

# Predicting the Present with Google Trends

---

**SF Fed, March 18**

**Hyunyoung Choi**

**Hal Varian**

## **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]



Google Insights for Search beta

hal@google.com | [My Account](#) | [Help](#) | [Sign out](#) | [Download as CSV](#) | [English \(US\)](#) ▼

Compare by	Search terms	Filter
<input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	Tip: Use quotation marks to match an exact phrase. ("table tennis") <input type="text" value="hangover"/> <a href="#">+ Add search term</a>	Web Search United States   All subregions   All metros Dec 2008 - Feb 2009 <a href="#">Reset</a> All Categories <input type="button" value="Search"/>

## 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...](#)

Totals ?

hangover  15

## Interest over time

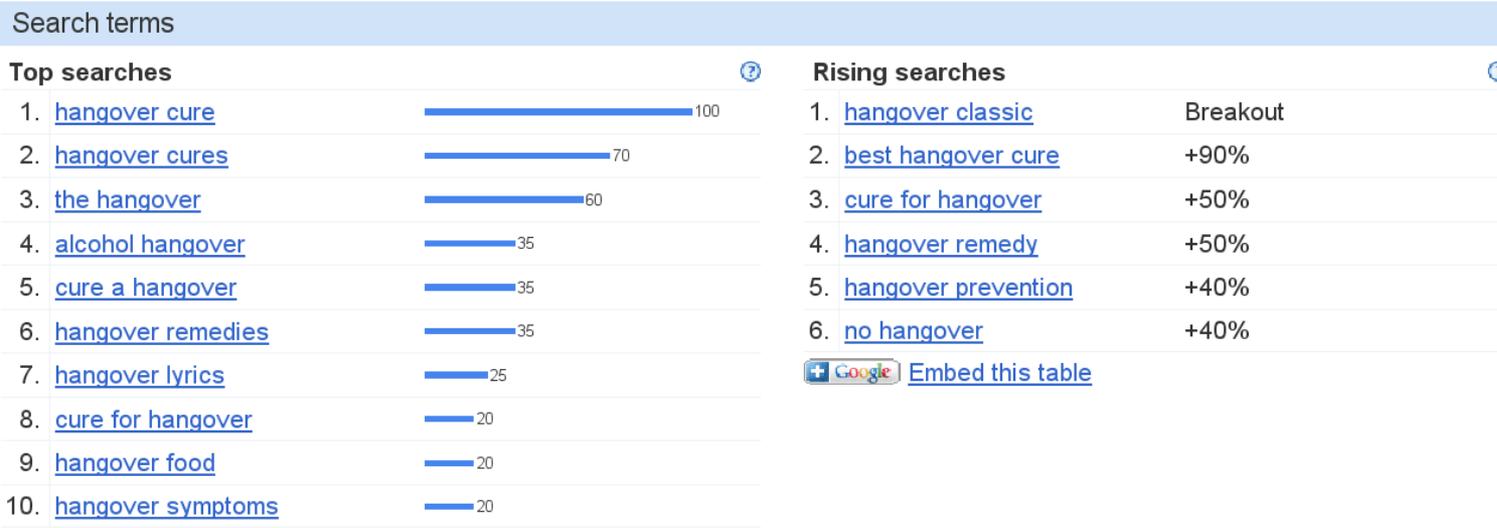
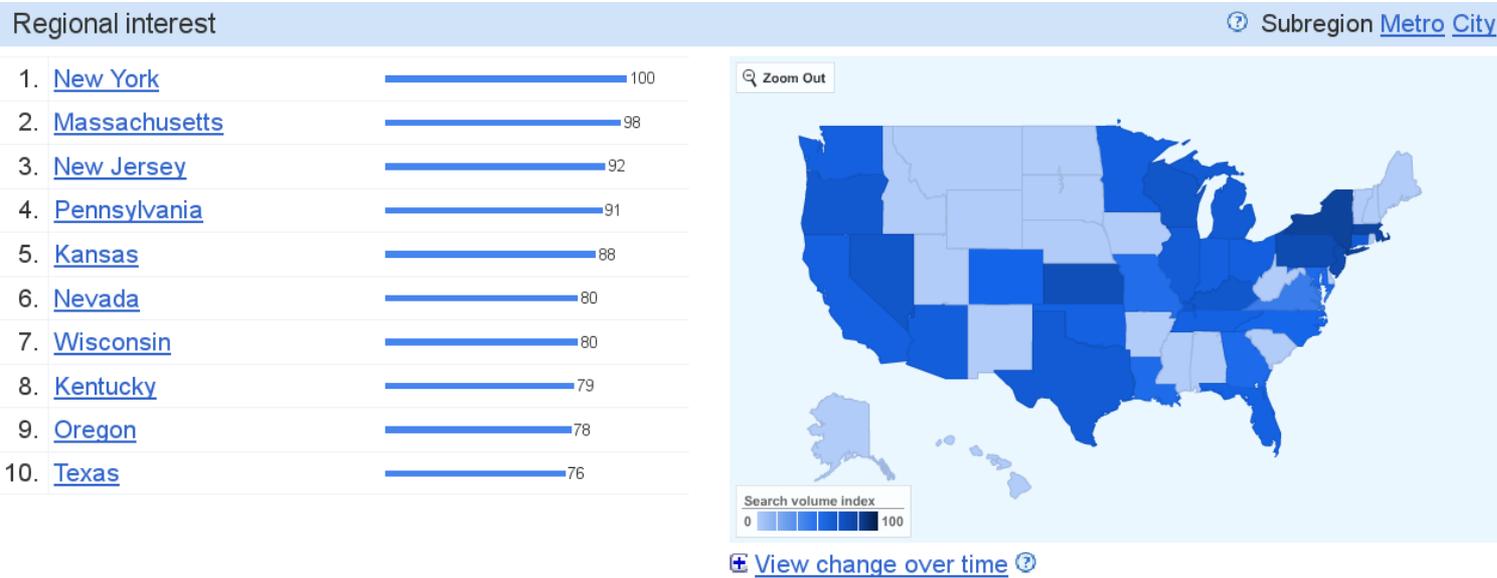
Forecast ?  News headlines

[Learn what these numbers mean](#)



 [Embed this chart](#)

# Hangover geo



# Hangover-vodka time series



Google Insights for Search  
beta

hal@google.com | [My Account](#) | [Help](#) | [Sign out](#) | [Download as CSV](#) | [English \(US\)](#) ▼

Compare by	Search terms	Filter
<input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	Tip: Use quotation marks to match an exact phrase. ("table tennis") <input type="text" value="hangover"/> <input type="text" value="vodka"/> <a href="#">+ Add search term</a>	Web Search United States   All subregions   All metros Dec 2008 - Feb 2009 <a href="#">Reset</a> All Categories <input type="button" value="Search"/>

## Web Search Interest: hangover, vodka

United States, Dec 2008 - Feb 2009

Categories: [Food & Drink](#), [Health](#), [Entertainment](#), [Local](#), [Reference](#)

