Predicting the Present with Bayesian Structural Time Series

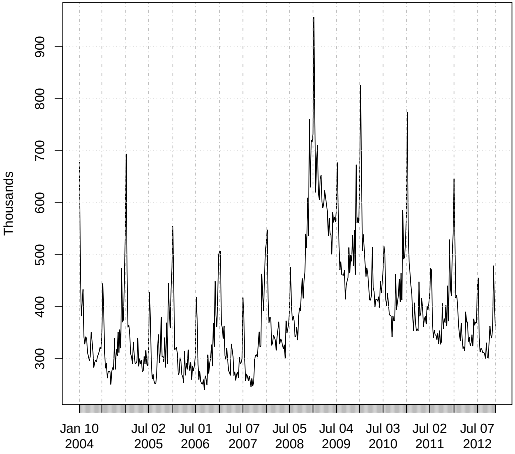
Steven L. Scott
Hal R. Varian

Google

November 22, 2013

Nowcasting

Maintaining “real time” estimates of infrequently observed time series.



- US weekly initial claims for unemployment.
- Recession leading indicator.
- Can we learn this week’s number before it is released?
- We’d need a real time signal correlated with the outcome.

Outline

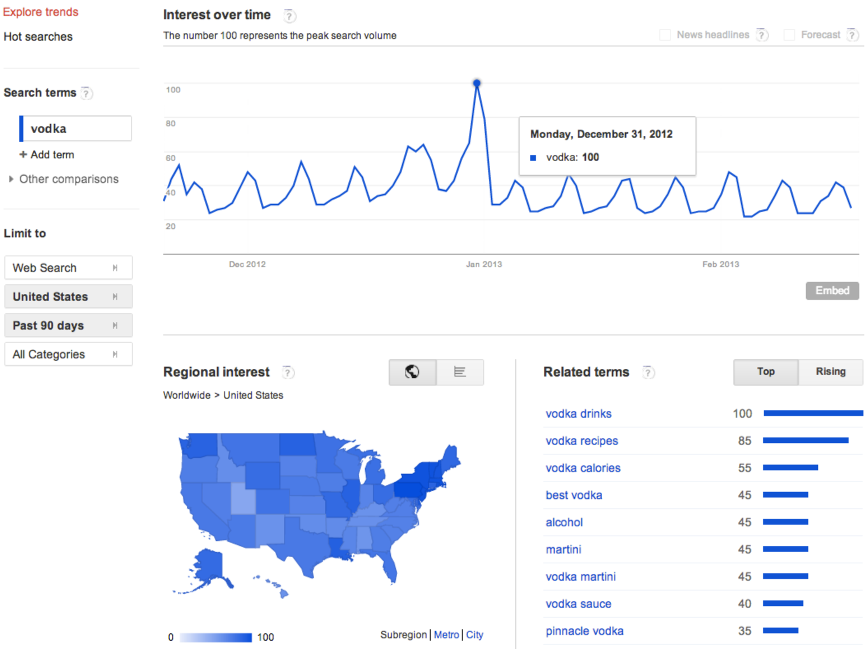
Google Trends and Google Correlate

Bayesian structural time series (with sparse regression)

