Data files for the paper Panel Data Nowcasting: The Case of Price-Earnings Ratios

Data sources

Public data sources:

- News data – http://www.structureofnews.com
- FRED – https://fred.stlouisfed.org
- FRED-MD – https://research.stlouisfed.org/econ/mccracken/fred-databases/

Proprietary data sources:

- Wharton Research Data Services (WRDS)
- Center for Research in Security Prices (CRSP)
- Institutional Brokers' Estimate System (I/B/E/S)
- RavenPack

Packages

All packages needed to sort the data are installed and loaded at the beginning of each code file.

Data

Target variables (proprietary data) – earnings_data_files.zip

Target variables are computed from the CRSP and I/B/E/S databases. Folder `earnings_data_files` contains `R` codes to sort the data.

Files that are need to be downloaded from CRSP and I/B/E/S through WRDS:

- `eps-wrds-new-ibes.csv` – full I/B/E/S data file from 1999-11-01 to at least 2017-07-01.
- `ibes-crsp-link.csv` – I/B/E/S-CRSP databases link table.
- `prices-crsp-daily.csv` – full CRSP daily stock price data from 1999-11-01 to at least 2017-07-01.
- `prices-crsp.csv` – full CRSP monthly stock price data from 1999-11-01 to at least 2017-07-01.
- `siccodes10.csv` – link table that links sic codes to Fama French 10 industry classification.
- `permno-ff10.csv` – link table that links PERMNO code to Fama French 10 industry classification.

In addition, `ticker_exclude.RData` file contains a list of tickers of the firms that we excluded from the analysis due to erratic time series for earnings and/or earnings forecasts. The file is available in the folder `earnings_data_files`.

Code

- `run_all.R` – R code that sorts the data files. All files should be kept in the folder `earnings_data_files`. The output file is `pe_ratio_ibes_crsp_ff.RData`. Paste this output to `compile-all-data` folder.

Predictor variables (proprietary & public data) – compile-all-data.zip

Folders:

- `AggregateTextual` (public) – aggregate news data. The data file `Monthly_Topic_Attention_Theta.xlsx` was downloaded from http://www.structureofnews.com. The data file is available in folder `compile-all-data/AggregateTextual`. The file in the folder `AggregateTextualData.R` sorts these data and outputs the file `AggregateTextualData.RData`.

- Firmdaily (proprietary) – daily firm-level data. The data was downloaded from CRSP. The full list of CRSP stock price data should be downloaded and stored in RETURNS_DAILY.csv, RETURNS_DAILY2.csv and RETURNS_DAILY3.csv files for the period 1999-11-01 to 2017-07-31. The file FirmDaily.R in the folder computes the required daily data. The output is data_daily_firm.RData file.
- FREDdaily (proprietary & public) – daily aggregate data. S&P 500 daily returns should be downloaded separately from CRSP due to it is a restricted series on FRED data source. The downloaded return series should cover 1999-11-01 to 2017-07-31 period and be stored in the file snp500ret.csv. FREDdaily.R code downloads and sorts the data from FRED data source and augments the data with the S&P 500 daily returns. The final output is daily_data.RData.
- MacroMD (public) – monthly real-time macro data. Data was downloaded from https://research.stlouisfed.org/econ/mccracken/fred-databases/.
- RavenPack2020 (proprietary) – daily firm level news data from RavenPack Done Jones and PR editions. Further details available in a separate readme file in the folder RavenPack2020.

Final data

The file data_full_midas.R computes the final data set merging all data files. The output is data_full_midas.RData. The final data set should contain 210[1] firms and 42 out-of-sample quarters. It should be used with the panel regression functions available in the R package midasml which can be install as follows:

```
install.packages("midasml")
```

[1]See the Online Appendix for the final list of firms with both PERMNO and RavenPack identifiers.