### Totals

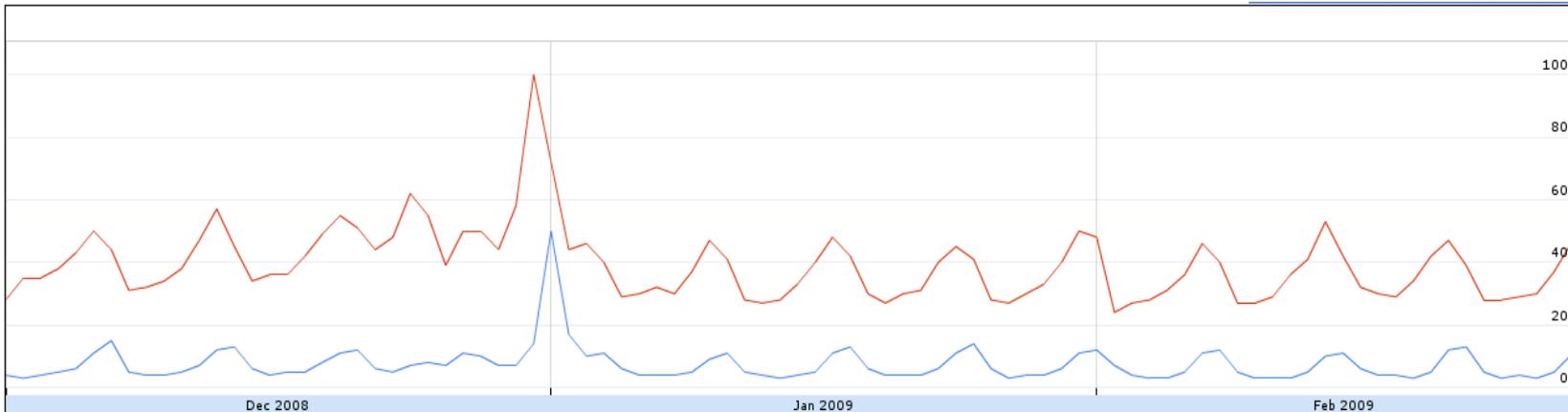
hangover  7

vodka  39

### Interest over time

Forecast  News headlines

[Learn what these numbers mean](#)



[Embed this chart](#)

# Searches for [civil war]



<b>Compare by</b> <input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	<b>Search terms</b> Tip: Use a comma as shorthand to add comparison items. (tennis, squash) ● civil war <a href="#">+ Add search term</a>	<b>Filter</b> Web Search ▼ United States ▼ All subregions ▼ All metros ▼ 2004 - present ▼ All Categories ▼
--	--	--

## Web Search Interest: civil war

United States, 2004 - present

Categories: [Arts & Humanities \(50-75%\)](#), [Society \(10-25%\)](#), [Local \(0-10%\)](#), [Entertainment \(0-10%\)](#), [more...](#)

**Totals** ⓘ  
civil war  35

### Interest over time

Forecast  News headlines

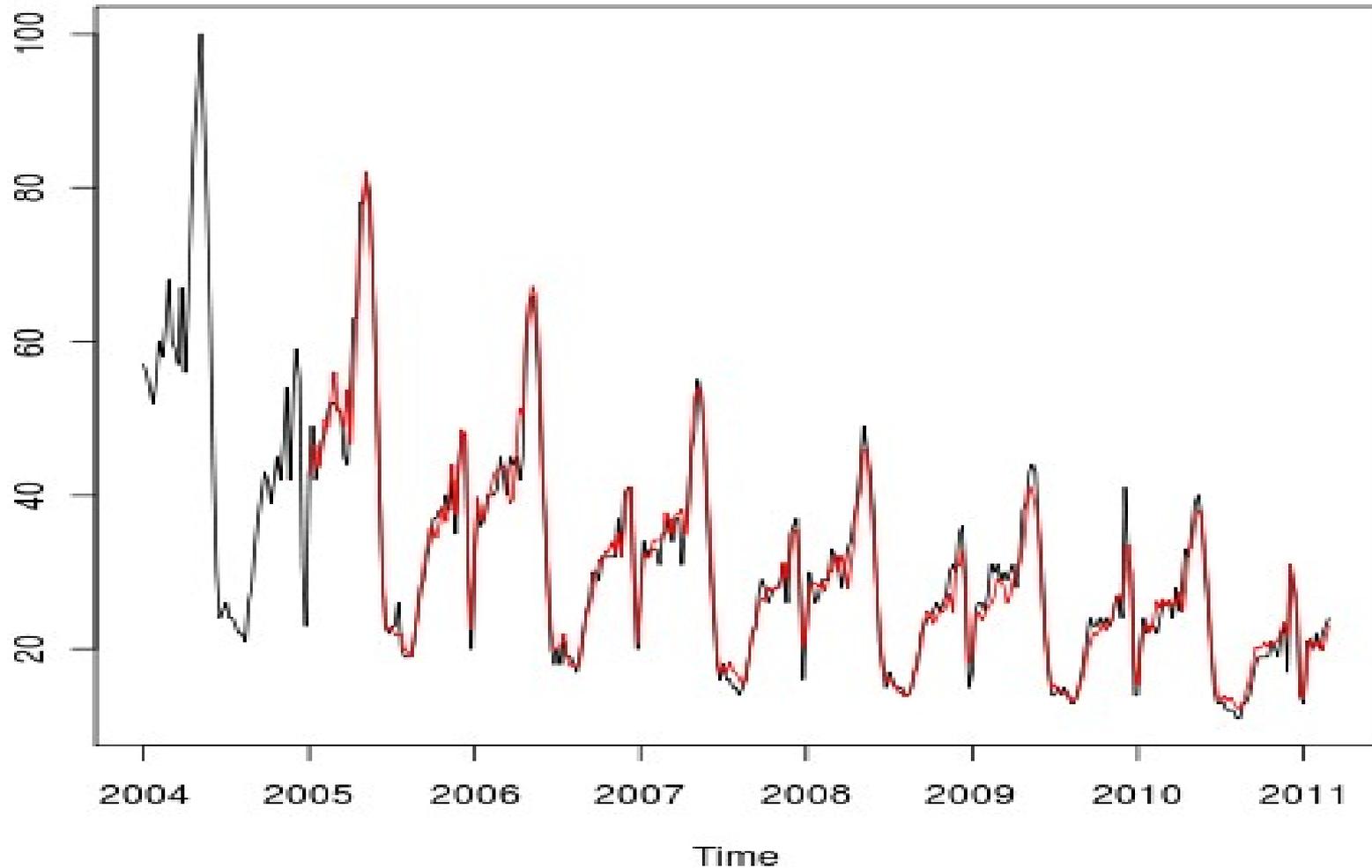
[Learn what these numbers mean](#)



