Predicting the Present with Google Trends: Summary

EE406: Contemporary Economic Issues
Semester 1/ 2021
Faculty of Economics, Thammasat University

ECONOMIC RECORD

THE ECONOMIC SOCIETY OF AUSTRALIA

THE ECONOMIC RECORD, VOL. 88, SPECIAL ISSUE, JUNE, 2012, 2–9

Predicting the Present with Google Trends

HYUNYOUNG CHOI and HAL VARIAN

Google, Inc., California, USA

In this paper we show how to use search engine data to forecast near-term values of economic indicators. Examples include automobile sales, unemployment claims, travel destination planning and consumer confidence.

Literature Review

Epidemiology

- Polgreen et al. (2008) and Ginsberg et al. (2009) showed that search data could help predict the incidence of influenza-like diseases.

- This work was widely publicised and stimulated several further findings in epidemiology, including Brownstein et al. (2009), Corley et al. (2009), Hulth et al. (2009), Pelat et al. (2009), Valdivia and Monge-Corella (2010) and Wilson (2009).

Source: Directly quoted from Choi and Varian (2012)

Literature Review (cont'd)

The US unemployment rate

- Ettredge et al. (2005)
- Suhoy (2009)
- Askitas and Zimmermann (2010)
- D'Amuri and Marcucci (2010)
- Baker and Fradkin (2011)

Retail sales

- Schmidt and Vosen (2009)
- Lindberg (2011)

Consumer sentiment

- Radinsky et al. (2009)
- Huang and Penna (2009)
- Preis et al. (2010)

Housing data

- Wu and Brynjolfsson (2010)

Source: Directly quoted from Choi and Varian (2012)

Example 1: Motor Vehicles and Parts

Motor Vehicles and Parts Dealers' series from the US Census Bureau 'Advance Monthly Sales for Retail and Food Services' report.

AR-1 model

$$y_t = b_1 y_{t-1} + b_{12} y_{t-12} + e_t$$

for the period 2004-01-01 to 2011-07-01

	Estimate	Standard Error	t value	$\Pr(>\|t\|)$
(Intercept)	0.67266	0.76355	0.881	0.381117
lag(y, −1)	0.64345	0.07332	8.776	3.59e−13***
lag(y, −12)	0.29565	0.07282	4.060	0.000118***

Multiple R-squared: 0.7185, Adjusted R-squared: 0.7111.

Source: Directly quoted from Choi and Varian (2012)

Example 1: Motor Vehicles and Parts (cont'd)

- Google Trends contains several automotive related categories.
- A little experimentation shows that two of these categories, **Trucks & SUVs** and **Automotive Insurance** significantly **improve in-sample fit** when added to this regression.

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	−0.45798	0.78438	−0.584	0.561081
lag(y, −1)	0.61947	0.06318	9.805	5.09e−15***
lag(y, −12)	0.42865	0.06535	6.559	6.45e−09***
suvs	1.05721	0.16686	6.336	1.66e−08***
insurance	−0.52966	0.15206	−3.483	0.000835***

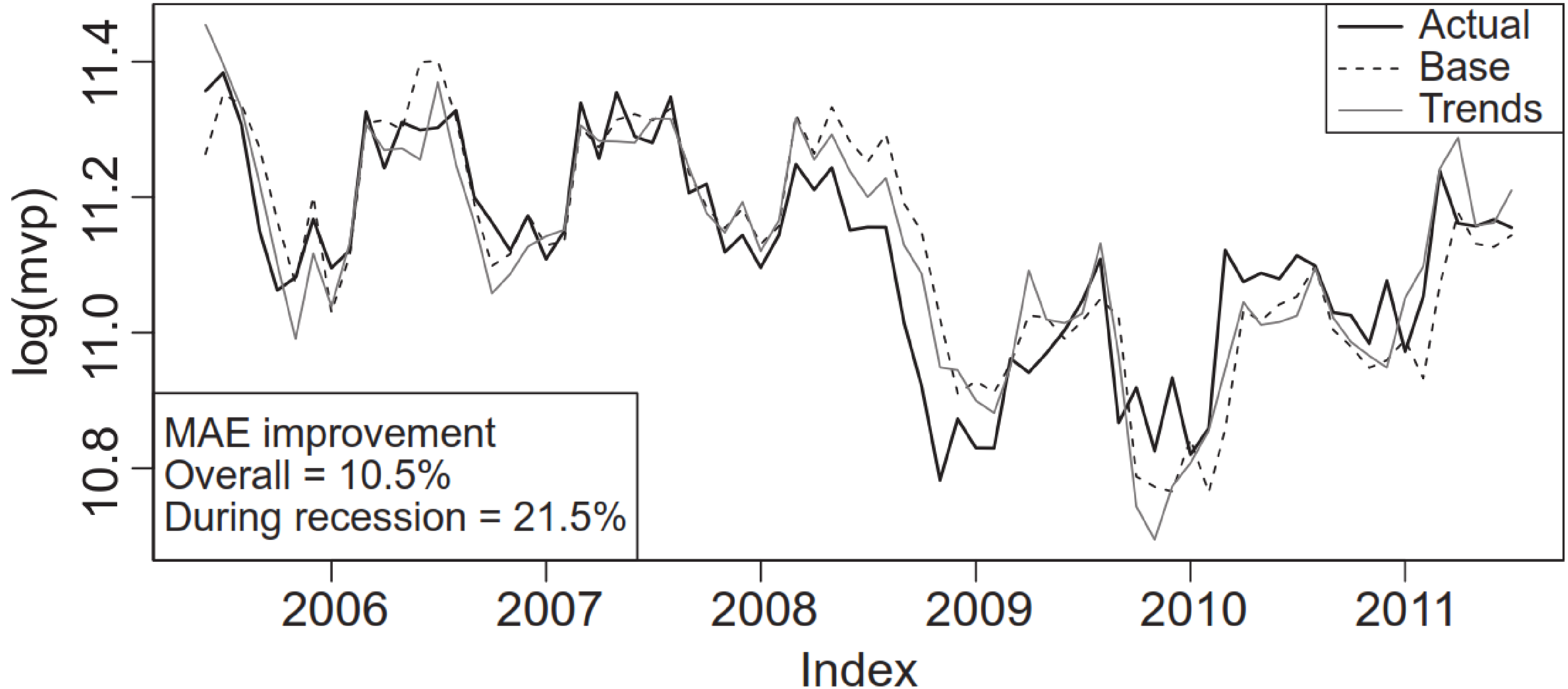
Multiple R-squared: 0.8179, Adjusted R-squared: 0.808.

