# Epidemiology of Inflation Expectations and Internet Search- An Analysis for India

J, Saakshi and Sahu, Sohini and Chattopadhyay, Siddhartha

IIT Kanpur, IIT Kanpur, IIT Kharagpur

6 March 2019

# Epidemiology of Inflation Expectations and Internet Search- An Analysis for India[*]

## Abstract

This paper investigates *how* inflation expectations of individuals are formed in India. We investigate if the news on inflation plays a role in the formation of inflation expectations following the epidemiology-based work by Carroll (2003). The standard literature on this topic considers news coverage by the print and audio-visual media as the sources of formation of inflation expectations. Instead, we consider the Internet as a potential common source of information based on which agents form their expectations about future inflation. Based on data extracted from Google Trends, our results indicate that during the period 2006 to 2018, the Internet has indeed been a common source of information based on which agents have formed their expectations about future inflation, and the Internet search sentiment has had some impact on inflation expectations. Additionally, based on the inflation expectations series derived from the Google Trends data, we find that there is presence of "information stickiness" in the system since only a small fraction of the population update their inflation expectations each period.

Keywords: Inflation expectations, Epidemiology, Internet search, Google Trends, India.

## 1. Introduction

That inflation expectations by microeconomic agents play a significant role in monetary policy and analysis, is a well-established fact. Taking this into cognizance, surveys like the Michigan Consumer Survey in the US, the Reserve Bank of New Zealand's Survey of Expectations, the Bank of England Inflation Attitudes Survey, the Inflation Expectations Survey of Household in India, etc. are conducted to obtain information on inflation expectations of the general public. Although these surveys are comprehensive in terms of recording various aspects of the quantitative and qualitative responses on inflation expectations of the general public, very little is known about *how* exactly the survey respondents form their expectations about future prices.

In this aspect, the success of the otherwise-dominant rational expectations hypothesis, in explaining how agents form their expectations about inflation, has been limited and it has been criticized due to its inadequacy in modelling real-life expectations (Pfajfar and Santoro, 2013). This led to alternate models for explaining expectations formation among agents based on behavioural, sociological and epidemiological factors. In this paper, we particularly focus on the empirical evidence of epidemiological sources of inflation expectations formation in India.

The literature on epidemiological sources of inflation expectations took off with the seminal work of Christopher Carroll in 2003. The basic idea, that draws inspiration from epidemiology, concerns how information about expected inflation emanates from a certain common source, like the forecasts made by professional forecasters (SPF), and the general public uses this information from experts to update their beliefs about future inflation. Carroll's work provides a micro foundation for an aggregate expectations equation in a sticky

information set-up consistent with Mankiw and Reis (2002). Since then, the epidemiological model has been explored by Lamla and Maag (2012), Pfajfar and Santoro (2013), Erhmann, Pfajfar and Santoro (2014) and Lei et. al. (2015). The first few studies induct media news as the common source of information based on which the general public form their own inflation expectations. Lei et. al. (2015) use newspaper reports on inflation as the common source and further categorize the newspapers (general, economic, political, etc.) to assess the impact of each type of newspaper report on inflation expectations formation in China.

Given this background, our paper focuses on the survey-based responses on inflation expectations by the general public in India and addresses two related issues. First, we explore if the Internet can be modelled as a common source of information based on which agents form their inflation expectations. Second, we ask if there is evidence of Carroll-type epidemiological sources of inflation expectations formation in India?

The contribution of this work to the existing literature is two-fold. This is the first study to our knowledge that uses the Inflation Expectations Survey of Household data to validate the presence of Carroll-type epidemiological sources of inflation expectations formation in India. Second, we depart from all the epidemiology-based studies mentioned above as far as the common source of information is concerned. Instead of considering the forecasts made by the professional forecasters or those provided by the news media, as the common source of information, we hypothesize that the Internet is the common source through which the general public update their future inflation expectations. This is not an implausible assumption in the Indian context since the number of Internet users in India is second[1] in the

---

[1] https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/ accessed on April 21, 2018.

world at approximately 481 million users[2] (about 24% of the total population) during December 2017. Additionally, the Inflation Expectations Survey of Households is currently conducted in the urban cities of India where the usage of Internet is pervasive. As per the report "Internet in India 2017" by the Internet and Mobile Association of India (IAMAI) and Kantar IMRB, urban India with an estimated population of 455 million already has 295 million using the Internet which is almost 65% of the urban population. To corroborate our hypothesis that the Internet is indeed a common source of information that people use to form their expectations about future inflation, we use the statistics of inflation-related searches conducted through Google in India between 2006 and 2018, the data on which is extracted through Google Trends.