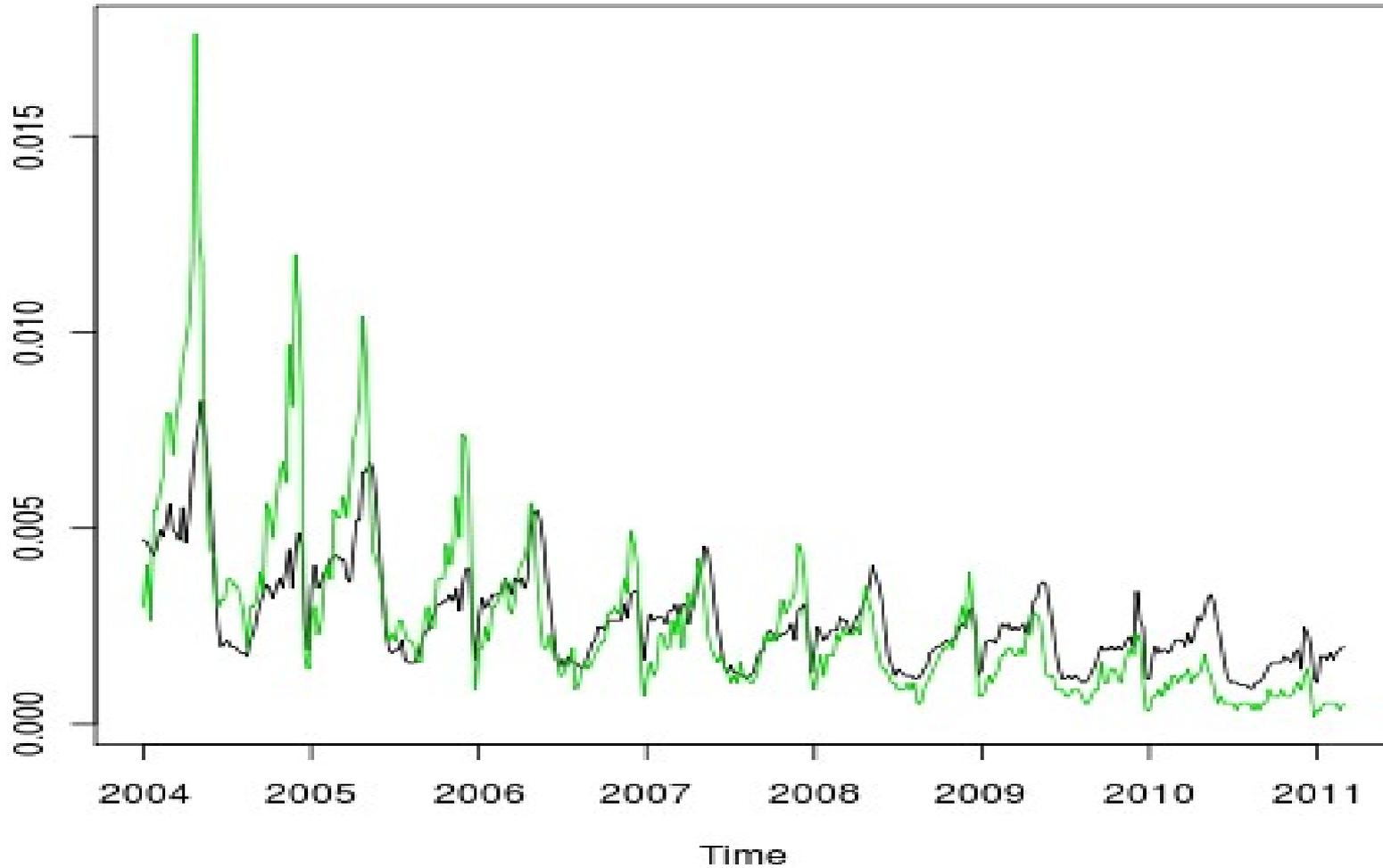
\* The last value on the graph is based on partial data and may change. [Learn more](#)

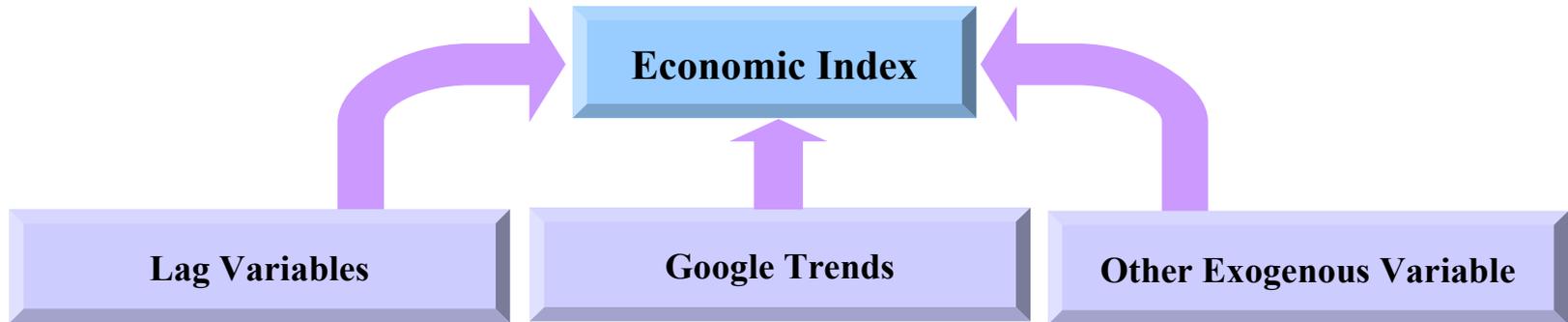
[Embed this chart](#)

US searches on [civil war]+ seasonal AR prediction



[civil war] and [term papers]





