by taking the average sentiment across articles that are tagged with the country's name.
- The resulting sentiment index has been expressed as z-scores, in units of standard deviations away from the average within each country.
- The news article-level data sourced from Factiva.com is proprietary. Researchers interested in reproducing the analysis will need to obtain a subscription for these data sets.

2.2 Local news sentiment indicator

- Local news sentiment isolates local news from country-specific news articles.
- Specifically, we recompute the sentiment index of each country after excluding any article mentioning any other country.
- Applying this filter allows us to capture the sentiment of purely local news.

2.3 Global news sentiment indicator

- Global news sentiment index captures the tone of news published in the world every day.
- We extract a common factor from our country-specific sentiment series using a Kalman filter.
- Formally, we estimate a single (latent) factor model in the spirit of Stock and Watson (2016).
  - We use an $AR(8)$ for the factor structure and estimate the model using Maximum Likelihood.
- The data coverage for news articles does not start at the same time for all countries.
  - In order to ensure that all countries in our sample have the longest time coverage, we impute these missing observations with zeroes when we estimate the factor model to extract the global news sentiment.

3 Documentation for other variables

The data for producing the main results is a Stata `.dta` file named `./data/regression_sample.dta`. We put together a country-level daily panel from the following sources:

1. *Daily equity returns* are based on each country's main stock market index.
2. *Daily equity trading volumes* are collected from reports by local stock exchanges.
   - Following Campbell, Grossman and Wang (1993) and Tetlock (2007), we compute the de-trended daily log trading volume where the trend is the 60-day rolling average.
3. *Stock market volatility* computed by
   (i) de-meaning each daily stock return,
   (ii) taking the square of this residual, and
   (iii) subtracting the past 60-day moving average of the squared residuals.
4. *VIX* proxies for global volatility.
5. *Dow Jones World Index* measures world equity returns.
6. *S&P Goldman Sachs Commodity Index* measures daily returns in commodity prices.

The data sourced from *Dow Jones World Index* and *S&P Goldman Sachs Commodity Index* are proprietary. Researchers interested in reproducing the analysis will need to obtain a subscription for these data sets.

References

Campbell, John Y., Sanford J. Grossman, and Jiang Wang, "Trading Volume and Serial Correlation in Stock Returns," *The Quarterly Journal of Economics*, 11 1993, *108* (4), 905–939.

Loughran, Tim and Bill Mcdonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, 2011, *66* (1), 35–65.

Stock, J.H. and M.W. Watson, "Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, 2016, pp. 415–525.

Tetlock, Paul C., "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of Finance*, 2007, *62* (3), 1139–1168.

Young, Lori and Stuart Soroka, "Affective News: The Automated Coding of Sentiment in Political Texts," *Political Communication*, 2012, *29* (2), 205–231.