In the process, this paper brings together two strands in the existing literature- the first one being the epidemiological sources of inflation expectations pioneered by Carroll (2003) as mentioned above, and the other is the use of meaningful data based on Internet search initiated by Choi and Varian (2012). Choi and Varian (2012) have demonstrated how the Google Trends data can be used to forecast automobile sales, forecast travel destination planning, unemployment, etc. More recently, researchers have used Google Trends data to predict spread of diseases (Cho. et. al., 2013), stock market movements (Dergiades et. al., 2015), oil price volatility (Afkhami et. al., 2017), unemployment rates (Naccarato et. al., 2018) etc. That the Internet-based data is gaining traction in terms of macroeconomic research in India is evident from the recent news that the Reserve Bank of India is all set to start a "Big Data Analytics" division by the end of the year 2018[3]. In this backdrop, this

---

[2] https://economictimes.indiatimes.com/tech/internet/internet-users-in-india-expected-to-reach-500-million-by-june-iamai/articleshow/63000198, accessed on April 21, 2018.

[3] https://www.livemint.com/Opinion/qHS04i31OfR8B4vskHVesJ/RBI-enters-the-exciting-new-world-of-Big-Data-analytics.html , accessed on April 27, 2018.

particular use of Google Trends data for epidemiological sources of inflation expectations formation, is a new addition to the existing literature, both from the point of view of a novel common source for information and also as a new application of the Internet search data.

Our findings indicate that the epidemiological sources of inflation expectations of the general public in India hold for both Carroll (2003) type specification based on the inflation expectations of professional forecasters, as well as the Internet search data. Thus, we infer that the Internet is indeed a significant source from where the agents derive information and update their future inflation expectations. Additionally, we find some limited evidence that the Internet search "sentiment", that is, the keywords used to conduct inflation-related searches, affect inflation expectations in the expected direction. All our findings are in line with the various works on epidemiology cited above, especially that of Lei et. al. (2015).

The rest of the paper is organized as follows. Section 2 outlines the benchmark epidemiology framework pioneered by Carroll (2003), including the modifications made to accommodate our model. Section 3 discusses the data sources. Section 4 analyses the results and Section 5 concludes.

## 2. **The Model**

Carroll's (2003) seminal work on the epidemiology of inflation expectations draws inspiration from Kermack-McKendrick (1927) model in the epidemiology literature that explains the process of transmission of disease in society from a "common source". At any given point in time, a constant fraction $\lambda$ of the population gets affected by the disease that spreads from a common source, like air pollution, while the remaining (1- $\lambda$) fraction of the population do not get affected but nevertheless remain susceptible. In the next time period, fraction $\lambda$ of those who escaped the disease in the previous period get infected, while the rest

do not. This is how a disease is transmitted from a common source over a period of time. Analogous to disease transmission from a common source, Carroll's (2003) model hypothesizes that news media is the "common source" of information (transmission) based on which a section λ of the population update their expectations about future inflation. The remaining (1-λ) proportion of the population, who have not been affected by the "common source" or the news media, continue with the inflation expectations from the last period. Thus, at any given point in time, there is a combination of agents who have updated their inflation expectations and those who continue with their inflation belief from the previous period. This is how an element of "information stickiness" is introduced in the model.

$$E_t(\pi_{t+1}) = \lambda N_t(\pi_{t+1}) + (1 - \lambda)\{\lambda N_{t-1}(\pi_{t+1}) + (1 - \lambda)(\lambda N_{t-2}(\pi_{t+1}) + \cdots)\} \qquad (1)$$

where, $E_t(\pi_{t+1})$ is the expectation of inflation rate for period t+1 conditional on the information available till time period $t$, $\pi_{t+1}$ is the realized inflation for period $t + 1$ and $N_t(\pi_{t+1})$ is the updated news at period $t$ to form inflation expectations for period $t + 1$. Quoting Carroll (2003), "The derivation of this equation is as follows. In period $t$ a fraction λ of the population will have been 'infected' with the current-period newspaper forecast of the inflation rate next quarter, $N_t[\pi_{t+1}]$. Fraction $(1 - \lambda)$ of the population retains the views that they held in period $t-1$ of period $t+1$'s inflation rate. Those period-$t-1$ views in turn can be decomposed into a fraction λ of people who encountered an article in period $t - 1$ and obtained the newspaper forecast of period $t + 1$'s forecast, $N_{t-1}[\pi_{t+1}]$, and a fraction $(1 - \lambda)$ who retained their period-$t - 2$ views about the inflation forecast in period $t + 1$. Recursion leads to the remainder of the equation".