Research Blog

## [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.)

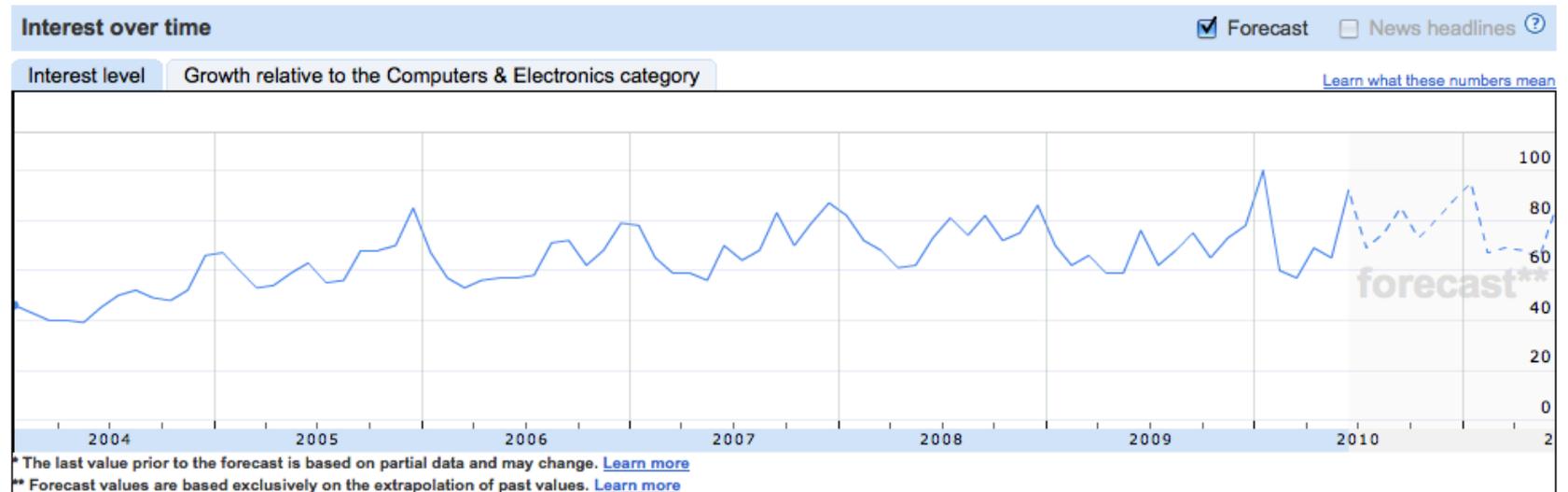
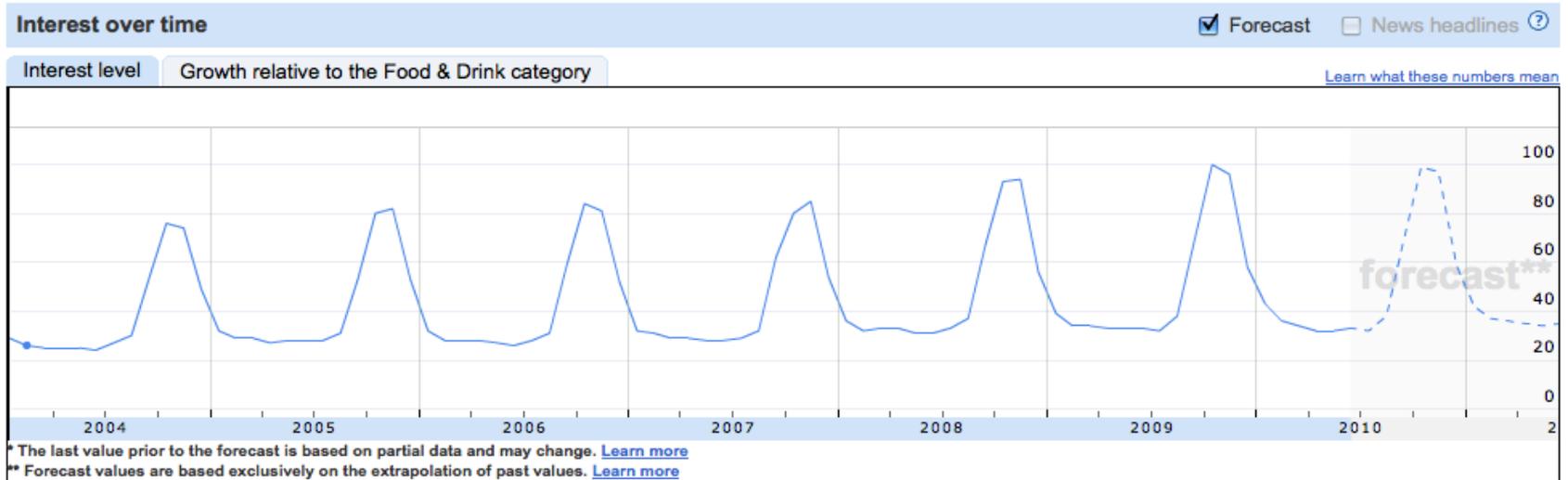
Google Trends is an index: *normalized query share using broad match*

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

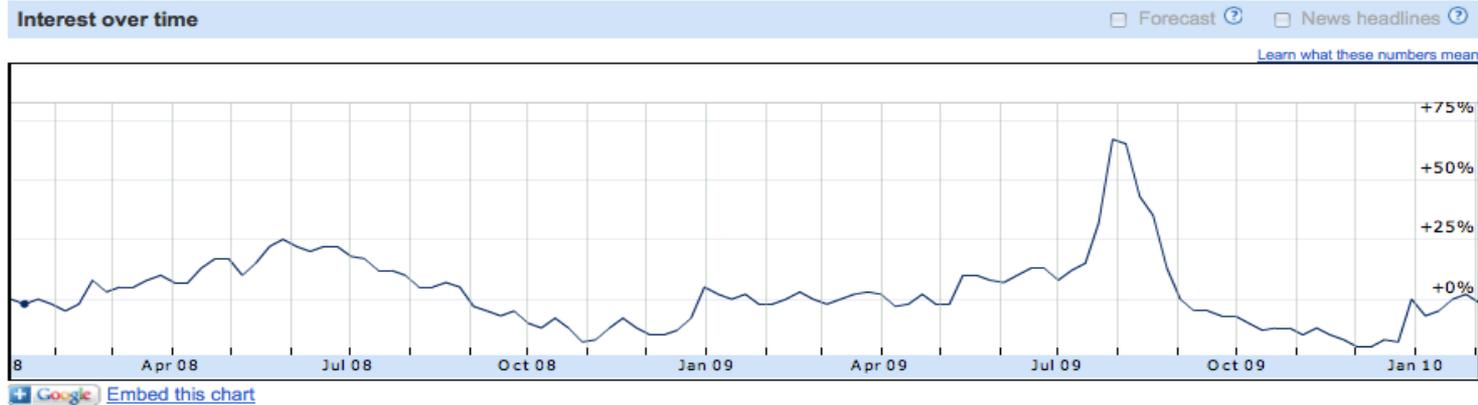


<b>Compare by</b> <input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	<b>Search terms</b> Tip: Use the plus sign to indicate OR. (tennis + squash) <input type="text" value="All search terms"/> <a href="#">+ Add search term</a>	<b>Filter</b> Web Search United States   All subregions   All metros Jan 2008 - Jan 2010 <a href="#">Reset</a> Vehicle Shopping <input type="button" value="Search"/>
--	---	---

## Web Search Interest

United States, Jan 2008 - Jan 2010

[All Categories](#) > [Automotive](#) > Vehicle Shopping



Top searches: blue book, cars, kelley blue book, used cars, etc.