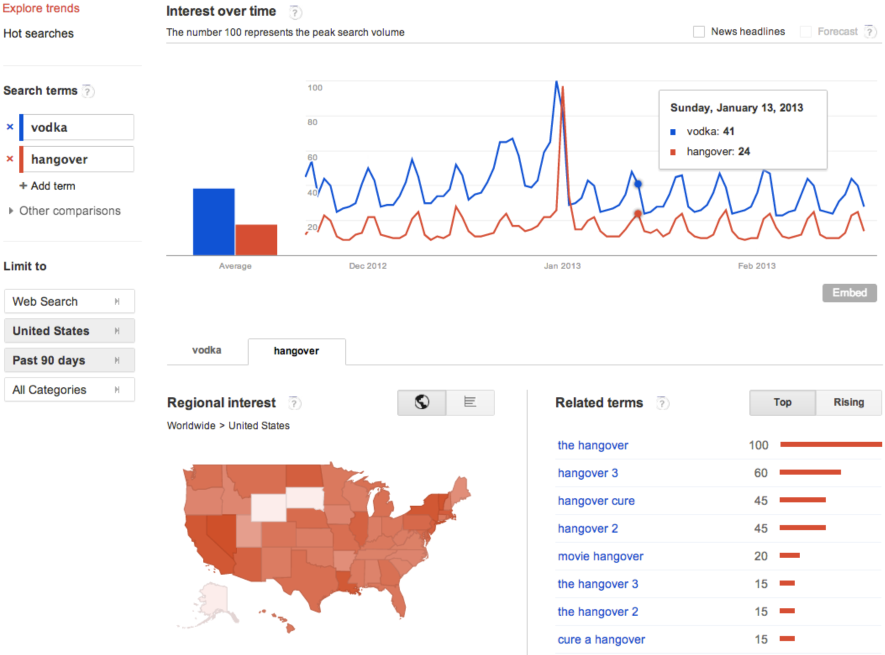
Examples
- Initial Claims
- Retail Sales

Conclusions

Google searches are a real time indicator of public interest

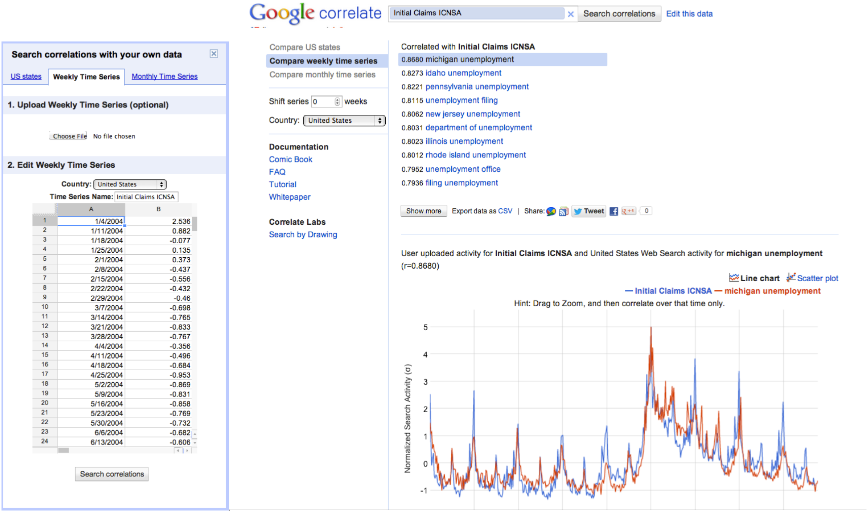


Google searches are a real time indicator of public interest



Individual search queries

Google correlate can provide the most highly correlated individual queries (up to 100)



	A	B
1	1/4/2004	2.536
2	1/11/2004	0.882
3	1/18/2004	-0.077
4	1/25/2004	0.135
5	2/1/2004	0.373
6	2/8/2004	-0.437
7	2/15/2004	-0.556
8	2/22/2004	-0.432
9	2/29/2004	-0.46
10	3/7/2004	-0.698
11	3/14/2004	-0.765
12	3/21/2004	-0.833
13	3/28/2004	-0.767
14	4/4/2004	-0.356
15	4/11/2004	-0.496
16	4/18/2004	-0.684
17	4/25/2004	-0.953
18	5/2/2004	-0.869
19	5/9/2004	-0.831
20	5/16/2004	-0.858
21	5/23/2004	-0.769
22	5/30/2004	-0.732
23	6/6/2004	-0.682
24	6/13/2004	-0.606

Google

Outline

Google Trends and Google Correlate

Bayesian structural time series (with sparse regression)

Examples

Conclusions

Structural time series models

State space form

There are two pieces to a structural time series model

Observation equation

$$y_t = Z_t^T \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, H_t)$$

- y_t is the observed data at time t.
- Z_t and H_t are structural parameters (partly known).
- α_t is a vector of latent variables called the "state".