The expression $\{\lambda N_{t-1}(\pi_{t+1}) + (1-\lambda)(\lambda N_{t-2}(\pi_{t+1}) + \cdots)\}$ can be expressed as the sum of an infinite series and equation (1) can be written succinctly as:

$$E_t(\pi_{t+1}) = \lambda N_t(\pi_{t+1}) + (1-\lambda)E_{t-1}(\pi_t) \tag{1a}$$

For the purpose of estimation, we use the following equation (2) that follows from equation (1a),

$$E_t(\pi_{t+1}) = \alpha_1 N_t(\pi_{t+1}) + \alpha_2 E_{t-1}(\pi_t) + \varepsilon_t \tag{2}$$

Most works on the epidemiological sources of inflation expectations, including that by Carroll (2003), use the forecasts made by the professional forecasters as the source from which the general public form their expectations (i.e. $N_t(\pi_{t+1})$). Lei et. al. (2015) assume that $N_t(\pi_{t+1})$ is directly obtained from news reports. Stepping aside from these assumptions, we hypothesize that agents draw their information from the Internet search represented by the Google Trends (GT) data[4]. Thus, we get equation (3) as follows:

$$N_t(\pi_{t+1}) = \gamma GT_t + \eta_t \tag{3}$$

Substituting equation (3) in equation (2), we get:

$$E_t(\pi_{t+1}) = \beta GT_t + \alpha_2 E_{t-1}(\pi_t) + \mu_t \tag{4}[5]$$

---

[4] It is to be noted that when we say that agents draw information from a common source like the Internet, it is not implied that people who have been surveyed about their inflation expectations have necessarily searched the Internet. In other words, there is no one-to-one correspondence between the response of the surveyed individuals and the individuals who searched the Internet. This is akin to the assumption in the standard epidemiology literature based on newspapers as a common source, where it is not the case that people who are surveyed about their inflation expectations have necessarily read the newspaper to derive information. We assume, like the newspaper is a source, the Internet is also a source of information and to show that the Internet is indeed a source based on which people form their expectations, we use the search statistic data provided by the Google Trends. Search statistics imply that people indeed have been using the Internet.

[5] $\beta = \gamma * \alpha_1$ and $\mu_t = \varepsilon_t + \gamma\eta_t$

Following Lei et. al. (2015), we categorize the inflation-related searches into three categories-

"Favorable" (like low inflation, low prices), "Unfavorable" (like high inflation, rising prices)

and "Neutral" (stable prices). The exact keywords searched for are outlined in the Data

description section. Accordingly, to assess the search sentiment and its impact on inflation

expectations, as in Lei et. al. (2015), we estimate the following equation:

$$E_t(\pi_{t+1}) = c + \alpha E_{t-1}(\pi_t) + \beta^+ GT\_Favorable_t + \beta^- GT\_Unfavorable_t +$$
$$\beta^0 GT\_Neutral_t + \mu_t \tag{5}$$

However, due to very low number of searches in the "neutral" category, no data points are

available. Thus, for search sentiment we estimate equation (6) which is a version of equation

(5) after excluding the variable $GT\_Neutral_t$:

$$E_t(\pi_{t+1}) = c + \alpha E_{t-1}(\pi_t) + \beta^+ GT\_Favorable_t + \beta^- GT\_Unfavorable_t + \mu_t \tag{6}$$

Having presented an outline of the epidemiological model suitable for our analysis, we

proceed to the following section that give an overview of the data used for this study,

along with a mention of the data sources.

## 3. Data description

### 3.1 Inflation expectations

Data on inflation expectations of households in India is available from the Inflation

Expectations Survey of Households (IESH) conducted quarterly by the Reserve Bank of India

(the Central Bank of India) since the second quarter of 2006.

The survey covers about 5000 urban individuals in each round across 18 major cities in India at present. The sample surveyed represents both genders, nine age categories and seven different broad occupational categories in each city. The survey respondents are asked to quote quantitative numbers based on their inflation perception, one-quarter ahead and one-year ahead inflation expectations. Figure 1 and Figure 2 plot the mean one-quarter ahead and mean one-year ahead inflation expectations of the general public in India against realized CPI and WPI respectively.

[Insert Figure 1 about here]

[Insert Figure 2 about here]

At a cursory glance, the above figures indicate that through the initial years until about the fourth quarter of 2011, the household inflation expectations have moved in tandem with WPI inflation while staying above the actual inflation. Post 2011, the household expectations series shows a co-movement with CPI inflation while staying above CPI inflation all along. It would not be out of place to mention here that the official inflation rate of India was calculated using WPI till 2014 and since then, it is being calculated using CPI (combined). Due to this switch in the official inflation series, it is difficult to infer as to which inflation numbers the general public have been following during the period covered by this study.

Apart from the IESH, there is the Survey of Professional Forecasters (SPF), which at present is conducted bi-monthly by the Reserve Bank of India. This survey was initiated in 2007 and was conducted at a quarterly frequency until April 2014. At present, 21 professional forecasters participate in this survey and give their quantitative forecasts on CPI-combined

inflation, CPI-core inflation and WPI-combined inflation rates for the current and future quarters, apart from other macroeconomic forecasts. Figure 3 shows the mean one-quarter ahead and mean-one-year ahead forecasts of the professional forecasters across different categories of inflation.

[Insert Figure 3 about here]

As evident from Figure 3, the inflation forecasts by the Survey of Professional Forecasters (SPF) in most periods have stayed below the WPI inflation. If we combine this observation with the conclusions from Figures 1 and 2 that the household inflation expectations have always remained above WPI inflation, it seems to indicate that households are deriving information or knowledge about future inflation from sources other than the forecasts made by the professional forecasters. This makes a case for our study where we hypothesize that there might be possibly some common source of information like the Internet, based on which the general public form their expectations about future inflation.