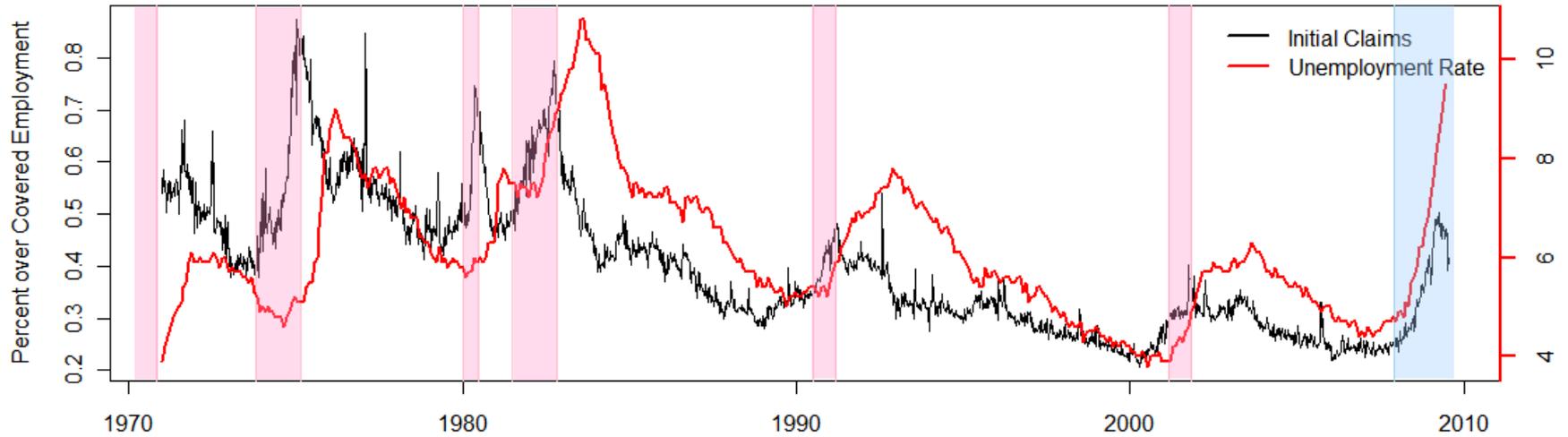
# UNEMPLOYMENT



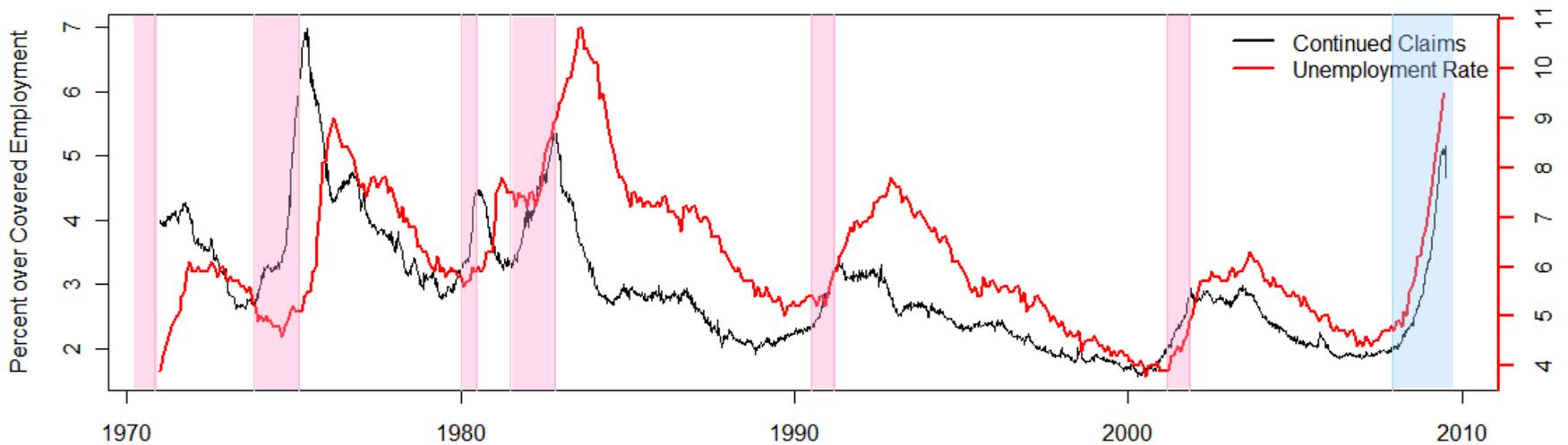
# Initial claims: good leading indicator for end of recession



### Initial Claims & Unemployment Rate



### Continued Claims & Unemployment Rate



# Google Trends data [Search Insights screenshot]



[My Account](#) | [Help](#) | [Sign out](#) | [Download as CSV](#)



<b>Compare by</b>	<b>Search terms</b>	<b>Filter</b>
<input checked="" type="radio"/> Search terms <input type="radio"/> Locations <input type="radio"/> Time Ranges	Tip: Use quotation marks to match an exact phrase. ("table tennis") <input type="text" value="All search terms"/> <a href="#">+ Add search term</a>	Web Search United States   All subregions   All metros Jan 2006 - Mar 2009 <a href="#">Reset</a> Welfare & Unemployment
		<input type="button" value="Search"/>

## Web Search Volume: Welfare & Unemployment

United States, Jan 2006 - Mar 2009

[All Categories](#) > [Society](#) > [Social Services](#) > Welfare & Unemployment

### Interest over time

News is unavailable for specific categories.

[Learn what these numbers mean](#)



### Search terms in Welfare & Unemployment *United States, Jan 2006 - Mar 2009*

#### Top searches

1.	<a href="#">unemployment</a>	100
2.	<a href="#">social security</a>	85
3.	<a href="#">social security administration</a>	25

#### Rising searches

1.	<a href="#">social security locations</a>	+350%
2.	<a href="#">mn unemployment</a>	+190%
3.	<a href="#">unemployment benefits</a>	+130%

# Keywords Examples



## Jobs

- Monster
- Indeed
- Jobs
- Job Search
- Resume
- Job Search Engines
- LinkedIn
- Hotjobs
- Cover Letter