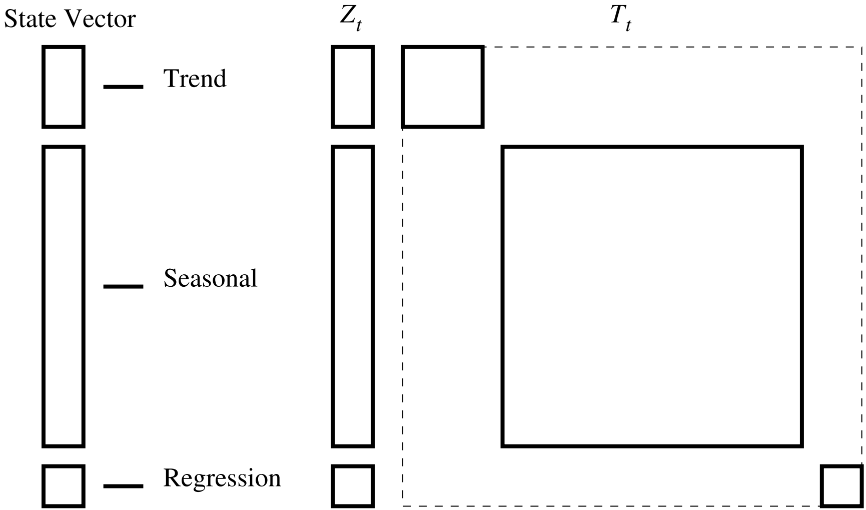
Transition equation

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \qquad \eta_t \sim \mathcal{N}(0, Q_t)$$

- T_t, R_t, and Q_t are structural parameters (partly known).
- η_t may be of lower dimension that α_t.

Structural time series models are modular

Add your favorite trend, seasonal, regression, holiday, etc. models to the mix



A good default model

The model with S seasons can be written

$$y_t = \underbrace{\mu_t}_{\text{trend}} + \underbrace{\gamma_t}_{\text{seasonal}} + \underbrace{\beta^T \mathbf{x}_t}_{\text{regression}} + \epsilon_t$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t$$

$$\delta_t = \delta_{t-1} + v_t$$

$$\gamma_t = -\sum_{s=1}^{S-1} \gamma_{t-s} + w_t$$

This is the "basic structural model" with an added regression effect.

- Trend: "level" μ_t + "slope" δ_t.
- Seasonal: $S - 1$ dummy variables with time varying coefficients. Sums to zero in expectation.

Google

MCMC

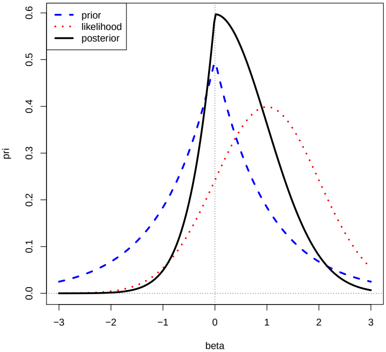
- The model parameters are $\theta = \{\sigma_\epsilon, \sigma_u, \sigma_v, \sigma_w, \beta\}$.
- The state is $\alpha = \{\alpha_1, \ldots, \alpha_n\}$.