Source: Directly quoted from Choi and Varian (2012)

Example 1: Motor Vehicles and Parts (cont'd)

FIGURE 2

Motor Vehicles and Parts



Source: Directly quoted from Choi and Varian (2012)

Example 2: Initial Claims for Unemployment Benefits

- Each **Thursday morning** the US Department of Labor releases a report describing the number of people who **filed for unemployment benefits** in the **previous week**.

- Initial claims have a **good record** as a **leading indicator**.

- Macroeconomist Robert Gordon indicates that there is a surprisingly **tight historical relationship** in past US **recessions** between the **cyclical peak** in new claims for unemployment insurance (measured as a four-week moving average) and the subsequent National Bureau of Economic Research **(NBER) trough**.

- Askitas and Zimmermann (2010), Suhoy (2009) and D'Amuri and Marcucci (2010) have confirmed the value of **search data** in **forecasting unemployment** in the **U.S.**, **Germany** and **Israel**.

Source: Directly quoted from Choi and Varian (2012)

Example 2: Initial Claims for Unemployment Benefits (cont'd)

AR-1 model

Start = 2004-01-17, End = 2011-07-02				
	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	0.25488	0.12951	1.968	0.0498*
L(y, 1)	0.98022	0.01007	97.368	<2e−16***

Multiple R-squared: 0.9607, Adjusted R-squared: 0.9606.

Using the **Google Trends** categories **Jobs** and **Welfare Unemployment** we find that these are marginally significant but have **little impact** on **in-sample fit**.

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	1.0563440	0.2686360	3.932	9.98e−05***
L(y, 1)	0.9183560	0.0208778	43.987	<2e−16***
Jobs	0.0007069	0.0003847	1.838	0.0669
Welfare... Unemployment	0.0003752	0.0001838	2.042	0.0418*

Multiple R-squared: 0.962, Adjusted R-squared: 0.9618.

Source: Directly quoted from Choi and Varian (2012)

Example 3: Consumer Confidence for Australia

- **Roy Morgan Consumer Confidence Index** for **Australia**
- It is **not obvious which categories** would be **most helpful** in predicting this series.
- There are a **variety of methods** one can use for **variable selection**.
- See **Castle et al. (2010)** for a recent discussion of this topic with emphasis on **nowcasting applications**.
- We used Google Trends category data for Australia, taking the average value of the category data for the first two weeks of the month and **seasonally adjusting** it using the **R command stl**.

Source: Directly quoted from Choi and Varian (2012)

Example 3: Consumer Confidence for Australia (cont'd)

- The **spike and slab technique** assigned high posterior probabilities to the categories **Crime & Justice**, **Trucks & SUVs** and **Hybrid & Alternative Vehicles**.
- The **last two** are not surprising as they are **highly correlated** with the **price of gasoline**, which is known to **impact consumer confidence** in the **United States**.
- We have **no explanation** for **the first predictor**.
- Plotting the **Crime & Justice** time series shows a definite **correlation** with **consumer confidence** for the period we examine.
- But there is **no way to know** if this correlation **will persist in the future**.