Both IESH and SPF data are publicly available at the Reserve Bank of India website (https://www.rbi.org.in/).

### 3.2 Internet search data

The Internet search data has been collated from Google Trends (https://trends.google.co.in/trends/). Google Trends is a public web facility made available by Google Inc. that gives Google search related statistics.

Search related data in Google Trends can be filtered according to geographic zones, time period, frequency and categories. For example, we looked into the search history for the keyword "Inflation" in India starting 2006: Q2 till 2018: Q2 at a quarterly frequency across "All categories". The data generated is an index known as the Google Trend Index for the searched term. It is to be noted that Google Trends does not generate data on search volume. Instead, it expresses the searched term as a fraction of the total number of searches conducted in the zone of our interest during a specified time period and assigns the number 100 against the highest fraction. The rest of the data series is rescaled with respect to the data point that corresponds to the number 100.

We draw attention to the fact that since the Google Trends Index is not a search volume data but relative to the total number of searches conducted and is rescaled based on the highest relative search, the index might display slight variation in the past data due to changes in the sample period.

We collected data on the search history of each term related to inflation, namely- "Inflation", "CPI Inflation", "WPI Inflation", "Core Inflation" and "Headline Inflation"; across three search categories, namely- "All Category Web Search", "Business and Industrial Web Search" and "News Web Search", that originated from India between 2006: Q2 and 2018: Q2.

For search sentiment, we collected data on the search history of terms to represent "Favorable" search, "Unfavorable" search and "Neutral" search in India during the time period 2006: Q2 and 2018: Q2.

Terms like "price fall", "price decrease", and "low inflation", were clubbed under "Favorable" search, reflecting positive search sentiment. "Price rise", "price increase", and "high inflation", were clubbed under "Unfavorable" search, thereby reflecting negative search sentiment. Not enough searches were recorded for terms related to "Neutral" search, thereby generating no data points by Google Trends. Data was generated by using Google Trends comparison that allowed search comparison across multiple keywords.

## 4. Results

In this section, we present our empirical analysis in two parts. In the first part, following Lei et. al. (2015), we explore if the Internet is indeed a common source from where the general public draws inflation-related information. In doing so, we employ the internet search data for the keyword "Inflation" and related terms based on the Google Trends data. In the second part, we go one step ahead and investigate if there is Carroll-type "information stickiness" where only a fraction of the population updates their beliefs about future expectations based on information drawn from the Internet. This requires us to calculate inflation expectations based on Google Trends data, following the methodology proposed by Guzman (2011). Next, with the calculated Google Trends based inflation expectations series, we estimate the epidemiological model of Carroll (2003) and assess the extent of prevalence of information stickiness in the economy.

Prior to starting our analysis, we check for the stationarity of all series considered in this work. Based on Augmented Dickey Fuller (ADF), KPSS and Phillips-Perron tests; the CPI inflation, WPI inflation, CPI of professional forecasters, WPI of professional forecasters, the

inflation expectations of households and the inflation expectations derived from Internet search- are all found to be I(1)[6].

Additionally, for each equation estimated in this section and presented in Table 1A through Table 10, we check for the stationarity of the estimated residuals. The estimated residuals are all stationary, thereby indicating that that variables considered are cointegrated and the estimated coefficients of OLS are cointegrating vector. Existence of cointegration implies estimation can be done at level even if individual variables are I(1).

**4.1 Internet as a common source of information**

We begin our analysis by estimating Equation (4) that checks if the Internet is a common source of information for forming inflation expectations. The use of Internet is proxied by the Google Trends data for search keywords like "Inflation", "CPI Inflation", "WPI Inflation", "Core Inflation" and "Headline Inflation". We further categorize the search keywords across three different web search categories, namely, "All Category Web Search", "Business and Industrial Web Search" and "News Web Search". The results of OLS estimations, as well as 2SLS estimation of Equation 4 are presented in Tables 1A through 1C.

[Insert Table 1A about here]

[Insert Table 1B about here]

[Insert Table 1C about here]

Results indicate that the coefficients of Google search of five inflation-related keywords across three web search categories (third column of Tables 1A-1C) are all positive and

---

[6] Results available upon request.

significant. This implies that the Internet, proxied by the Google search statistics of agents, is indeed a common source from where the public draws information to form future expectations about inflation. This conclusion is robust to the choice of inflation-related keyword, the Google search category and the method of estimation.

**4.2 Internet search sentiment**

Having established the role played by Internet search in inflation expectations formation, we next check if the Internet search sentiments affect inflation expectations of both the households and the professional forecasters, as represented by Equation 6. Following Lei et. al. (2015) who had estimated how news sentiment (how news on inflation are reported) affect expectations formed about future inflations, we try to find if search sentiments affect inflation expectations. It is expected that favorable searches should lead to lower inflation expectations while unfavorable searches are associated with higher inflation expectations. Search sentiments are "Favorable" when searches involve terms like "price fall", "price decrease", and "low inflation". "Unfavorable" search sentiment reflects searches like "price rise", "price increase", and "high inflation". Not enough searches were recorded for terms related to "neutral" search, thereby generating no data points by Google Trends.