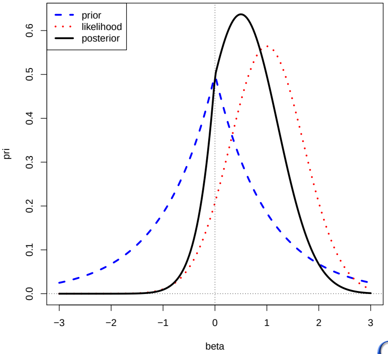
- MCMC algorithm:
  - Draw α given $\mathbf{y}$, θ
    - Kalman filter "forward filter - backward sampler" [Carter and Kohn(1994)], [Frühwirth-Schnatter(1995)], [de Jong and Shepard(1995)], [Durbin and Koopman(2002)].
    - Draws α directly
  - Draw θ given α.
    - Given α, then $[\sigma_u], [\sigma_v], [\sigma_w], [\beta, \sigma_\epsilon]$ are conditionally independent.
    - Independent priors on the time series σ's. Boring.
    - "Spike and slab" prior on β.

The “lasso prior” is not sparse

It induces sparsity at the mode, but not in the posterior distribution

$$p(\beta) \propto \exp\left(-\sum_j |\beta_j|\right)$$





Spike and slab priors

[George and McCulloch (1997)]

- We think most elements of β are zero.
- Let $\gamma_j = 1$ if $\beta_j \neq 0$ and $\gamma_j = 0$ if $\beta_j = 0$.

$$\gamma = \boxed{1}\boxed{0}\boxed{0}\boxed{1} \cdots \boxed{1}\boxed{0}\boxed{0}$$

- Now factor the prior distribution

$$p(\beta, \gamma, \sigma^{-2}) = p(\beta_\gamma | \gamma, \sigma^2) p(\sigma^2 | \gamma) p(\gamma)$$

$$\gamma \sim \prod_j \pi_j^{\gamma_j} (1 - \pi_j)^{1-\gamma_j} \qquad \text{"Spike"}$$

$$\beta_\gamma | \gamma, \sigma^2 \sim \mathcal{N}\left(b_\gamma, \sigma^2 \left(\Omega_\gamma^{-1}\right)^{-1}\right) \qquad \text{"Slab"}$$

$$\frac{1}{\sigma^2} \sim \Gamma\left(\frac{df}{2}, \frac{ss}{2}\right) \qquad \text{does not depend on } \gamma$$

Prior elicitation

$$\pi_j = \text{“expected model size”} \; / \; \text{number of predictors}$$
$$b = 0 \text{ (vector)}$$
$$\Omega^{-1} = \kappa\{\alpha \mathbf{X}^T\mathbf{X} + (1-\alpha)\text{diag}\mathbf{X}^T\mathbf{X}\}/n$$
$$ss/df = (1 - R^2_{\text{expected}})s_y^2$$
$$df = 1$$

- The Ω^{-1} expression is κ observations worth of prior information.
- It can help to average Ω^{-1} with its diagonal.
- Prior elicitation is 4 numbers: expected model size, expected R^2, beta weight (κ), and sigma weight (df).

MCMC for spike and slab regression

For each variable j, draw $\gamma_j | \gamma_{-j}, \mathbf{y}$.

$$\gamma | \mathbf{y} \sim C(\mathbf{y}) \frac{|\Omega_\gamma^{-1}|^{\frac{1}{2}}}{|V_\gamma^{-1}|^{\frac{1}{2}}} \frac{p(\gamma)}{SS_\gamma^{\frac{DF}{2}-1}}$$

- Each γ_j only assumes the values 0 or 1.
- V_j is the posterior variance of model γ.
- SS_γ is a "sum of squares," whose expression I will spare you.
- A $|\gamma| \times |\gamma|$ matrix needs to be inverted to compute $p(\gamma|\mathbf{y})$. Cheap! (if there are lots of 0's).

Section summary

The following steps comprise one MCMC iteration:

- Draw state given model parameters and $\mathbf{y}$.
- Draw state component parameters given α.
- Loop over j, drawing each $\gamma_j|\gamma_{-j}, \mathbf{y}, \alpha$ (but integrating out β and σ_ϵ).
- Draw β and σ given γ, α and $\mathbf{y}$.

Repeat for many iterations.

Comment: The discussion here is about "predicting the present" but time series models with many contemporaneous predictors arise frequently.