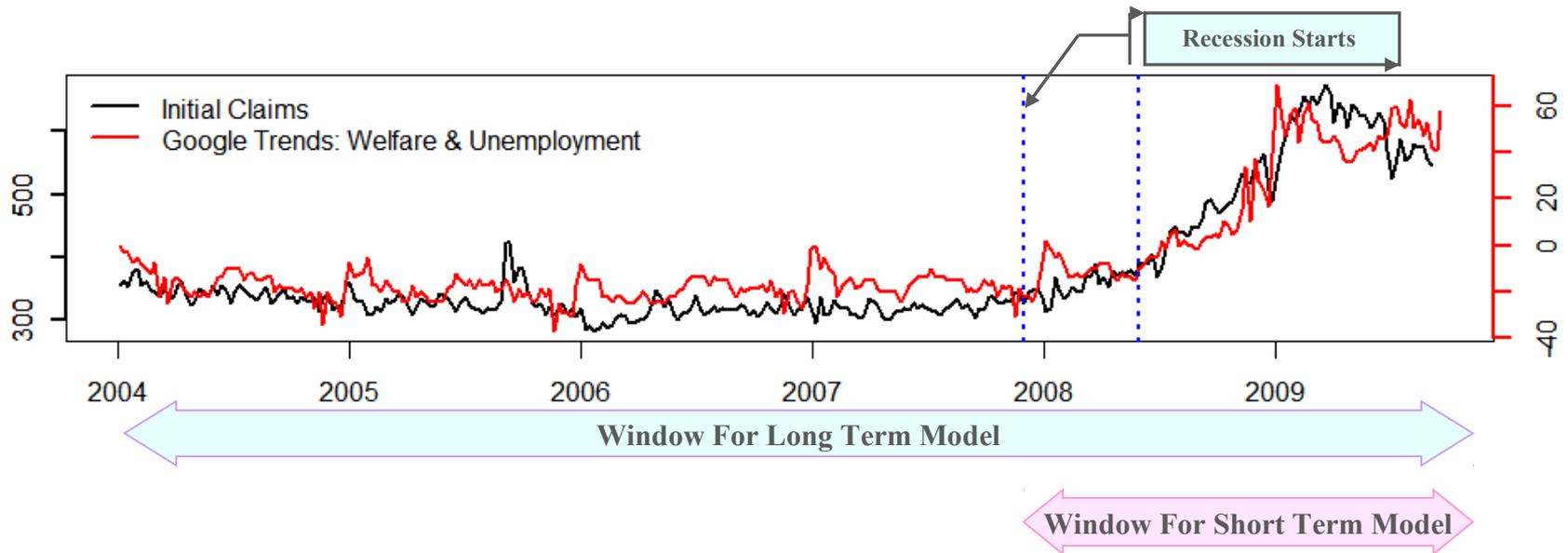
## Welfare & Unemployment

- Social Security
- Social Security Office Locations
- Social Security Administration
- Unemployment Benefits
- Social Security Disability
- Social Security Gov
- Unemployment Office
- Food Stamps
- Department of Labor

## Recruiting & Staffing

- CareerBuilder
- Kelly Services
- Manpower
- Temp Agencies
- Robert Half
- Spherion
- Aerotek
- Walmart jobs
- Appleone

# Initial Claims vs. Google Trends



- 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.

	Week	5/24/09 - 5/30/09	5/31/09 - 6/6/09	6/7/09 - 6/13/09	6/14/09 - 6/20/09	6/21/09 - 6/27/09	6/28/09 - 7/4/09
US Dept of Labor	Initial Claims(K)	625	605	612	630	617	Release at 7/9/09
	Continued Claims(MN)	8.84	6.71	6.76	6.72	6.90	
Google Trends	Jobs	-1%	-1%	0%	-1%	-2%	-3%
	Recruitment & Staffing	-33%	-30%	-27%	-28%	-29%	-37%
	Welfare & Unemployment	38%	41%	39%	43%	44%	44%

## Reference AR(1) Model

$$\log(y_t) = \text{Intercept} + \phi \log(y_{t-1}) + e_t$$

## AR(1) Model With Google Trends

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

		Baseline Model				Alternative Model					
		Intercept	$\phi$	$\sigma$	MAE	Intercept	$\phi$	Jobs	Welfare	$\sigma$	MAE
LT	Est	0.1269	0.9902	0.0443	3.24%	1.6498	0.8727		0.0014	0.0429	2.73%
	SE	0.1618	0.0126			0.3754	0.029		0.0003		
ST	Est	0.2174	0.9839	0.0432	3.10%	1.792	0.8632	0.0014	0.0010	0.0398	2.70%
	SE	0.2632	0.0202			0.5541	0.0427	0.0006	0.0004		

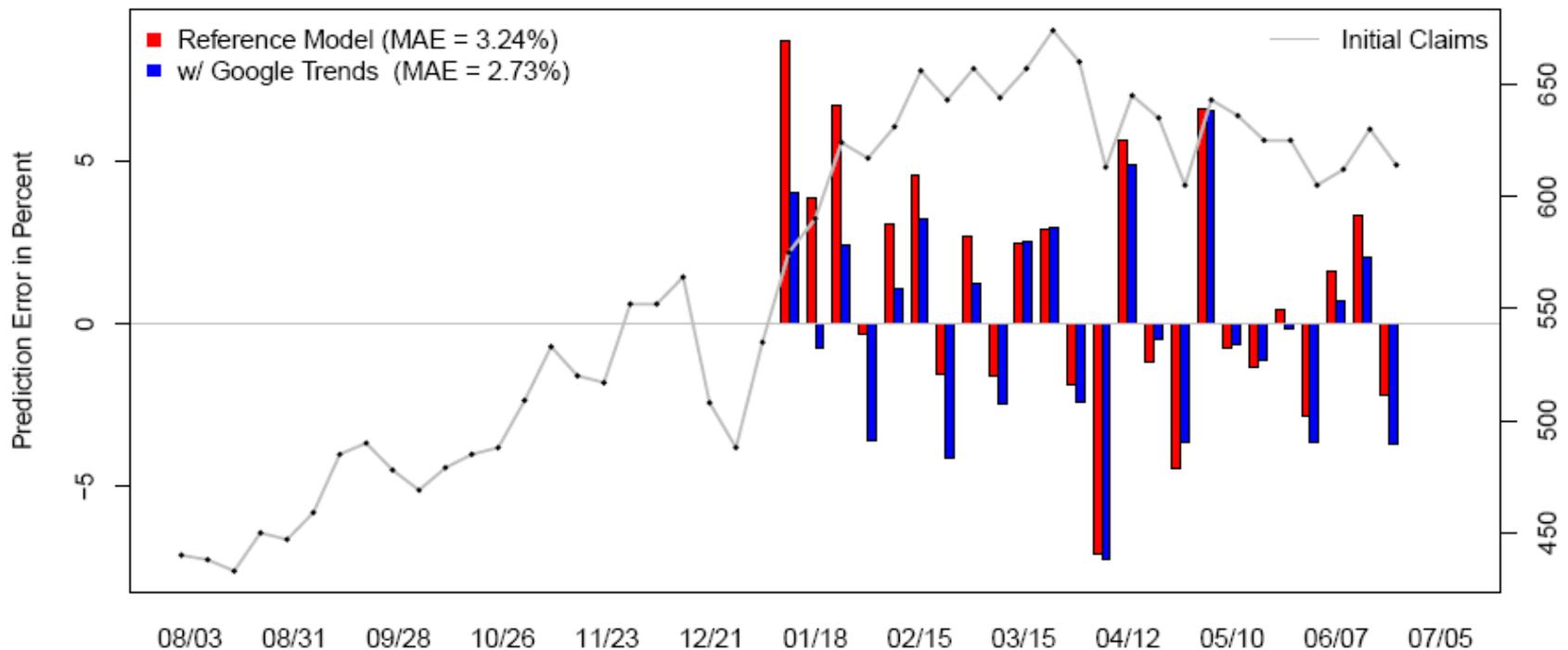
Model fit improved significantly – smaller standard deviation, high log likelihood and smaller AIC

Initial claims are positively correlated with searches on Jobs and Welfare.

