U.S.A. Independent Study Class Proposal

**Main Research Question:**

How do non-financially related news stories associated with a publicly traded company impact short and medium term price changes?

**Motivation:**

On Wednesday October 22, 2014, a complaint was filed to the Supreme Court of the State of New York by Christina Di Mauro Kelly against her husband Sage Kelly, CEO of Jefferies & Co., for allegedly practicing illegal drug use and engaging in extramarital affairs to, in one case, coerce Marc Beer, CEO of Aegerion Pharmaceuticals into a business relationship. During this incident, following a night of drinking, Marc Beer was reported to have been with his girlfriend and both Kelly’s in the same bedroom. Beer then suggested switching partners with Sage Kelly’s wife. Knowing her husband’s wish to secure a business relationship with Beer, Christina Kelly “felt responsible not to disappoint Marc” and followed through with the suggestion. This story’s impact, along with the original complaint, fell heaviest on AEGR (Jefferies was not a publicly traded company) whose stock price fell 8% on Friday October 24, 2014 before rebounding back to its pre-scandal levels on Tuesday October 28, 2014 at $33.69.

Sources:

<http://www.smallcapnetwork.com/Kelly-vs-Kelly-Sex-Drugs-Lies-and-Two-Biotech-Earnings-Report-AEGR-SGEN-LUK/s/via/3414/article/view/p/mid/1/id/1983/>

<http://www.thestreet.com/story/12931859/1/aegerion-board-expresses-support-for-embattled-ceo-amid-drug-use-accusations.html>

**Rationale:**

News articles can reflect initial consumer sentiment towards publicly traded companies and can potentially be used to predict a downward or upward trend in the corresponding stock price. However, most news stories represent noise – biased viewpoints, speculation, etc. – or surface to the media only after the fact. This research question aims to drill down to news that, while having an impact on the company’s stock performance, suggest no medium-term implications on the company’s fundamental growth. Examples of such stories include news involving drugs, scandals, alleged accusations, etc. that tarnish a company’s image but not necessarily its products. These signals may then be captured and used to predict future price direction.

**Learning Objectives:**

Committee members, by the end of this project, would be able to do a variety of text mining tasks (text extraction, cleaning, etc.), understand basic finance theory, understand natural language processing, understand machine learning techniques, be able to do exploratory analysis, develop better presentation skills, become more experienced using LaTeX, and become more experienced R users.

**Fall 2015 Schedule**

There will be weekly meetings on Wednesdays from 8-9pm at: location TBD. Note that these are rough week breakdowns and tasks assigned will very likely take longer than one week to be completed.

Week 1: Team introductions, goals & expectations, Present project schedule and layout, Discuss main project, Outline of project, go over text mining concepts and WRDS intro, assign two people to extracting news sources (SeekingAlpha, Yahoo Finance, Reuters) and assign two people to extract stock prices (WRDS database – SAS may be required).

Week 2-3: Individual presentations on first week’s tasks (go through code, understand how to use WRDS). News people start cleaning and subsetting news articles based on key terms (note the time of publication!), prices people get prices on relevant stocks (stand by for news people) at relevant times.

Week 4: One group look at short-term stock returns. Another group look at medium-term stock returns (few days after). Visualization. Interpretation.

Week 5: Compile into Powerpoint slides. Assign everyone to document their methodology and results using LaTeX. Prepare for club presentation.

If there is extra time:

Week 6: Change project perspective to that of prediction, assign two people to extract, clean, and format all news articles into a DTM using tfidf as metric (and/or any other ngram), assign another two people to prepare stock price changes for a variety of different times (1 min., 5 min., etc.) for each news article to be used as response variables

Week 7: Assign machine learning tasks such as linear regression, random forest, and other models (one model per member is an option) and determine error metric (MSE is an option). Visualization and interpretation.

Week 8: Presentation of findings for each member. Compile into Powerpoint slides. Assign everyone to document their methodology and results using LaTeX, and prepare for club presentation.

Potential Readings: <https://www.bc.edu/content/dam/files/schools/csom_sites/finance/Giannetti102914.pdf>

To understand big picture effect of corporate frauds on stock market equity holdings of a household.

<http://www.rinfinance.com/agenda/2012/talk/Nagar+Hahsler.pdf>

To understand how to do crude sentiment analysis of news articles.

<http://www.diva-portal.org/smash/get/diva2:440508/FULLTEXT01.pdf>

Great overview on introductory finance concepts, text preprocessing of news articles, and sentiment analysis of those articles.

**Evaluation**

Members will be evaluated based on their final presentation. Since this is a highly collaborative project with many intermediary steps, what each member has learned, how well he/she has learned, and the quality of overall analysis will be well-represented in the final presentation and research paper.