Outline

Google Trends and Google Correlate

Bayesian structural time series (with sparse regression)

Examples

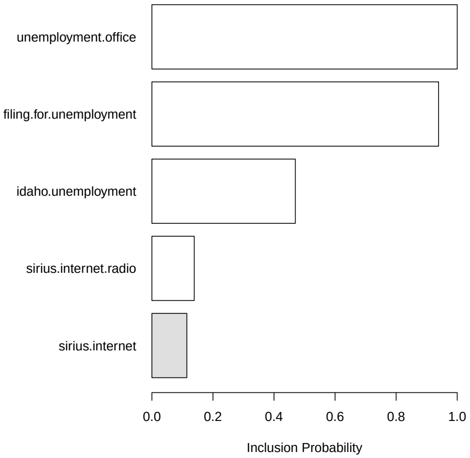
- Initial Claims
- Retail Sales

Conclusions

Posterior inclusion probabilities

With expected model size = 3, and the top 100 predictors from correlate

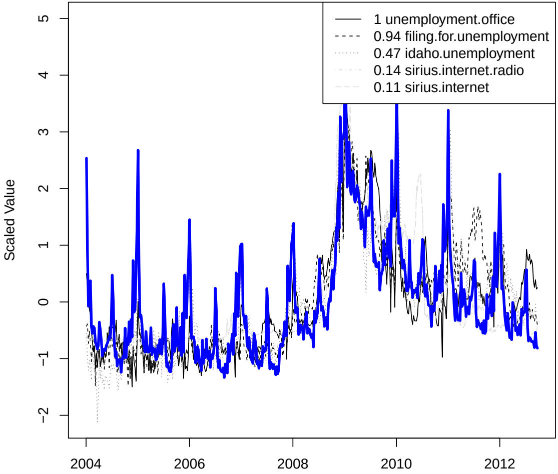
```
plot(model, "coef", inc = .1)
```



- Only showing inclusion probabilities $< .1$.
- Shading shows $Pr(\beta_j > 0|\mathbf{y})$.
  - White: positive coefficients
  - Black: negative coefficients

What got chosen?

```
plot(model, "predictors", inc = .1)
```

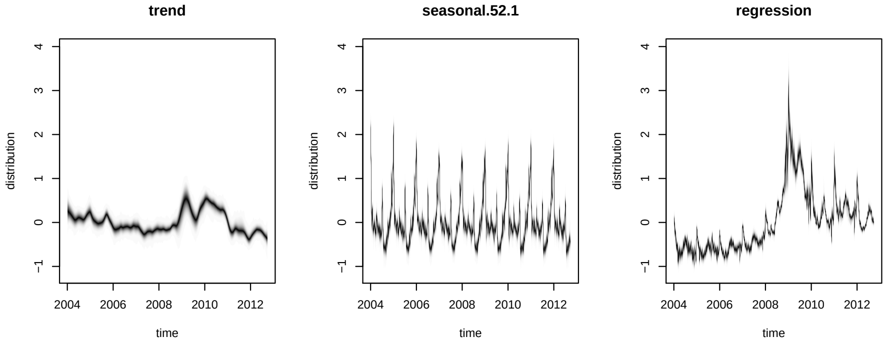


- Solid blue line: actual
- Remaining lines shaded by inclusion probability.

How much explaining got done?

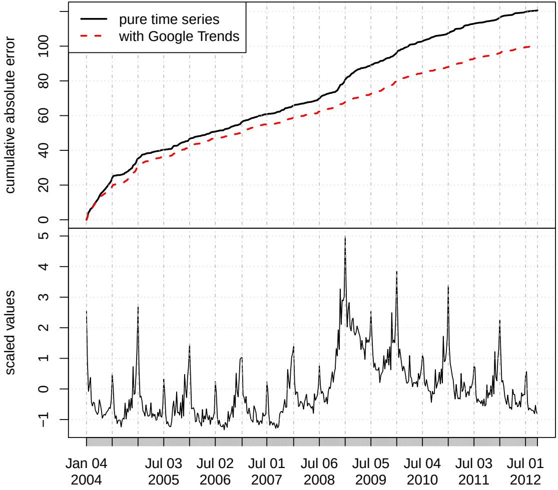
Dynamic distribution plot shows evolving pointwise posterior distribution of state components.

```
plot(model, "components")
```



Did it help?