# Long Term Model: Prediction Comparison with MAE

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)

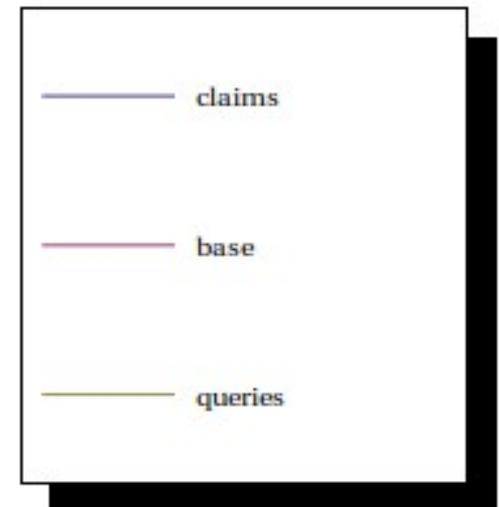
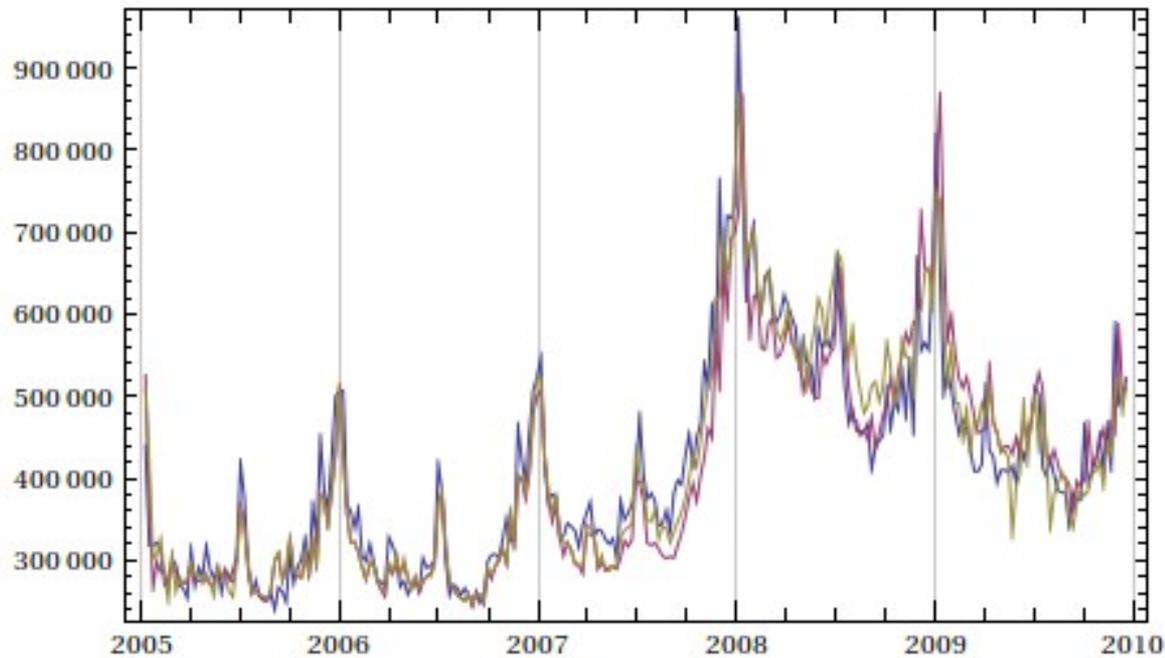


# Seasonally unadjusted data



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

Rolling forecasts



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

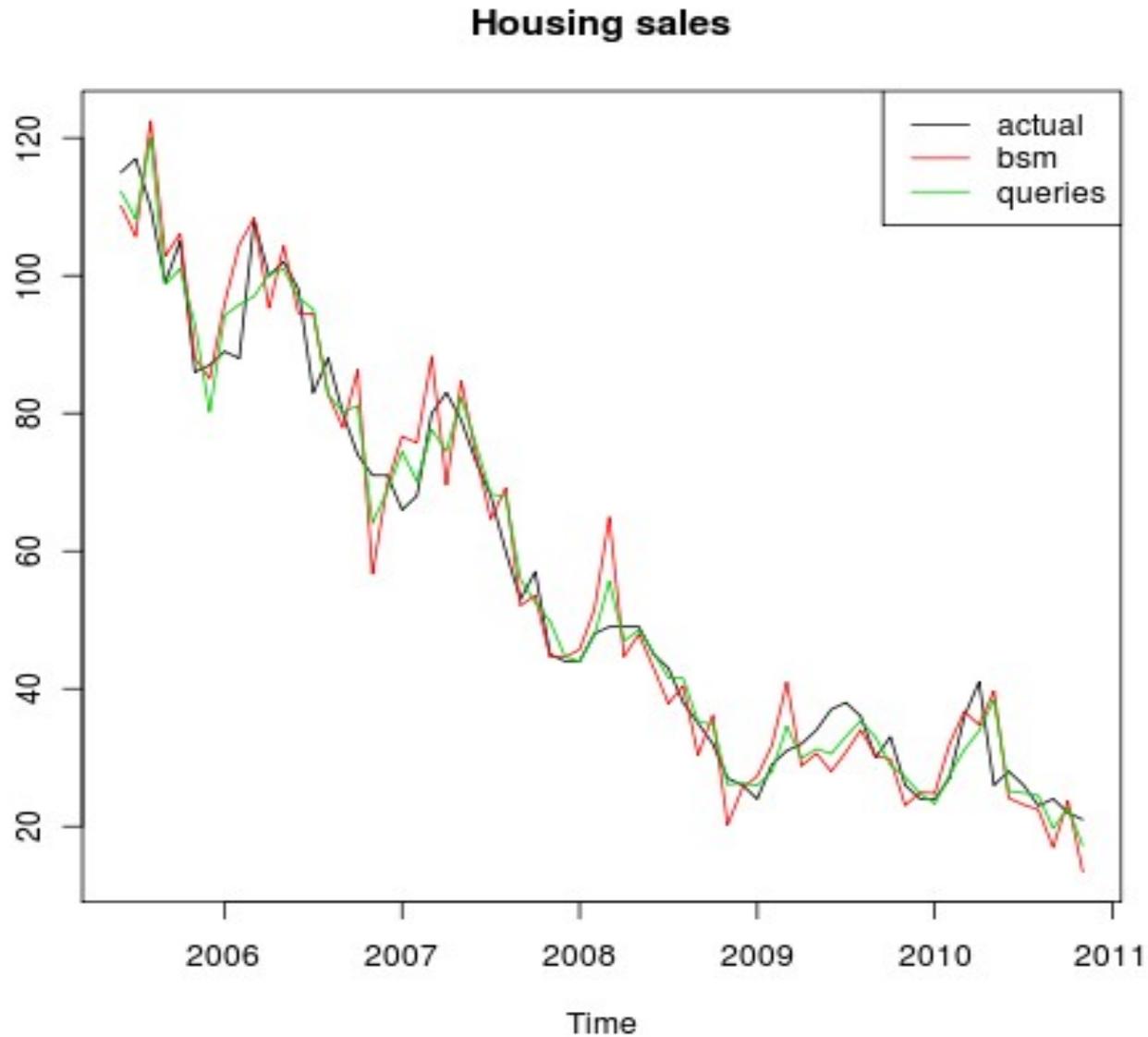
Example

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.

