The data file consists of monthly Lexis-Nexis article counts for publicly listed stocks from January, 2010 through December, 2016. I received the raw files from David Rakowski at UT-Arlington. These files were collected but not used in:

Rakowski, Shirley ans Stark, 2020. Twitter activity, investor attention, and the diffusion of information, *Financial Management*, 50 (1).

The files were not used because they required extensive processing which I have done, reading the 400 csv files using Matlab, and collecting them in a SAS data set for public use.

The file consists of article counts in 109,263 stock months. Average number of stocks per month is 1,301.

I had some overlapping Ravenpack data from 2011, and the article counts in this data have a correlation of 0.325 with the number of Dow Jones news articles in Ravenpack, and a correlation of 0.322 with all articles in that database.

How that correlation is viewed may depend on how you squint, or what your budget is. The data clearly is not perfectly correlated with the Ravenpack product, but could also be viewed as a reasonable instrument for media activity.

Some entries were not readable by Matlab, and one file had corrupt data. Overall, I estimate that I was able to process > 99% of the data Rakowski was kind enough to send me.

All errors in the data should be attributed to me.

Paul Irvine
Kleinheinz Professor of Finance
Neeley School of Business, TCU
May 18, 2021