Predicting the Present with Google Trends

SF Fed, March 18

Hyunyoung Choi
Hal Varian
Searches for [hangover]

Which day of the week are there the most searches for [hangover]?

1: Sunday
2: Monday
3: Tuesday
4: Wednesday
5: Thursday
6: Friday
7: Saturday
Search index for [hangover]

Web Search Interest: hangover
United States, Dec 2008 - Feb 2009
Categories: Food & Drink (25-50%), Health (10-25%), Entertainment (0-10%), Local (0-10%), more...

Interest over time

Learn what these numbers mean

Embed this chart
Hangover geo

Regional interest
1. New York
2. Massachusetts
3. New Jersey
4. Pennsylvania
5. Kansas
6. Nevada
7. Wisconsin
8. Kentucky
9. Oregon
10. Texas

Search terms
Top searches
1. hangover cure
2. hangover cures
3. the hangover
4. alcohol hangover
5. cure a hangover
6. hangover remedies
7. hangover lyrics
8. cure for hangover
9. hangover food
10. hangover symptoms

Rising searches
1. hangover classic
2. best hangover cure
3. cure for hangover
4. hangover remedy
5. hangover prevention
6. no hangover
Searches for [civil war]
US searches on [civil war] + seasonal AR prediction
[civil war] and [term papers]
Predicting the Present

Economic Index

Lag Variables  Google Trends  Other Exogenous Variable

Predicting the Present with Google Trends
Thursday, April 02, 2009 at 4/02/2009 02:10:00 PM

Predicting Initial Claims for Unemployment Benefits
Wednesday, July 22, 2009 at 7/22/2009 05:00:00 PM
Proposed procedure for using Trends data

Fit the best model you can using the data you have (which may often be past values of the time series itself.)

Add Google Trends data as an additional predictor

See how the *out of sample* forecast improves using mean absolute error using a rolling window forecast.

Particularly interest in turning points since they are the hardest thing to forecast.

**Issues with Google Trends**

Mixed frequency: Trends is available daily/weekly basis while series of interest may be weekly or monthly. (This is a plus.)

Google Trends is an index: *normalized query share using broad match*

Must have at least 50 observations to appear in Google Trends due to privacy policy.

Google Trends is sampled data, and changes slightly from day to day.

Can look at session context (Apple as Food v Apple as Consumer Electronics)
Search for [apple] in context
Can also examine searches by category

Top searches: blue book, cars, kelley blue book, used cars, etc.
Unemployment
Initial claims: good leading indicator for end of recession
Web Search Volume: Welfare & Unemployment

United States, Jan 2006 - Mar 2009

All Categories > Society > Social Services > Welfare & Unemployment

Interest over time

Search terms in Welfare & Unemployment

Top searches
1. unemployment
2. social security
3. social security administration

Rising searches
1. social security locations +350%
2. mn unemployment +190%
3. unemployment benefits +130%
<table>
<thead>
<tr>
<th>Jobs</th>
<th>Welfare &amp; Unemployment</th>
<th>Recruiting &amp; Staffing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monster</td>
<td>Social Security</td>
<td>CareerBuilder</td>
</tr>
<tr>
<td>Indeed</td>
<td>Social Security Office Locations</td>
<td>Kelly Services</td>
</tr>
<tr>
<td>Jobs</td>
<td>Social Security Administration</td>
<td>Manpower</td>
</tr>
<tr>
<td>Job Search</td>
<td>Unemployment Benefits</td>
<td>Temp Agencies</td>
</tr>
<tr>
<td>Resume</td>
<td>Social Security Disability</td>
<td>Robert Half</td>
</tr>
<tr>
<td>Job Search Engines</td>
<td>Social Security Gov</td>
<td>Spherion</td>
</tr>
<tr>
<td>Linkedin</td>
<td>Unemployment Office</td>
<td>Aerotek</td>
</tr>
<tr>
<td>Hotjobs</td>
<td>Food Stamps</td>
<td>Walmart jobs</td>
</tr>
<tr>
<td>Cover Letter</td>
<td>Department of Labor</td>
<td>Appleone</td>
</tr>
</tbody>
</table>
According to the NBER, the current recession started December 2007.

