62nd ISI WORLD STATISTICS CONGRESS 2019

18 – 23 August 2019, Kuala Lumpur

Come | Connect | Create

IPS 188

Does Google Search Index Help Track and Predict Inflation Rate? An Exploratory Analysis for India

G. P. Samanta

Reserve Bank of India

21st August 2019, 10.30 am – 12.30 am

Outline/Content

- Background
- Google Search Index
- Data and Methodology
- Empirical Results
- Concluding Remarks

Background

Introduction (1/2)

- **Forward-looking Assessment** of Inflation is an important input to Monetary Policy, particularly in the era of Inflation Targeting

- **Traditional Approaches**
  - Macro-econometric Model
  - Time Series Modelling and Forecasting
  - Small Economic Models (Output-Gap; P-Star, etc.)
  - Survey-Based Estimation of Market Expectation

Introduction (2/2)

- **Emerging Approaches**
  - **Text Mining (**Media Sentiment**)**
  - **Twitter Data/Message Analyses**
  - **Web-Crawling/Scrapping** (Indices-online prices)
  - **Online Search Index (**i.e. Google Search Index**)**

Google Search Indicators - Price and Inflation

Google Search Index : Broad Features

- Google search index for specified keywords reflects the volume of search on the given keywords in Google Search Engine.
- Important features of Google Search Indices

(1) Time series data provide rescaled/relative (not absolute) volume of search on given keywords;

(2) Maximum search interest over a set of time points/periods is assigned a value 100 & that in remaining time points are rescaled accordingly;

(3) Time series replica of the search index vary depending on the date when the enquiry/search were made.

Google Search Index : Link with Economic Variables

- **Many researchers argue that**

  (1) People usually gather information on a topic on which they are anxious or concerned with;

  (2) Google search index/volume reflects individuals' ***revealed expectations*** about the variables/aspects (*represented by the search key words*);

  (3) So, such indices may track or predict corresponding variable/aspect.

Example:

(a) Before ***purchasing consumer goods***, one may be interested in searching for information on the item.

(b) People may be more ***concerned/worried about rising or higher inflation*** and their search interest on the keyword 'inflation' may rise.

Data and Methodology

Data – Google Search & Price Indices

- **Data Frequency**: Monthly
- **Data Period**: From Apr 2012 to Mar 2019
- **Data Sources**:

Google Trend Data:

- Search Indexes for Keywords "Inflation" & "Price"

National Statistics Office (NSO):

- Consumer Price Index – Urban (CPI-U)
- Consumer Price Index – Combined (CPI-C)

Data – Google Search Strategy

- **Search Location:** India
- **Data Frequency / Period:** Monthly / Apr 2012 to Mar 2019
- **Enquiry Dates:** Three randomly chosen days in May 2019
- **Two Search Words:** 'Inflation' & 'Price'

Basic Search Indices/Replicas for a Keyword:

- On each enquiry date, a replica of monthly time series of search index was drawn.
- **Denote** the value in t-th day on i-th replica by $\widetilde{Y}_t^{(i)}$ (i),i=1,2,3.

Data – Google Search Strategy

- **Given**, $\widetilde{Y}_t^{(i)} = \dfrac{\widetilde{X}_t^{(i)}}{\delta_i} \text{ x } 100;$

  where $\widetilde{X}_t^{(i)}$ = Unknown absolute volume of search in t-th day on i-th replica;

  δ_i = maximum no of search over t on $\widetilde{X}_t^{(i)}$, i=1,2,3

- **Overall/Pooled search indicator**:

  Y_t = Geometric-Mean (GM) of $\widetilde{Y}_t^{(i)}$s, i=1,2,3

  $\propto$ GM of $\widetilde{X}_t^{(i)}$s, i=1,2,3 (i.e. *GM of actual but unknown search volumes*)

- We call the overall search index as follows
  - **Keyword:** "Inflation" → **Overall Search Index**: GM-Infl
  - **Keyword**: "Price" → **Overall Search Index**: GM-Price

Data – K-Period Inflation/Change

- **K-period Percentage Change of a Variable X_t**

$$\boldsymbol{\pi}_{\boldsymbol{t}} = \left[\log_e(X_t) - \log_e(X_{t-k})\right] \text{ x } 100;$$

$X_t = $ CPI-C; CPI-U; GM-Price and GM-Infl

Empirical Results

Unit-Root Tests – Different Transformations (1/2)

