

1116-00-298

**Jeremy Tillay\*** (jrt5493@gmail.com), **Adela Yang**, **Yilun Chen** and **Xiudi Li**. *Text-Mining and Topic Modeling the Wall Street Journal to Find Market Inefficiencies.*

In this project, we aim to find semi-strong form inefficiencies in the market and delays in stock price changes following the release of publicly available information online. Specifically, we analyze all publicly available articles on the online archives of *The Wall Street Journal*. We make predictions about stock behavior following a day by assuming it will be similar to days when similar news are published online.

In order to analyze text, we implement Natural Language Processing (NLP) to refine our text documents and Latent Semantic Indexing (LSI) to group words into topics and analyze the frequency these topics appear in a given day. Then, we make predictions about the behavior of the stock market under the assumption that its behavior will mimic days when there is a similar frequency distribution of these topics. (Received August 23, 2015)