Stock prices and Search engines

Nick Lepley
Capstone project

Background

- Asset price is information
  - Price=expected future value of the asset
  - The only way to beat the market is to bring information to it

- Before people trade, they research
  - Search volume potentially related to trading volume

Plan

- Compare historical stock data from Quandl to search queries in Google Trends data

- Likely timescale is days-weeks.

- Questions:

  - Is search volume lagging/leading trading volume and price?

  - Are volatility and extreme events predicted?

 - Can we build a portfolio strategy that consistently outperforms S&P index?

Difficulties

- Google Trends
  - No API
  - Requires a Google Account to access
  - Enforces request limits
- Search on names is noisy
  - "Taco Bell" searches surge each year on May 5th, unlikely to be due to investor interest

Progress so far

- Created a pseudoAPI for Google Trends using python requests library and process calls to download thousands of CSV files

  - Scraped data for searches on 3000+ stock tickers and company names from 2004-Present

- Suggestive preliminary results, but poor visuals

- Currently developing a Heroku app for visualizing the data

Sample Results
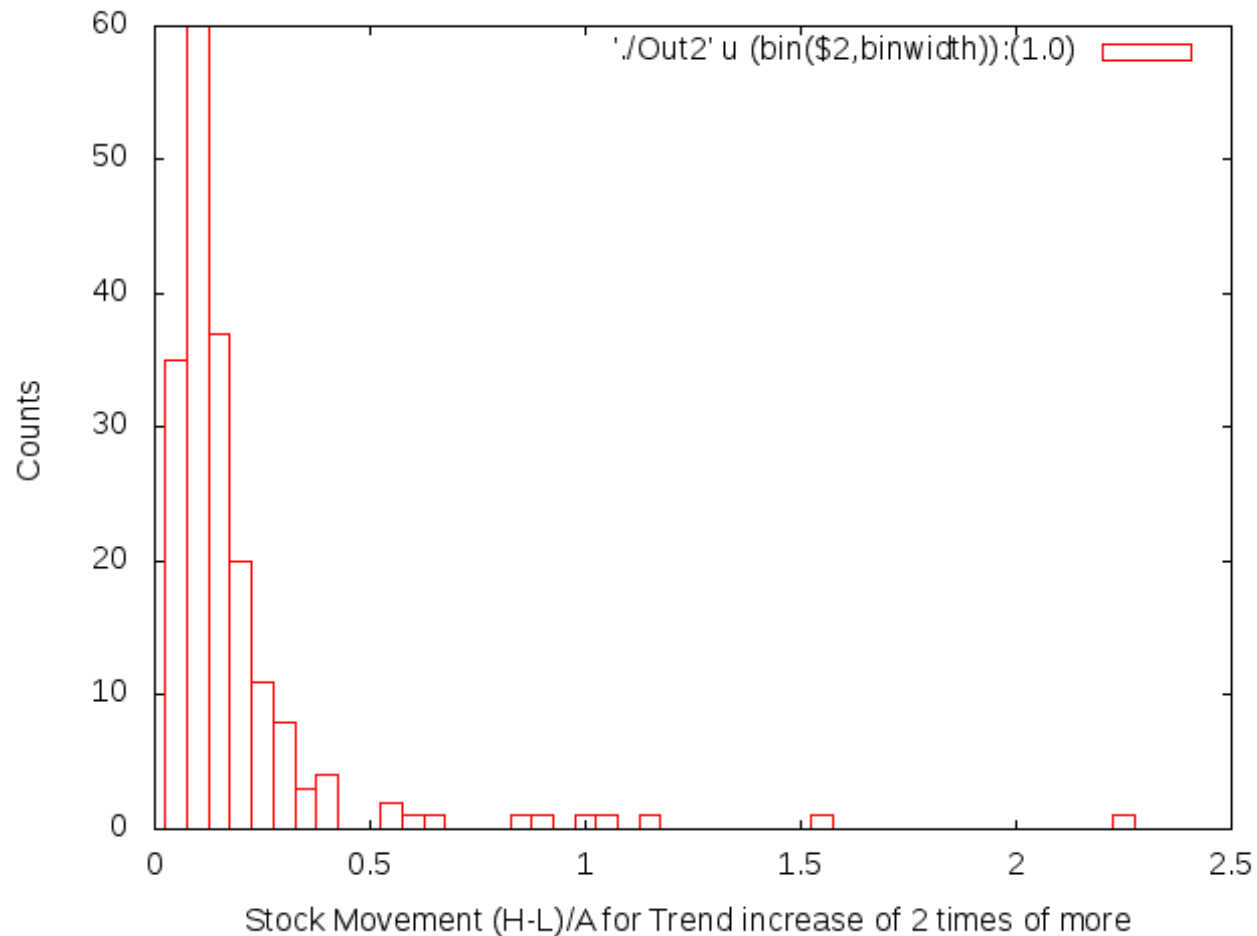
Stock movement histogram for high volatility in search