Variable	Augmented Dickey-Fuller					Phillips-Perron				
	Optimal Lag	Unit-Root Test		Test for Trend		Band-width	Unit-Root Test		Test for Trend	
		Test Statistics	p-value	Test Statistics	p-value		Test Statistics	p-value	Test Statistics	p-value
(A) Annual Inflation Rate/Annual Percentage Change										
lnCPI-C	7	-2.8429	0.1869	1.7515	0.0843	3	-2.2117	0.4769	1.3288	0.1876
lnCPI-U	7	-3.7490	0.0247	2.8582	0.0056	3	-2.6516	0.2592	1.9785	0.0512
lnGMPrice	6	- 3.9981	0.0125	3.0147	0.0036	1	-3.4091	0.0569	1.8850	0.0629
lnGMInfl	0	-5.8940	0.0000	-5.2213	0.0000	2	-5.8860	0.0000	-5.2213	0.0000
(B) First-Difference Series of Variables at (A) above										
ΔlnCPI-C	6	-7.0514	0.0000	-4.3018	0.0001	0	-5.6672	0.0000	-2.0093	0.0478
ΔlnCPI-U	6	-6.4774	0.0000	-3.4508	0.0010	2	-6.0054	0.0000	-1.7980	0.0759
ΔlnGMPrice	3	-5.9215	0.0000	-0.9223	0.3593	8	-8.1457	0.0000	-0.6333	0.5283
ΔlnGMInfl	1	-9.4089	0.0000	-0.0126	0.9900	13	-20.5903	0.0000	0.1838	0.8546
(C) Second-Difference Series of Variables at (A) above										
Δ^2lnCPI-C	7	-7.8157	0.0000	0.3484	0.7287	8	-15.7029	0.0000	0.1071	0.9150
Δ^2lnCPI-U	7	-7.7350	0.0000	0.5843	0.5610	4	-11.9371	0.0000	0.2176	0.8283
Δ^2lnGMPrice	4	-8.7383	0.0000	-0.0139	0.9889	6	-51.1038	0.0001	0.0135	0.9892
Δ^2lnGMInfl	4	-7.7786	0.0000	-0.0493	0.9913	6	-51.2341	0.0001	0.0234	0.9852

Unit-Root Tests – Different Transformations (2/2)

Variable	Augmented Dickey-Fuller					Phillips-Perron				
	Optimal Lag	Unit-Root Test		Test for Trend		Band-width	Unit-Root Test		Test for Trend	
		Test Statistics	p-value	Test Statistics	p-value		Test Statistic	p-value	Test Statistics	p-value
(D) De-Trending Linear-Time-Trend of Variables at (A) above										
elnCPI-C	7	-2.8428	0.1869	-5.1194	0.0000	1	-5.3040	0.0002	0.0916	0.9273
elnCPI-U	7	-3.7490	0.0247	-4.8328	0.0000	3	-2.6516	0.2592	-2.9244	0.0044
elnGMPrice	6	-3.9981	0.0125	-0.5160	0.6075	1	*-3.4091*	0.0569	-0.6357	0.5267
elnGMInfl	0	-5.8940	0.0000	0.1911	0.8489	2	-5.8860	0.0000	0.1911	0.8489
(E) De-Trending Quadratic-Time-Trend of Variables at (A) above										
e2lnCPI-C	1	-4.4954	0.0027	-0.5439	0.5880	3	*-3.2540*	0.0811	-0.4989	0.6191
e2lnCPI-U	1	-3.6238	0.0336	-0.4053	0.6864	3	-2.8794	0.1743	-0.3621	0.7182
e2lnGMPrice	3	-3.6125	0.0348	-0.4351	0.6647	1	-3.5711	0.0384	-0.3378	0.7364
e2lnGMInfl	0	-6.0223	0.0000	0.0601	0.9522	1	*-6.0519*	0.0601	0.0601	0.9522