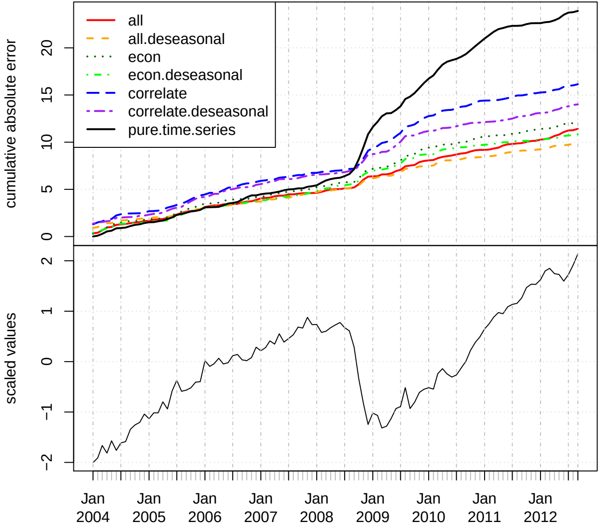
```
CompareBstsModels(list("pure time series" = model1,
                       "with Google Trends" = model2))
```



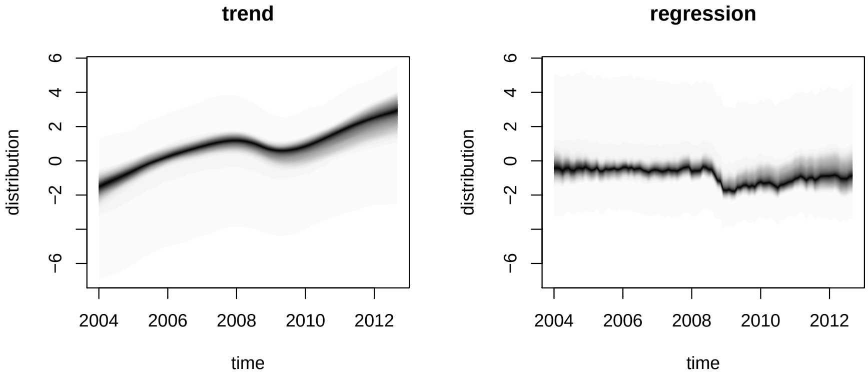
- Plot shows cumulative absolute one-step-ahead prediction error
- The regressors are not very helpful during normal times.
- They help the model to quickly adapt to the recession.

Retail Sales (excluding food services, deseasonalized)

This example incudes query verticals in addition to Correlate queries.

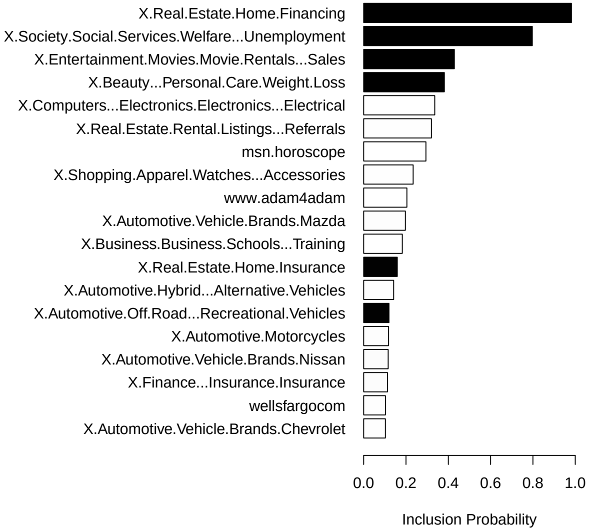


The regression component captures the big disruption



Which predictors are important?