[Insert Table 2 about here]

Table 2 presents the estimation results of search sentiments based on Equation 6. The upper panel of the table considers survey-based inflation expectations of households as $E_t(\pi_{t+1})$ (the dependent variable in Equation 6), while the lower panel of the same table considers the CPI inflation forecasts by the professional forecasters as $E_t(\pi_{t+1})$. Row 1 of Table 2 indicates that for favourable searches, although the coefficient is of the desired negative sign

14

it is statistically not significant and hence no conclusion can be drawn about positive sentiment reducing inflation expectations of households. Unfavorable searches increase inflation expectations of households since the coefficient is both positive and significant. However, the results are not particularly robust for the method of estimation since 2SLS results (row 2) give the correct coefficient signs for search sentiments but they are statistically not significant.

The lower panel of Table 2 gives partial evidence of the impact of search sentiments on inflation expectations of professional forecasters. Row 3 indicates that positive search sentiment is negative, as desired, and is statistically significant, while the negative search sentiment, although of the correct positive sign, is statistically not significant. However, if 2SLS (row 4) is considered as the method of estimation, then none of the search sentiments are statistically significant.

To conclude the discussion on Internet search sentiment and its impact on inflation expectations, we believe that there is some evidence of the same, since the results are sensitive to the choice of the dependent variable $(E_t(\pi_{t+1}))$ and also to the method of estimation. For those particular specifications that give evidence of the impact of search sentiment on inflation expectations, the inferences drawn here match with Lei et. al. (2015) that finds the similar impact of media reports on inflation expectations.

**4.3 Internet search-based inflation expectations and information stickiness**

Having established that the Internet is indeed a source of information for agents' inflation expectations formation, we next investigate the extent of expectations updation that happens in the economy in a Carroll-like framework. We first estimate the benchmark Carroll (2003)

model represented by Equation 2, followed by our version of the same equation that uses inflation expectations based on Google Trends data.

Table 3 reports the OLS results of the benchmark Carroll (2003) equation, represented by various specifications of Equation 2, where the WPI inflation expectations of the professional forecasters (SPF) are the epidemiological source of inflation expectations for the general public. In all versions of Equation 2, the coefficient of SPF is positive and significant, thereby implying that epidemiological sources of inflation expectations formation exist in the Indian economy. Past inflation expectations by the general public are also positive and significant in all versions of Equation 2, thereby highlighting the contribution of agent's past expectations while the current expectations about future inflation. Past actual inflation (WPI inflation) matters as well for inflation expectations formation, but it has a negative relationship with the latter (row 5). This negative relation with past WPI inflation seems counter-intuitive, but if we look at the households' inflation expectations series vis-à-vis the WPI inflation (Figure 1), we observe that post 2012, the two expectations have moved in opposite directions. When we run the same regressions for a smaller sample between 2008: Q1 and 2013: Q1, lagged value of WPI inflation turns out to be negative and statistically not significant[7].

[Insert Table 3 about here]

We repeat the exercise as done in Table 3, to check for epidemiological sources of expectation formation when the CPI forecasts of the professional forecasters are considered.

[Insert Table 4 about here]

---

[7] Results available upon request.

Similar to the results obtained in Table 3, in Table 4 we find that the coefficient on the CPI forecasts made by the professional forecasters is positive and significant, thereby implying Carroll-type the epidemiological sources of inflation expectations formation by the general public. Past inflation expectations by agents are also positive and significant, thereby indicating that not all agents update their expectations based on current information.

Thus overall, using the two types of inflation numbers (WPI and CPI) provided by the professional forecasters as the source of news for the general public, we do find evidence of Carroll-like epidemiological sources of inflation expectations formation in India.

Next, we depart from the SPF-based Carroll type specification of the epidemiology equation and substitute it by inflation expectations derived from Google Trends. Our contention is that some fraction of the population updates their beliefs based on the inflation-related information that they obtain from the Internet, while the rest of the population continue with their previous beliefs about expected inflation.

We follow the methodology outlined by Guzman (2011) in deriving inflation expectations based on Internet search. The expected inflation series based on Internet search, $(GT\_InfExp)_t$, is calculated as the sum of the directional change in Internet search, that is $[ln(GT_t) - ln(GT_{t-1})]$, where $GT$ is Google Trends data and a proxy for Internet search; and realized inflation from the last period[8]. This is represented by Equation 7 below.

$$(GT\_InfExp)_t = [ln(GT_t) - ln(GT_{t-1})] + \pi_{t-1} \tag{7}$$

---

[8] Here, we assume that in time period $t$, when agents are forming expectations about next period's inflation, they do not have information on current period's official inflation since it is published with a time lag. On the other hand, Google Trends is real-time data, and hence it is present in the current information set.

