

Let's Google it

Can Google search indices nowcast
Norwegian retail sales and unemployment rate?

Jon Ellingsen

Masteroppgave
Universitetet i Oslo
Økonomisk institutt

21. juni 2017

Bakgrunn

Konjunkturanalyse

- ▶ Utfordring: tidsetterslep
- ▶ Løsning: Nowcasting \equiv prediksjon av nåtiden
- ▶ Implementering: samvarierte indikatorer for økonomien

Big data

- ▶ Teknologisk utvikling → mer data
- ▶ Google Trends: **realtidsdata** på **dissaggregert** nivå fra søkemotoren tilbake til 2004

Hypotese

"I den grad folk bruker Google som en informasjonskilde når de skal ta valg, kan Googles søkeindekser avsløre intensjoner som driver økonomiske hendelser før de fanges opp i offentlige statistikker"

Data

Målvariable

- ▶ Detaljomsetning og arbeidsledighet
 - ▶ Nært knyttet til husholdningers atferd
 - ▶ Publiseres månedlig
 - ▶ Følges tett av analytikere
- ▶ Sesongjusterte data
- ▶ Transformasjoner:
 - ▶ Detaljomsetning: $\Delta \log(RS_t)$
 - ▶ Arbeidsledighetsrate: ΔU_t

- ▶ $SVI \equiv$ Indeks av relativ søkefrekvens for søketermer
- ▶ To typer SVIer
 - ▶ Kategorier
 - ▶ Individuelle søketermer
- ▶ Månedlig data
 - ▶ Men tilgjengelig fra første dag i måneden
- ▶ Transformasjoner:
 - ▶ $\Delta \log(SVI_t)$
 - ▶ Winsorizing på 95 % nivå: begrense outliers
 - ▶ Sesongjustering: månedlige dummier

Variabelseleksjon

- ▶ (Stor) modell med alle potensielle prediktorer:

$$y_t = \sum_{i=1}^m \sum_{j=0}^p \beta_{i,j} x_{i,t-j} + \varepsilon_t$$

- ▶ Kun noen av x 'ene er relevante prediktorer
- ▶ Ønsker ikke strukturell modell, kun prediksjon

LARS

Efron et al. (2004)

- ▶ Algoritme som bruker least angle regression for å tilpasse lineære modeller til data av høy dimensjonalitet
- ▶ Gelper and Croux (2008): Utvidelse av lars for tidsserier
- ▶ To steg:
 1. Rangerer de potensielle prediktorene
 2. Velger modellspesifikasjonen

Least angle regression

Rangering

- ▶ $\beta_i = 0 \ \forall \ i$
- ▶ Finn prediktor, x_i , høyst korrelert med y
- ▶ Øk $\beta_i = 0$ i retning av fortegnet på $\text{corr}(y, x_i)$
 - ▶ Behold $r = y - \hat{y}$ på veien
- ▶ Stopp når $\text{corr}(r, x_i) = \text{corr}(r, x_k)$
- ▶ Øk (β_j, β_k) i felles LS retning, $u_{i,k}$
 - ▶ Behold $r = y - \hat{y}$ på veien
- ▶ Stopp når $\text{corr}(r, u_{i,k}) = \text{corr}(r, x_m)$

Least angle regression

Seleksjon

- ▶ Etter hver iterasjon: OLS med alle rangerte prediktorer
 - ▶ Lagrer BIC
- ▶ Minimerer BIC over alle modellene

$$\text{BIC} = -2\log\text{-likelihood} + \ln(n)k$$

Nowcasting

- ▶ Treningsample: februar 2004 - desember 2009
- ▶ Testsample: januar 2010 - januar 2017

Nowcasting

Estimering og prediksjon

Estimering:
$$y_t = \beta_0 y_{t-1} + \sum_{i=1}^{m^*} \sum_{j=0}^{p^*} \beta_{i,j} x_{i,t-j} + \varepsilon_t$$

Prediksjon:
$$\hat{y}_{t+1} = \hat{\beta}_0 y_t + \sum_{i=1}^{m^*} \sum_{j=0}^{p^*} \hat{\beta}_{i,j} x_{i,t+1-j}$$

AR

SVI

AR-SVI

Nowcasting

Evaluering

$$\text{RMSE} = \sqrt{\frac{1}{P} \sum_{i=1}^P \left(y_{t+i} - \hat{y}_{t+i} \right)^2} = \sqrt{\frac{1}{P} \sum_{i=1}^P \left(e_{t+i} \right)^2}$$

DM-test:

$$d_i (= e_{t+i,1}^2 - e_{t+i,2}^2) = \beta + u_i$$

$$H_0 : \beta = 0 \quad \text{eller} \quad H_1 : \beta \neq 0$$

Resultater

Detaljomsetning

	Benchmark models		SVI models		AR-SVI models	
	AR(1)	Random walk	Categories	Queries	Categories	Queries
Expanding window	0.9826	1.8293	1.1752	1.1963	0.9841	1.0433
% change from AR(1)			(+20)**	(+22)*	(0)	(+6)**
% change from random walk			(-36)**	(-35)**	(-46)**	(-43)**
Rolling window	0.9825	1.8293	1.2315	1.3583	1.0103	1.0525
% change from AR(1)			(+25)**	(+38)**	(+3)	(+7)
% change from random walk			(-33)**	(-26)**	(-45)**	(-42)**

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Resultater

Arbeidsledighet

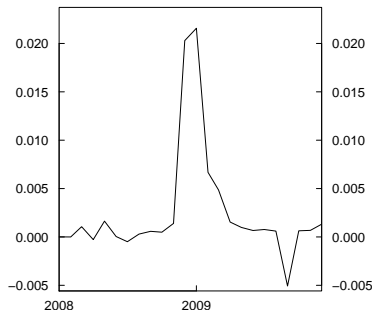
	Benchmark models		SVI models		AR-SVI models	
	AR(1)	Random walk	Categories	Queries	Categories	Queries
Expanding window	0.06129	0.08612	0.06311	0.05925	0.06907	0.06566
% change from AR(1)			(+3)	(-3)	(+13)**	(+7)*
% change from random walk			(-27)**	(-31)***	(-20)**	(-24)***
Rolling window	0.06133	0.08612	0.05692	0.0597	0.06876	0.06443
% change from AR(1)			(-7)	(-3)	(+12)*	(+5)
% change from random walk			(-34)***	(-31)***	(-20)**	(-25)***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Resultater

Finanskrisen

(a) Expanding window.



(b) Rolling window.

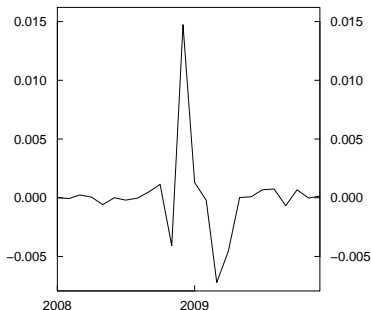


Figure: Unemployment rate. Difference in the squared prediction error between an AR(1) and the AR-SVI_c model during the financial crisis. When the series is above zero, the AR-SVI_c model outperforms the AR(1) model. January 2008 - December 2009.

Diskusjon

- ▶ Subjektiv seleksjon i starten
- ▶ Trekke ut relevante signaler fra SVIene
 - ▶ DFM, PCA etc.
- ▶ Bedre i urolige tider?
- ▶ SVI_t tilgjengelig før y_{t-1}