National unemployment rate passed 5% in mid 2008 and search queries on [Welfare and Unemployment] increased at same time.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>US Dept of Labor</td>
<td>Initial Claims(K)</td>
<td>625</td>
<td>605</td>
<td>612</td>
<td>630</td>
<td>617</td>
</tr>
<tr>
<td></td>
<td>Continued Claims(MN)</td>
<td>8.84</td>
<td>6.71</td>
<td>6.76</td>
<td>6.72</td>
<td>6.90</td>
</tr>
<tr>
<td>Google Trends</td>
<td>Jobs</td>
<td>-1%</td>
<td>-1%</td>
<td>0%</td>
<td>-1%</td>
<td>-2%</td>
</tr>
<tr>
<td></td>
<td>Recruitment &amp; Staffing</td>
<td>-33%</td>
<td>-30%</td>
<td>-27%</td>
<td>-28%</td>
<td>-29%</td>
</tr>
<tr>
<td></td>
<td>Welfare &amp; Unemployment</td>
<td>38%</td>
<td>41%</td>
<td>39%</td>
<td>43%</td>
<td>44%</td>
</tr>
</tbody>
</table>
Model

Reference AR(1) Model

\[ \log(y_t) = \text{Intercept} + \phi \log(y_{t-1}) + \epsilon_t \]

AR(1) Model With Google Trends

\[ \log(y_t) = \text{Intercept} + \alpha \text{Jobs}_t + \beta \text{Welfare}_t + \phi \log(y_{t-1}) + \epsilon_t \]

<table>
<thead>
<tr>
<th></th>
<th>Baseline Model</th>
<th>Alternative Model</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>(\phi)</td>
<td>(\sigma)</td>
<td>MAE</td>
<td>Intercept</td>
<td>(\phi)</td>
<td>Jobs</td>
<td>Welfare</td>
</tr>
<tr>
<td><strong>LT</strong></td>
<td>Est</td>
<td>0.1269</td>
<td>0.9902</td>
<td>0.0443</td>
<td>3.24%</td>
<td>1.6498</td>
<td>0.8727</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.1618</td>
<td>0.0126</td>
<td>0.0163</td>
<td></td>
<td>0.3754</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td><strong>ST</strong></td>
<td>Est</td>
<td>0.2174</td>
<td>0.9839</td>
<td>0.0432</td>
<td>3.10%</td>
<td>1.792</td>
<td>0.8632</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.2632</td>
<td>0.0202</td>
<td>0.0163</td>
<td></td>
<td>0.5541</td>
<td>0.0427</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

Model fit improved significantly – smaller standard deviation, high log likelihood and smaller AIC

Initial claims are positively correlated with searches on Jobs and Welfare.
With Google Trends, the out-of-sample prediction MAE decreases by 15.74%.

- Prediction with rolling window from 1/18/2009 to 6/27/2009 (24 weeks)
Seasonally unadjusted data

[file for unemployment] query
MAE goes down by 15% overall
Structural models

Can use your favorite forecasting model, e.g., Kalman filters
   Attractive since they are adaptive
   BSM = “basic structural model” = trend + seasonal + residual
   BSM + Kalman regression seems to work well

Example
   Monthly housing sales from Census
   Estimate BSM for 2004-2009, forecast 2010, with and without query data
   Rolling 1-step ahead forecast MAE goes from 16% to 12%, a decline of 25%
Housing sales and predicted
Model Selection
For US unemployment rate and initial claims

Hyunyoung Choi
Hal Varian
Nowcasting work by LSE/Oxford group: Jennifer Castle, Jurgen Doornick, David Hendry

Contemporaneous forecasting – as in “predicting the present”

Updating forecasts as new information becomes available – mixed frequent estimation