Unit-Root Tests –Annual Changes & Their Differences

Variable	Augmented Dickey-Fuller					Phillips-Perron				
	Optimal Lag	Unit-Root Test		Test for Trend		Band-width	Unit-Root Test		Test for Trend	
		Test Statistics	p-value	Test Statistics	p-value		Test Statistic	p-value	Test Statistics	p-value
(A) Annual Inflation Rate/Annual Percentage Change										
gCPI-C	1	-3.1547	0.1018	-2.1941	**0.0316**	3	-2.3619	0.3961	-1.4354	0.1556
gCPI-U	4	-1.8559	0.6669	-0.7415	0.4610	5	-1.8075	0.6912	-0.5764	0.5662
gGMPrice	0	-5.3040	0.0002	0.0916	0.9273	1	-5.3040	0.0002	0.0916	0.9273
gGMInfl	0	-2.2835	0.4373	-0.8676	0.3885	4	-2.5811	0.2900	-0.8676	0.3885
(B) First-Difference Series of the Variables at (A) above										
ΔgCPI-C	11	-4.3419	0.0052	1.5623	0.1248	13	-6.2063	0.0000	0.7070	0.4819
ΔgCPI-U	1	-6.6947	0.0000	1.0733	0.2869	13	-5.9064	0.0000	1.0946	0.2774
ΔgGMPrice	1	-9.1522	0.0000	0.1058	0.9161	26	-20.2929	0.0001	-0.1534	0.8785
ΔgGMInfl	0	-7.8708	0.0000	-0.2613	0.7946	2	-7.8397	0.0000	-0.2613	0.7946

Predictive Power – Granger's Causality (1/2)

Google Search Data	Null Hypothesis	Obs	Lag	F-Statistics	P-Value
gGMPrice	gGMPrice does not Granger Cause gCPI-C	66	9	1.9125	0.0732
	gCPI-C does not Granger cause gGMPrice	66	9	2.0230	0.0575
	gGMPrice does not Granger Cause gCPI-U	72	3	0.6434	0.5899
	gCPI-U does not Granger cause gGMPrice	72	3	3.3392	0.0246
ΔgGMPrice	ΔgGMPrice does not Granger Cause ΔgCPI-C	72	2	0.9296	0.3937
	ΔgCPI-C does not Granger cause ΔgGMPrice	72	2	3.4142	0.0387
	ΔgGMPrice does not Granger Cause ΔgCPI-U	72	2	1.7898	0.1749
	ΔgCPI-U does not Granger cause ΔgGMPrice	72	2	3.3384	0.0415
ΔgGMInfl	ΔgGMInfl does not Granger Cause ΔgCPI-C	62	12	1.6213	0.1279
	ΔgCPI-C does not Granger cause ΔgGMInfl	62	12	2.0394	0.0483
	ΔgGMInfl does not Granger Cause ΔgCPI-U	62	12	1.6005	0.1341
	ΔgCPI-U does not Granger cause ΔgGMInfl	62	12	2.8420	0.0073

Predictive Power – Granger's Causality (2/2)

Google Search Data	Null Hypothesis	Obs	Lag	F-Statistics	P-Value
Δ2lnGMInfl	Δ2lnGMInfl does not Granger Cause Δ2lnCPI-C	66	7	2.0053	0.0721
	Δ2lnCPI-C does not Granger cause Δ2lnGMInfl	66	7	2.2713	0.0431
	Δ2lnGMInfl does not Granger Cause Δ2lnCPI-U	63	10	1.4382	0.1974
	Δ2lnCPI-U does not Granger cause Δ2lnGMInfl	63	10	2.0800	0.0483
e2lnGMPrice	e2lnGMPrice does not Granger Cause e2lnCPI-C	78	9	2.7336	0.0098
	e2lnCPI-C does not Granger cause e2lnGMPrice	78	9	2.6099	0.0131
	e2lnGMPrice does not Granger Cause e2lnCPI-U	78	9	2.8349	0.0077
	e2lnCPI-U does not Granger cause e2lnGMPrice	78	9	3.4623	0.0017
e2lnGMInfl	e2lnGMInfl does not Granger Cause e2lnCPI-C	78	9	1.1596	0.3334
	e2lnCPI-C does not Granger cause e2lnGMInfl	78	9	2.0825	0.0456
	e2lnGMInfl does not Granger Cause e2lnCPI-U	78	9	0.6332	0.7640
	e2lnCPI-U does not Granger cause e2lnGMInfl	78	9	2.2178	0.0332