Sample Results (cont'd)
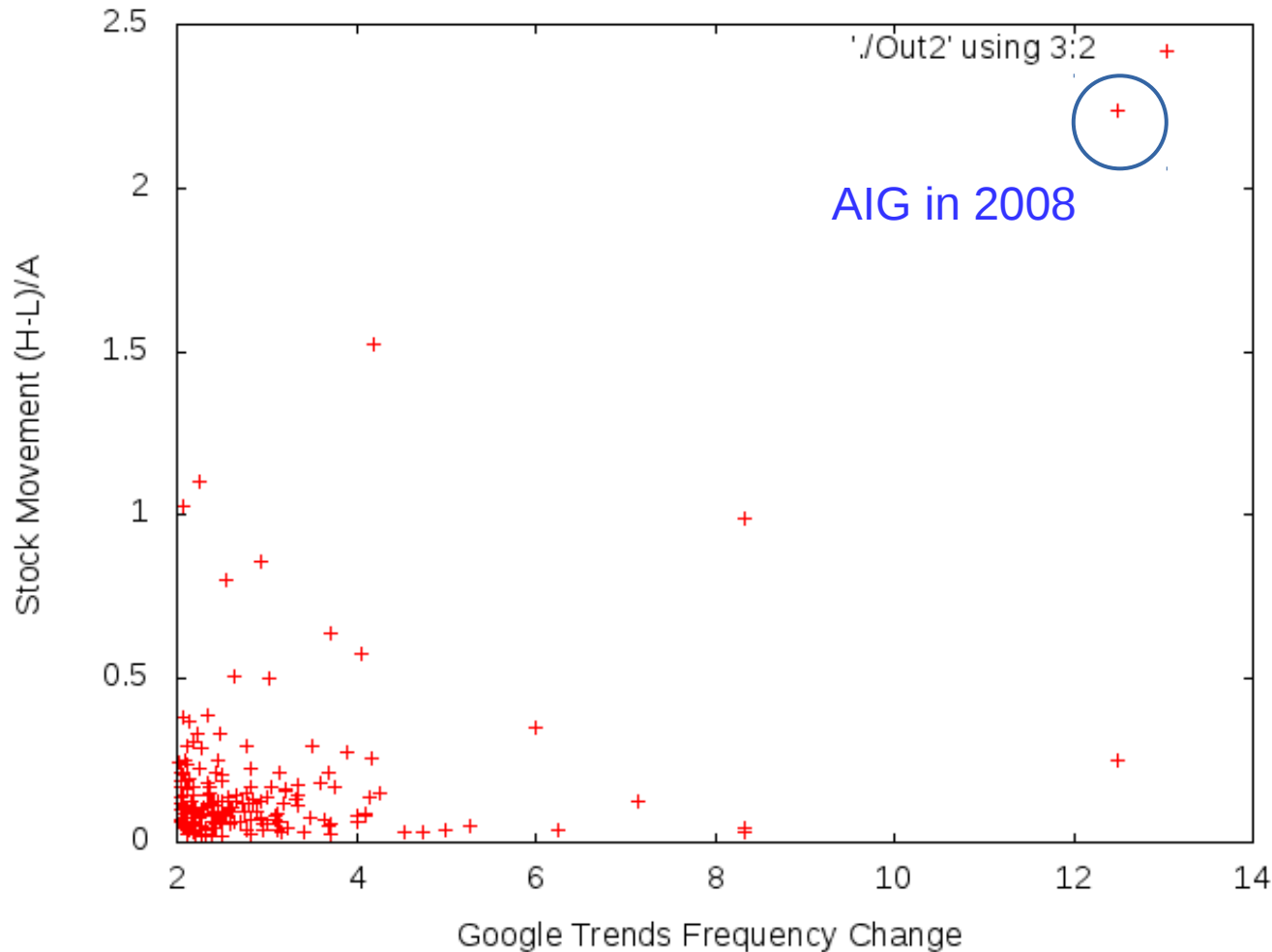
Tickers search volatility versus stock volatility

Actually interested in outliers

Coming Soon

- Better visualization

- More search terms

- More sophisticated models

    - Leading/lagging time series analysis

    - Expected $ value of indicators