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

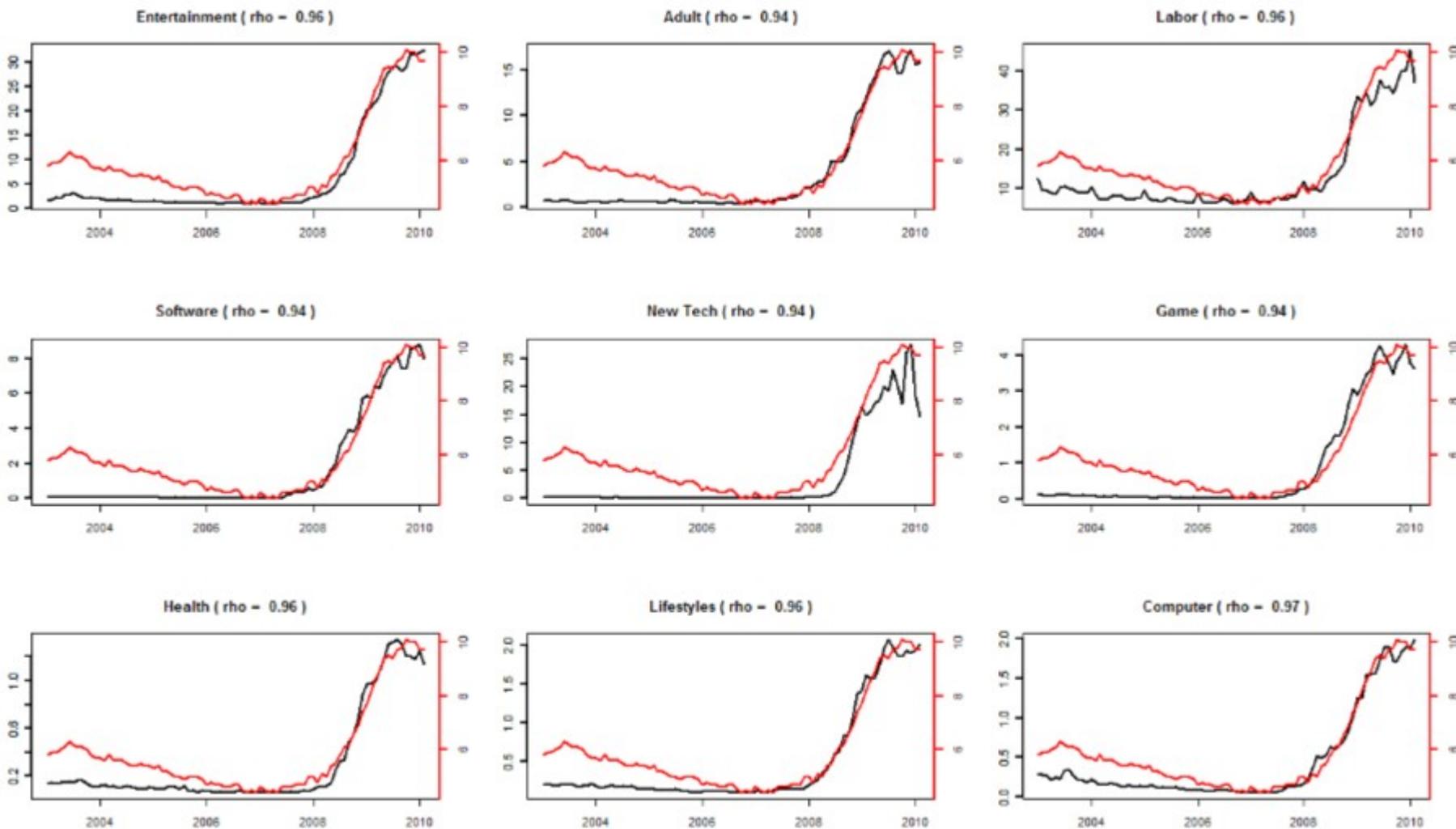
## Entertainment:

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

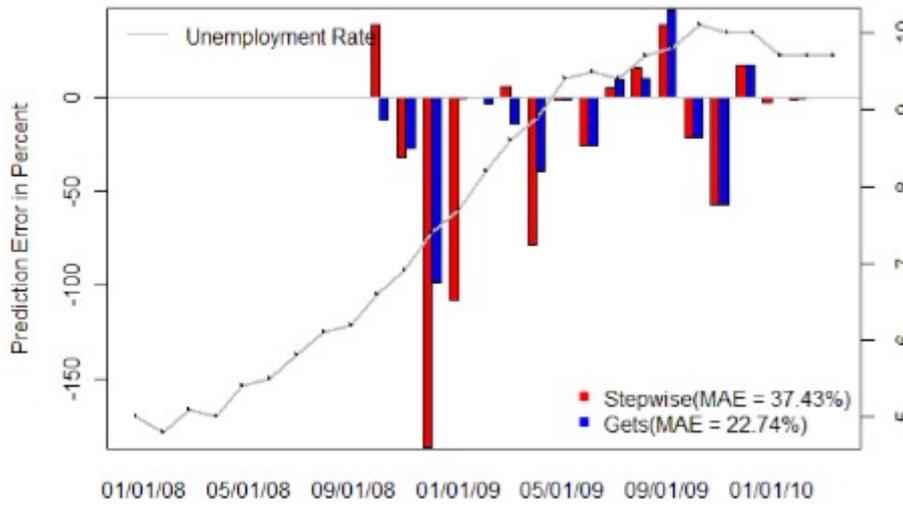
# Correlation between Unemployment Rate and Aggregate over the Keywords Group



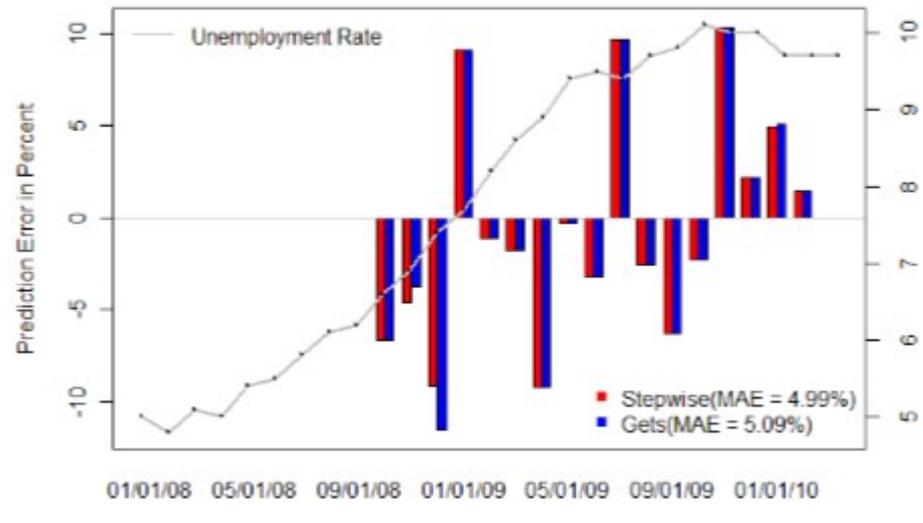
# Predicting *Unemployment* with selected queries



Starting w/ Top 60 predictors with high correlation



Starting w/ 29 labor market related predictors



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( $\rho = 0.96$ )