Tracking Power – Correlation Coefficient

Variable Pair	Correlation Coefficient	Variable Pair	Correlation Coefficient
gGMPrice & gCPI-C	0.1805 (0.1212)	gGMInfl & gCPI-C	0.3591 (0.0016)
gGMPrice & gCPI-U	0.1641 (0.1595)	gGMInfl & gCPI-U	0.3509 (0.0020)
ΔgGMPrice & ΔgCPI-C	0.0150 (0.8985)	ΔgGMInfl & ΔgCPI-C	*0.1608 (0.1704)*
ΔgGMPrice & ΔgCPI-U	-0.0082 (0.9447)	ΔgGMInfl & ΔgCPI-U	*0.1710 (0.1450)*
ΔlnGMPrice & ΔlnCPI-C	0.1695 (0.1186)	ΔlnGMInfl & ΔlnCPI-C	0.0373 (0.7333)
ΔlnGMPrice & ΔlnCPI-U	0.1401 (0.1981)	ΔlnGMInfl & ΔlnCPI-U	-0.0280 (0.7977)
e2lnGMPrice & e2lnCPI-C	0.3977 (0.0001)	e2lnGMInfl & e2lnCPI-C	0.2557 (0.0168)
e2lnGMPrice & e2lnCPI-U	0.3701 (0.0013)	e2lnGMInfl & e2lnCPI-U	0.2913 (0.0062)

Figures within ( ) are p-values.

Concluding Remarks

Summary and Conclusions

Time Series Properties

- Annual change of each series is I(1) process.
- Each log-transformed series belongs to I(2) class ***or*** TS class (*Quadratic trend*).

Predictive Ability (*Granger's Causality Framework*)

- Bi-directional causal-relationship (*predictive ability*) between CPI-C & GMPrice and CPI-C & GMInfl.
- Bi-directional ***predictive ability:*** CPI-U and GMPrice.
- GM-Infl is influenced by past CPI-U.

Summary and Conclusions (2/2)

Tracking Ability

- Both GM-Price and GM-Infl have ***strong ability to track*** inflation based on both CPI-C and CPI-U

- Annual percentage change in GMInfl is strongly correlated with inflation rates based on both CPI-C and CPI-U.

Select References

Agarwal, Aprrov, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau (2011), "Sentiment Analysis of Twitter Data", ***Proceedings of the Workshop on Language in Social Media (LSM 2011), pages 30–38, Portland, Oregon***, 23 June 2011.

Cavallo, Alberto (2013), "Online and Official Price Indexes: Measuring Argentina's Inflation", ***Journal of Monetary Economics***, Vol. 60, pp. 152-65.

Cavallo, Alberto (2015), "Scraped Data and Sticky Prices", ***NBER Working Paper Series***, Working Paper 21490.

Cavallo, Alberto (2016), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", ***NBER Working Paper*** 22142, March.

Cavallo, Alberto (2017), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", ***American Economic Review***, Vol. 107, No. 1, pp. 283-303.

Cavallo, Alberto, Brent Neiman and Roberto Rigobon (2015), "The Price Impact of Joining a Currency Union: Evidence from Latvia", ***IMF Economic Review***, Vol 63, No. 2.

Cavallo, Alberto and Roberto Rigobon (2011), "The Distribution of the Size of Price Changes", ***NBER Working Paper Series***, Working Paper 16760.

Cavallo, Alberto and Roberto Rigobon (2016), "The Billion Prices Project: Using Online Prices for Measurement and Research", ***Journal of Economic Perspectives***, Vol. 30, No. 2, Spring, pp. 151-78.

Choi, Hyunyoung and Hal Varian (2009a), "Predicting the Present with Google Trends", ***Technical Report***, Google Inc. (**Website:** https://research.google.com/googleblogs/pdfs/ google_predicting_the_present.pdf).

Choi, Hyunyoung and Hal Varian (2009b), "Predicting Initial Claims for Unemployment Insurance Using Google Trends", ***Technical Report***, Google Inc. (**Website**: https://research.google.com/archive/papers/initialclaimsUS.pdf).

Choi, Hyunyoung and Hal Varian (2012), "Predicting the Present with Google Trends", ***Economic Record, The Economic Society of Australia***, Vol. 88, Issue S1 (Special Issue), pp 2-9.

Ettredge, Michael, John Gerdes and Gilbert Karuga (2005), "Using Web-based Search Data to Predict Macroeconomic Statistics", ***Communications of the ACM***, Vol 48, No. 11, November, pp. 87-92.

Ellen Tobback, Stefano Nardelli, David Martens (2017), "Between Hawks and Doves: Measuring Central Bank Communication", **Working Paper Series, European Central Bank**, No. 2086, July.

Guzmán, Giselle (2011), "Internet Search Behavior as an Economic Forecasting Tool: the Case of Inflation Expectations", ***Journal of Economic and Social Measurement***, Vol. 36, No. 3, IOS Press, pp. 119-167.

Seabold, Skipper and Andrea Coppola (2015), "Nowcasting Prices Using Google Trends – An Application to Central America", ***Policy Research Working Paper 7398, World Bank Group***, August.

THANK YOU