For a visual reference, we plot the calculated series of inflation expectations based on Google Trends data in Figure 4, against the survey-based expectations of households as well as realized CPI inflation. The correlation between inflation expectations based on Google Trends and survey-based inflation expectations is 0.42 (standard error is 0.13) and it is significant at 1% level of significance, while the correlation between Google Trends based inflation expectations and realized CPI inflation is 0.89 (standard error is 0.07) and it is significant at 1% level of significance. Hence, the Google Trends based inflation expectations series follows realized CPI inflation closely.

The calculated series of inflation expectations based on Google Trends has a mean and standard deviation of 7.68 and 3.01 respectively, between 2006: Q2 and 2018: Q2. Compared to this, the mean and standard deviation of household inflation expectations are 9.66 and 2.51 respectively (during 2006: Q2 to 2018: Q2), of CPI by professional forecasters are 9.29 and 1.78 respectively (during 2008: Q1 and 2018: Q2), and of WPI by professional forecasters are 6.86 and 3.38 respectively (during 2008: Q1 and 2018: Q2). It may be noted that the forecasts by professional forecasters commenced from 2008: Q1.

We test for the nature of expectations (adaptive or rational expectations), and find that inflation expectations based on Google Trends data follows rational expectations.[9]

Once we obtain the $(GT\_InfExp)_t$ series, we next estimate the benchmark Carroll (2003) equation (Equation 2) while replacing the inflation expectations of the professional forecasters by this new series. It is worth noting here, that we do not consider the version of the Carroll equation that includes the $\pi_{t-1}$ term as an explanatory variable, since this term

---

[9] Results available upon request.

$(\pi_{t-1})$ is already embedded in the calculation of inflation expectations based on Internet search, $(GT\_InfExp)_t$.

As in the first part of our analysis, we consider five key search words- "Inflation", "CPI Inflation", "WPI Inflation", "Core Inflation" and "Headline Inflation", across three different search categories- "All Category Web Search", "Business and Industrial Web Search" and "News Web Search", between 2006: Q2 and 2018: Q2. While "All Category Web Search" is general in nature (suppose a student is searching for "what is inflation"), "News Web Search" caters to searches specific within the news section. For our context, news search is more relevant, although we present the result for all categories for robustness check.

Table 5 reports the OLS and 2SLS estimation results for Carroll-type equation where the Internet is deemed as a common source of information for the general public. Hence, inflation expectations based on Google search for the keyword "Inflation" across three search categories is considered in place of inflation expectations of experts (SPF) in the benchmark Carroll model.

[Insert Table 5 about here]

Results indicate that Carroll-type epidemiological sources of inflation expectations, where the Internet is deemed as a common source for drawing information on inflation, is valid in the Indian context. A positive and significant coefficient of $(GT\_InfExp)_t$ gives this evidence (model 1 in each panel of Table 5). Following Carroll (2003), we ran a restricted version model 1 (model 2 in each panel of Table 5) and results indicate that since the coefficient of Google Trends-based inflation expectations is positive and significant at 0.10 across all search categories for both OLS and 2SLS methods, about 10% of the population update their

beliefs about future inflation. On the other hand, the positive and significant coefficients of $E_{t-1}(\pi_t)$ indicate that a significant proportion of the economy do not update their future inflation expectations, thereby indicating the presence of "information stickiness" in the system.

For robustness check, we repeat the estimation exercise done in Table 5, for the various inflation-related searches under specific search categories.

[Insert Table 6 about here]

[Insert Table 7 about here]

[Insert Table 8 about here]

[Insert Table 9 about here]

Tables 6, 7, 8 and 9, present the results for searches for four inflation-related keywords- "WPI Inflation", "CPI Inflation", "Core Inflation" and "Headline Inflation", respectively. As in case of Table 5, all the search words are divided across three search categories and the corresponding $(GT\_InfExp)_t$ is calculated in each case.

Similar to the results in Table 5, we find that irrespective of the search keyword and search category, the epidemiological source of inflation expectations based on Internet search, is positive in all cases and significant in all but eight cases (model 1 for all panels for all tables). A restricted version of model 1 (model 2 for all panels across all tables) indicate that about 9% to 11% of the population update their inflation expectations when we consider the searches in the "All Category Web Search". Inflation expectations formed in the previous period also continue to be positive and significant, thereby implying the agents' stickiness in updating expectations. However, for other categories of search, the restricted version across

various keywords, do not give a statistically significant coefficient for internet search-based inflation expectations.

When compared with the benchmark Carroll (2003) epidemiology-based equation where the forecast of professional forecasters is considered (Table 4), while 14% of the population update their inflation expectations based on the former, for Internet as a source of information, this number is around 10%. Thus, the Internet as a source of information for formation of inflation expectations, comes second to possibly news or newspaper reports (where forecasts of professional forecasters are reported), nevertheless it's impact on future expectations formation is significant and cannot be ignored.