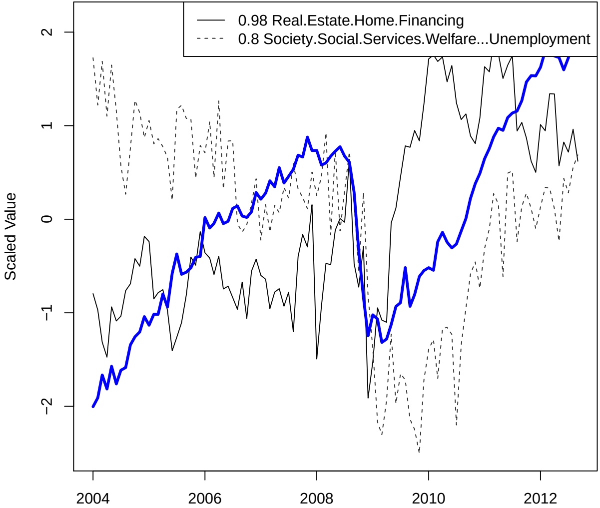
Out of 100 Correlate queries and 100+ “economically relevant” vericals



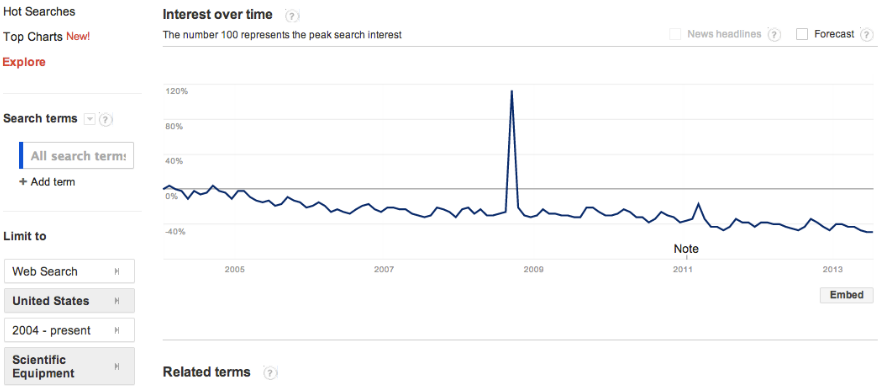
- 24 auto verticals
- 61% of iterations included ≥ 1 auto vertical.

Strong partial correlations beat strong correlations

The top two predictors aren't highly correleated, but have "shocks" in the right places.



Fun with variable selection



- With the full model the most important variable was the vertical for "Scientific Equipment"

Outline

Google Trends and Google Correlate

Bayesian structural time series (with sparse regression)

Examples

Conclusions

Conclusions

- Google Trends and Google Correlate give nearly real time predictors showing public interest in a wide variety of topics.
- Prediction is easy when nothing is changing. Gaussian process handles slow changes. Google trends data helps describe sudden changepoints.
- There's lots of them (even after aggregation). Some should obviously be included/excluded. Some are not so obvious. Average the models.
- A similar "spike and slab" trick can be used to select the time series state components as well.

Carter, C. K. and Kohn, R. (1994).
On Gibbs sampling for state space models.
Biometrika **81**, 541–553.

de Jong, P. and Shepard, N. (1995).
The simulation smoother for time series models.
Biometrika **82**, 339–350.

Durbin, J. and Koopman, S. J. (2002).
A simple and efficient simulation smoother for state space time series analysis.
Biometrika **89**, 603–616.

Frühwirth-Schnatter, S. (1995).
Bayesian model discrimination and Bayes factors for linear Gaussian state space models.
Journal of the Royal Statistical Society, Series B: Methodological **57**, 237–246.

George, E. and McCulloch R. (1997).
Approaches for Bayesian Variable Selection.
Statistica Sinica **7**, 339–374.