Media Index

Background

One of the more fascinating sections of the WSJ is on the inside of the back page under the standing headline "Abreast of the Market." There you can read each day what the market did yesterday, whether it went up, down or sideways as measured by indexes like the Dow Jones Industrial Average . . . . In that column, you can also read selected post-mortems from brokerage houses, stock analysts and other professional track watchers explaining why the market yesterday did whatever it did, sometimes with predictive nuggets about what it will do today or tomorrow. This is where the fascination lies. For no matter what the market did–up, down or sideways–somebody will have a ready explanation.
— Vermont Royster, *Thinking Things Over Abaft of the Market*, 1986/1/15

Are the media just a sideshow to financial markets – or do the dire pronouncements of newshounds have any real effect on share prices? Economist Paul Tetlock used the Harvard psychosocial dictionary, which classifies words by the moods they convey, to test whether pessimistic words affect the stock market. He found that a pessimistic Abreast column with many words that communicate weakness and negativity not only describes what happened in yesterday's trading but also foreshadows the next day's activity. As he explained in an interview, "*Some of the words that we read in the financial news have greater entertainment value than information value. By the time news is printed in the WSJ, it is already known to most sophisticated institutional investors. Some WSJ readers may not realize this point, especially when a news column presents stale information in a compelling and colorful way. Thus, individuals may trade based on news that has already been incorporated in stock prices, leading the market to overreact to past information. Eventually, sophisticated investors restore order in the market, bringing prices back in line with fundamentals.*"

Data & Task

The data set includes 6,600 articles on the WSJ column "Abreast of the Market." Your task is to derive a **media index** for each day based on the categories in the General Inquirer's Harvard psychosocial dictionary.

- Use the two large valence categories, *Positiv* and *Negativ*, in the General Inquirer to define a media index. The *Positiv* category contains over one thousand words of positive outlook and the *Negativ* category contains over two thousand words of negative outlook.

- There are 182 categories. Use LSA to extract the dimension with the largest variation to define a media index.