Documentation for "Media Sentiment and International Asset Prices"

Samuel P. Fraiberger Do Lee Damien Puy Romain Ranciere

June 16, 2021

1 Data Documentation

This document describes the replication archive for the above paper. This archive contains the public-use data files used in the paper and the programs to run the analysis. Three main datasets:

1. *News articles and sentiment measures.*

   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. We focused on English articles published by Reuters between 1991 and 2015 and tagged with either "economic news" or "financial market news" as well as with one of the 25 countries in our sample – 9 AE and 16 EM.

   To measure news sentiment, we use a "bag-of-words" model, allowing us to reduce complex and multi-dimensional text data into a single number.

   We define the sentiment of an article j as:

   $$s_j = \frac{\sum_i w_{ij} p_{ij} - \sum_i w_{ij} n_{ij}}{\sum_i w_{ij} t_{ij}},$$

   where p_{ij} is the number of occurrences of positive word i in article j, n_{ij} is the number of occurrences of negative word i in article j, t_{ij} is the number of occurrences of word i in article j, and w_{ij} is the weight associated with word i in article j. In our baseline estimates, we take $w_{ij} = 1$, allowing each word to contribute to the sentiment measure proportionally to its frequency of occurrence.

   Next, we compute a daily sentiment index for each country by taking the average sentiment across articles that are tagged with the country's name. Finally, we normalize each country sentiment index by computing its z-score.

   The data coverage for news articles does not start at the same time for all countries. For example, our coverage of news articles from Russia starts only from the end of 1991, whereas the news coverage for most countries start from the beginning of the same year. In order to ensure that all countries in our sample have the longest time coverage, we impute these

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.

missing observations with zeroes when we estimate the factor model to extract the global news sentiment. However, our main results remain unchanged if we restrict the sample into a balanced panel across countries. Results are available upon request.

2. *Asset prices and related variables.*

   We assemble a country-level daily panel from the following sources:

   (a) *Daily equity returns* are from each country's main stock market index. Source: Bloomberg and Datastream.

   (b) *World equity returns* are from the Dow Jones World Index. Source: Haver Analytics.

   (c) *Daily equity trading volumes* proxy for market liquidity and are collected from reports by local stock exchanges. Source: Bloomberg and Datastream. Following Campbell, Grossman and Wang (1993) and Tetlock (2007), we compute the de-trended daily log trading volume using a rolling average of the past 60 days to define the trend.

   Not all countries have full data coverage on equity trading volumes. For example, to our knowledge, volume data on Chile's stock market is not available for our sample period. Although controlling for volumes can restrict the sample over which we estimate, our results are not sensitive to including or not including this control. These results are available on request.

   (d) *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.

   (e) *S&P Goldman Sachs Commodity Index* measures daily percentage changes in commodity prices. Source: Haver Analytics.

   (e) *CBOE VIX* proxies for global volatility. Source: Chicago Board Options Exchange. Retrieved from FRED: https://fred.stlouisfed.org/series/VIXCLS/

3. *Capital flows.* We collected data on daily equity fund flows from EPFR Global, which contains information on the asset allocation of a large number of international equity funds.

   Most funds followed by EPFR Global are (i) located in AE and (ii) account for a significant share of the external funding received by EM. As a result, the country flows dataset has proved to be a good proxy of total gross inflows in (or out) of EM.

   We focused on the "equity country flows" dataset, which reports the estimated daily amount of equity funding in US dollars that entered or left each country due to international funds' portfolio reallocation. Our dataset of equity flows covers 16 EM between 2008 and 2015.

   EPFR Global also provides a breakdown of the country flows based on the (i) domicile (foreign vs. local) or (ii) type (ETF vs. non-ETF (active)) of the funds.

4. We also draw from other datasets:

   (a) *US Economic Policy Uncertainty Index (EPU)* from Baker, Bloom and Davis (2016). Retrieved from: https://policyuncertainty.com/media/US_Policy_Uncertainty_Data.xlsx

   (b) *MSCI Emerging Markets (EM) Index.* Retrieved from Datastream.