**Variable selection** – which predictors should be used out of a rich set of predictors?

Variable selection (Castle examines 21 different methods)

- Judgment based on model, implicit or explicit
- Penalized fit (AIC, BIC, various overfitting corrections), Bayesian selection
- Machine learning techniques (lasso et al)
- Significance testing – retain variables that are significantly different from 0
  - Stepwise regression
  - Gets (Generalized to Specific) from LSE/Oxford team

**Applications**

- Genetic markings, econometrics, etc.
Applications to unemployment forecasting

Find 1000 queries that have highest contemporaneous correlation with unemployment rate [not initial claims]

Use some variable selection methods to build a forecasting model, see what performs best

Important economic fact: Unemployment rate among young men in July 2009 was 19.7%
Four stages of unemployment searches

Labor Market Related:
companies that are hiring, jobs classifieds, who's hiring, department of labor, working in oregon, unemployment eligibility, file for unemployment, go2ui, unemployment, unemployment claim, unemployment benefits, unemployment compensation, unemployment office
Four stages of unemployment searches

Labor Market Related:
companies that are hiring, jobs classifieds, who's hiring, department of labor,
working in oregon, unemployment eligibility, file for unemployment, go2ui,
unemployment, unemployment claim, unemployment benefits, unemployment
compensation, unemployment office

New Tech Trends:
linux netbook, top netbooks, ipod digitizer, free apps, free ringtone downloads for
cell phones, good ipod apps, good ipod touch apps, good itouch apps
Four stages of unemployment searches

Labor Market Related:
companies that are hiring, jobs classifieds, who's hiring, department of labor, working in oregon, unemployment eligibility, file for unemployment, go2ui, unemployment, unemployment claim, unemployment benefits, unemployment compensation, unemployment office

New Tech Trends:
linux netbook, top netbooks, ipod digitizer, free apps, free ringtone downloads for cell phones, good ipod apps, good ipod touch apps, good itouch apps

Entertainment:
what are some good screamo bands, atlanta sports cards, quotes and sayings, guitar scales beginner, poker hands order, home workout routines, sweepstakes and contests, american film institute top 100 films, best movies of the 90's, movie theater locator, where can you download free music, ameristar casino st charles
Four stages of unemployment searches

Labor Market Related:
companies that are hiring, jobs classifieds, who's hiring, department of labor, working in oregon, unemployment eligibility, file for unemployment, go2ui, unemployment, unemployment claim, unemployment benefits, unemployment compensation, unemployment office

New Tech Trends:
linux netbook, top netbooks, ipod digitizer, free apps, free ringtone downloads for cell phones, good ipod apps, good ipod touch apps, good itouch apps

Entertainment:
what are some good screamo bands, atlanta sports cards, quotes and sayings, guitar scales beginner, poker hands order, home workout routines, sweepstakes and contests, american film institute top 100 films, best movies of the 90's, movie theater locator, where can you download free music, ameristar casino st charles

Adult Content:
adult video, freepornhub, anchor babes, kissing games, porn tube, jailbait teen
Correlation between Unemployment Rate and Aggregate over the Keywords Group

- Entertainment (rho = 0.96)
- Adult (rho = 0.94)
- Labor (rho = 0.96)
- Software (rho = 0.94)
- New Tech (rho = 0.94)
- Game (rho = 0.94)
- Health (rho = 0.96)
- Lifestyles (rho = 0.96)
- Computer (rho = 0.97)
Predicting *Unemployment* with selected queries

Much smaller prediction error with labor market predictors. *Gets* model selection is more effective with top-60 model than stepwise regression.

Top 60 predictors with high correlation includes:
- error code 0 (rho = 0.97)
- afk acronym (rho = 0.97)
- amateur xxx (rho = 0.97)
- austin pets alive (rho = 0.97)
- inetinfo-exe what is (rho = 0.96)
- washington state unemployment (rho = 0.96)
- hacker news (rho = 0.96)
- colorado unemployment (rho = 0.96)
- secure server (rho = 0.96)