## 4.4 Internet search and inflation expectations gap

In the concluding part of our analysis, we check if there is evidence of increase in Internet search leading to "near-rational" inflation forecasts by the households. In other words, does the increase in the use of Internet search, that might be considered as procuring more information on inflation, reduce the gap between inflation expectations of the general public and that of the professional forecasters?

To test this premise, Carroll (2003) considered a news index based on the first-page coverage of two newspapers in the US of inflation-related reports. He conjectured that with higher inflation-related news reports, the general public would be better informed and thus the gap between their inflation expectations and that of the professional forecasters would narrow down.

Methodology-wise, the news index of Carroll (2003) matches with the nature of the Google Trends data since both are about inflation-related news coverage/Internet search relative to total newspaper reports/total searches conducted. Thus, the explanatory variable in our case is the Google Trends data for the search word "Inflation".

To test our premise of more Internet searches (more information) leading to "near rational" inflation expectations of households, we estimate equation 8 below-

$$|GAP_t| = \alpha + \beta(GT_t) + u_t \qquad (8)$$

Where the dependent variable is the modulus of inflation expectations of households minus the inflation forecasts of professional forecasters and the explanatory variable on the right-hand side is the Google Trends based index.

[Insert Table 10 about here]

Akin to Carroll (2003), results in Table 10 show that the coefficient on Google Trends is negative and significant, for both WPI forecasts and CPI forecasts of professional forecasters, thereby implying that as more Internet searches are conducted, that is equivalent to obtaining more inflation-related information, the inflation expectations gap between the general public and that of the professional forecasters, reduce.

To sum up the main findings, we show that there is evidence of Carroll-type epidemiological sources of inflation expectations formation in India where the Internet is a common source of information for the general public. However, there is presence of "information stickiness" since not all agents update their expectations each period. Additionally, there is some limited

evidence of the impact of Internet search sentiment on inflation expectations of the general public in India.

## 5. Conclusion

This paper explores the epidemiological sources of inflation expectations formation of the general public in India. While Carroll's (2003) benchmark study considers the inflation expectations of professional forecasters as the "common source" based on which people form their expectations, we propose that the Internet is also a potential "common source" of information. Based on data extracted from Google Trends that represents Internet search, our results indicate that during the period 2006 to 2018, the Internet has indeed been a common source of information based on which agents have formed their expectations about future inflation, and the Internet search sentiment has had some impact on inflation expectations. Additionally, based on the inflation expectations series derived from the Google Trends data, we find that there is presence of "information stickiness" in the system since only a small fraction of agents update their inflation expectations each period. Overall, our findings conform to the findings of the epidemiology literature and conclude that the epidemiological sources of inflation expectations formation do exist for the Indian economy.

## Reference

Afkhami, M., Cormack, L., & Ghoddusi, H. (2017). Google Search keywords that best
    predict energy price volatility. *Energy Economics*, 17-27.

Carroll, C. (2003). Macroeconomic Expectations of Households and Professional Forecasters.
    *The Quarterly Journal of Economics, 118*(1), 269-298.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic
    Record, 88*(s1), 2-9.

Dergiades, T., Milas, C., & Panagiotidis, T. (2015). Tweets, Google Trends and Sovereign. *Oxford Economic Papers, 67*(2), 406-432.

Ehrmann, M., Pfajfar , D., & Santoro, E. (2014). Consumer Attitudes and the Epidemiology of Inflation Expectations. *Discussion Paper No. 2014-029, Tilburg University, Center for Economic Research.*

Guzman, G. (2011). Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations. *The Journal of Economic and Social Measurement, 36*(3), 119-167.

Kermack, W. O., & McKendrick, A. O. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 115*(772), 700--721.

Lamla, M., & Maag, T. (2012). The Role of Media for Inflation Forecast Disagreement of Households and Professional Forecasters. *Journal of Money, Credit and Banking, 44*(7), 1325-1350.

Lei, C., Lu, Z., & Zhang, C. (2015). News on inflation and the epidemiology of inflation. *Economic Systems, 39*(4), 644-653.

Mankiw, N. G., & Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics, 117*(4), 1295-1328.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change, 130*, 114-122.