   (c) *Citi Index of Economic Surprises (CESI)* captures deviations between actual macroeconomic data releases and the Bloomberg survey median in key countries. We use

information from economic data surprises in the US, the Euro Area, China, and the G10 countries, which are available at the daily frequency since 2001. Retrieved from Haver Analytics.

(d) *International Financial Statistics (IFS).* Quarterly data on real GDP growth for each country in our sample. Retrieved from IMF Data: https://www.imf.org/en/Data

(e) *US Dollar Exchange Rates.* Daily exchange rates retrieved from BIS Statistics: https://www.bis.org/statistics/full_webstats_xru_current_d_dataflow_csv_row.zip

(f) *NBER Recession Indicators for the US.* Monthly dummy variables where a value of 1 indicates a recessionary period. Retrieved from FRED: https://fred.stlouisfed.org/series/USREC

The data sourced from Factiva.com, Datastream, EPFR, Haver Analytics, and Bloomberg are proprietary. Researchers interested in reproducing the analysis will need to obtain a subscription for these data sets.

2 Code Documentation

1. The code is stored in `./code`

   (a) Contains all programs needed to go from the raw input to the results in the paper.

   (b) `./code/main.do` is the master do-file. Running this program executes the replication by calling the other do-files in this folder. The do-files are for Stata version 14 or above.

   (c) `./code/ado` contains any additional Stata modules called by `./code/main.do`.

2. Stata log files in `.txt` format stored in `./log`

3. Input to the programs are stored in `./data`

   (a) `./data/news.csv` contains the news articles and sentiment measures. Additional articles from `./data/features-1991-1996-missing-countries` are appended to this dataset.

   (b) `./data/INDEX.csv` contains each country's daily stock market index.

   (c) `./data/global proxies.xlsx` contains the Dow Jones World Index.

   (d) `./data/Trading volumes.xlsx` contains each country's daily equity trading volumes.

   (e) `./data/Commodity indexes.xlsx` contains the S&P Goldman Sachs Commodity Index

   (f) `./data/VIXCLS.csv` contains the daily CBOE VIX.

   (g) `./data/epfr` contains the raw data files for equity fund flows downloaded from EPFR.

   (h) `./data/EPU_All_Daily_Policy_Data.csv` is the Economic Policy Uncertainty Index (EPU).

   (i) `./data/MSCI.csv` contains the MSCI Emerging Markets (EM) Index.

   (j) `./data/CESI.xlsx` contains the Citi Index of Economic Surprises (CESI).

   (k) `./data/IFS_09-25-2020 21-57-15-95.csv` contains the International Financial Statistics (IFS) from the IMF. Metadata: `./data/Metadata_IFS_09-25-2020 21-57-15-95.csv`.

   (l) `./data/WEBSTATS_XRU_CURRENT_D_DATAFLOW_csv_row.csv` contains the daily US Dollar Exchange Rates data from the BIS.

4. Data Preparation:

(a) `./code/news.do` prepares the news articles and sentiment measures.

(b) `./code/data.do` prepares the asset prices and related variables.

(c) `./code/epfr.do` prepares the capital flows data from EPFR Global.

(d) `./code/dfm.do` estimates the factor model to extract the common factor ("global news sentiment") from our initial sentiment measures.

(e) `./code/clean.do` puts together the sentiment measures, asset prices and related variables, and capital flows data into a single country-level daily panel.

(f) `./code/table_#.do` makes tables where `#` is the table number in the paper.

(g) `./code/figure_#.do` makes figures where `#` is the table number in the paper.

5. Output to the programs: `./results` contains

   (a) Figures in `./figure_#.pdf` format where `#` is the table number in the paper.

   (b) Tables in `./table_#.tex` format where `#` is the table number in the paper.
   Some `.tex` tables have been converted from Excel format from `./results/results.xlsx`.

References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1593–1636.

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.

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.