Source: Directly quoted from Choi and Varian (2012)

Example 3: Consumer Confidence for Australia (cont'd)

- The Trends predictors **reduce MAE** of the simple AR-1 model by about **12.7 per cent** for **in-sample forecasts**.
- **One-step-ahead** MAE goes from 3.63 per cent to 3.29 per cent, an **improvement** of **9.3 per cent**.

	Estimate	Standard Error	t value	Pr(>|t|)
(Intercept)	1.5172617	0.2795356	5.428	5.67e−07***
lag(y, −1)	0.6839436	0.0584158	11.708	<2e−16***
Crime ... Justice	−0.0009664	0.0002404	−4.020	0.000129***
Trucks ... SUVs	0.0010600	0.0005346	1.983	0.050735.
Hybrid ... Alternative. Vehicles	−0.0007869	0.0001482	−5.308	9.26e−07***

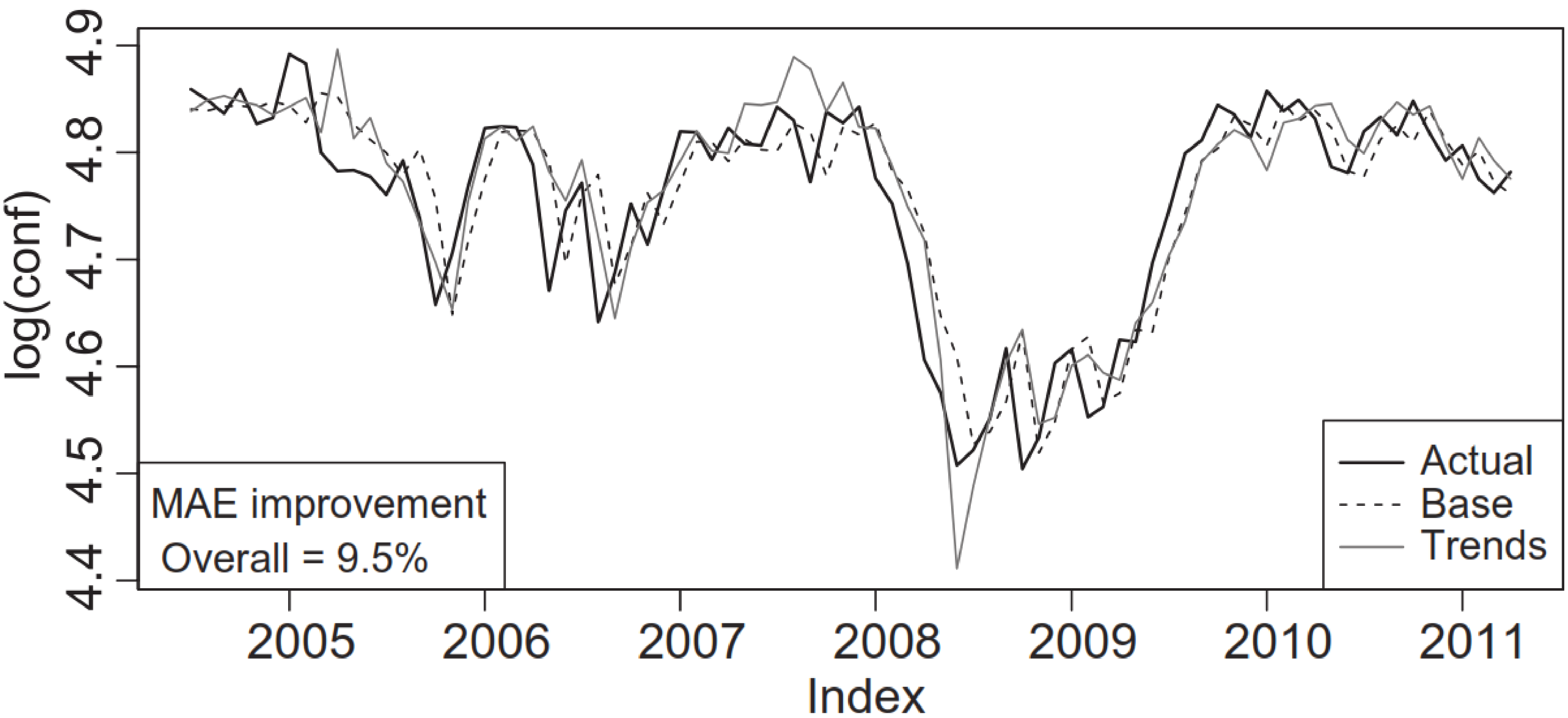
Multiple R-squared: 0.8583, Adjusted R-squared: 0.8514.

Source: Directly quoted from Choi and Varian (2012)

Example 3: Consumer Confidence for Australia (cont'd)

FIGURE 6

Australia Consumer Confidence



Source: Directly quoted from Choi and Varian (2012)