Construction of Google Search Indices by Applying Principal Component Analysis

Sile Li & Shannon Callan, CFA

Contacts:
enquiries@eaglealpha.com
(646) 843-6048

20th May 2016
Eagle Alpha Enables Asset Managers to Obtain Alpha from Alternative Data
Internet Search Data in Economics and Finance Research

- Internet Search Data (Google, Baidu)
  - Advantages: new information, continuous information, broadness, easy access
  - Disadvantages: relatively short history, sampling bias, diverse and dynamic search behaviors

- Studies Have Shown Predictive Value for Economic and Financial Metrics
  - Unemployment rate, house price, consumption, tourism, inflation, commodity price volatility, stock market return and volatility etc. in various countries.

- Evolution of Methodology

  **Most popular terms** (e.g. "job search", "jobs", "monster.com", "resume", "employment" and "job listings" - Ettredge et al. (2005))

  **Highly correlated terms** (e.g. "JSA" - McLaren & Shanbhogue (2011))

  **Highly correlated terms** (e.g. "Free shipping" - Choi & Varian (2009))

  **Google categories & Factor modeling** (e.g. 56 categories for US consumption - Vosen et al. (2009))

  **The FEARS index** (30 terms such as "crisis", "inflation" - Da et al. (2011))

  **Bayesian variable selection** (e.g. 151 terms such as "unemployment office", "filing for unemployment" - Scott & Varian (2014))

  **Topic extraction & PCA index** (e.g. 45 terms such as "municipal bond", "over capacity" - Dong & Bollen (2015))
Eagle Alpha Approach

- Term Selection (packages *BMA, BSTS*)
  - Primitive list
  - Harvard dictionary economic keywords
  - Find related searches in Google Trends and Google Correlate
  - Filtering via correlation, Bayesian variable selection, term loading in PCA analysis
- Obtain Data and Process (packages *gtrendsR, zoo, robustHD, seas*)
  - Convert time frequency, remove outliers, winsorize, and seasonal adjustment
- Apply Principal Component Analysis (function *prcomp*)
- Index Creation
  - Create index from top principal component(s)
  - Use of rolling window
  - Track category loadings within the index

Eagle Alpha’s Online Search Index gauges search behaviours related to specific economic activities. It is a composite indicator which measures the co-movement of multidimensional and dynamic search terms.
Eagle Alpha Online Search Indices

**EA US Employment Index vs. Nonfarm Payrolls**

- **Index Breakdown**
  - Job Loss: 33%
  - New Job: 42%
  - Job Search: 16%
  - Others: 5%
  - Correlation: 0.54
  - Out of Sample MAE Improvement vs baseline ARIMA model: 11%

**Source:** Eagle Alpha, Google, BLS, Bloomberg

**EA US Autos Index vs. US Light Vehicle Sales**

- **Index Breakdown**
  - Auto Brands: 10%
  - Auto Retail: 5%
  - Auto Finance: 85%
  - Correlation: 0.52
  - Out of Sample MAE Improvement vs baseline ARIMA model: 9%

**Source:** Eagle Alpha, Google, BEA, Bloomberg

**EA UK Employment Index vs. Claimant Count Change**

- **Index Breakdown**
  - Job Loss: 22%
  - Job Search: 43%
  - New Job: 35%
  - Correlation: 0.68
  - Out of Sample MAE Improvement vs baseline ARIMA model: 3.5%

**Source:** Eagle Alpha, Google, ONS, Bloomberg

**EA UK Housing Index vs. RICS House Price Index**

- **Index Breakdown**
  - Mortgage: 24%
  - Property Search: 26%
  - State Agents: 25%
  - Rent: 25%
  - Correlation: 0.8
  - Out of Sample MAE Improvement vs baseline ARIMA model: 3%

**Source:** Eagle Alpha, Google, RICS, Bloomberg
- Currently: data, reports and analyst access.
- Under Construction: tool for clients.