Pfajfar, D., & Santoro, E. (2013). News on Inflation and the Epidemiology of Inflation Expectations. *Journal of Money, Credit and Banking, 45*(6), 1045-1067.
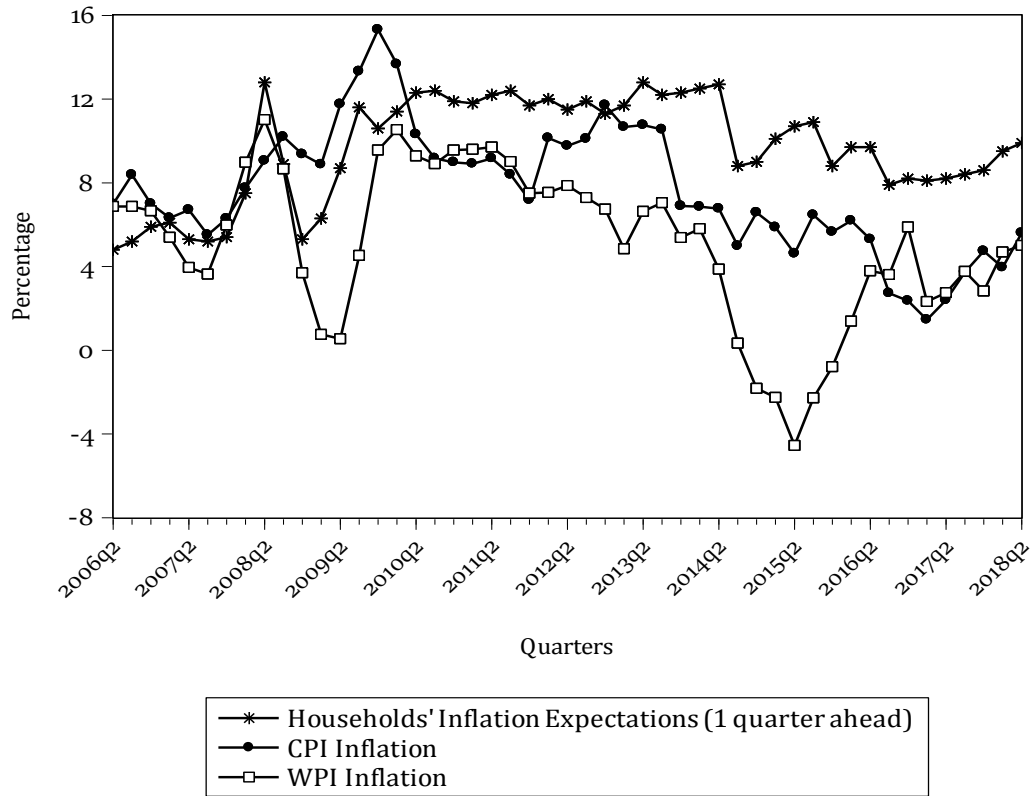
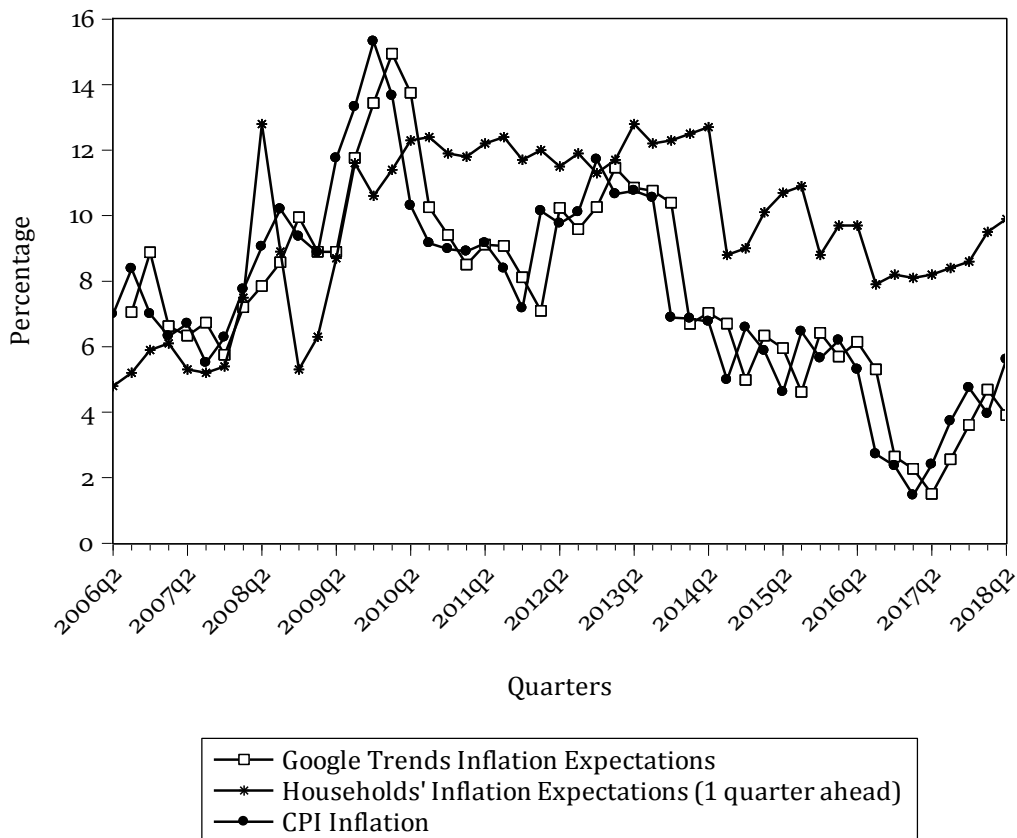**Figure 1: Mean 1-Quarter-Ahead Households' Inflation Expectations and Actual Inflation in India**
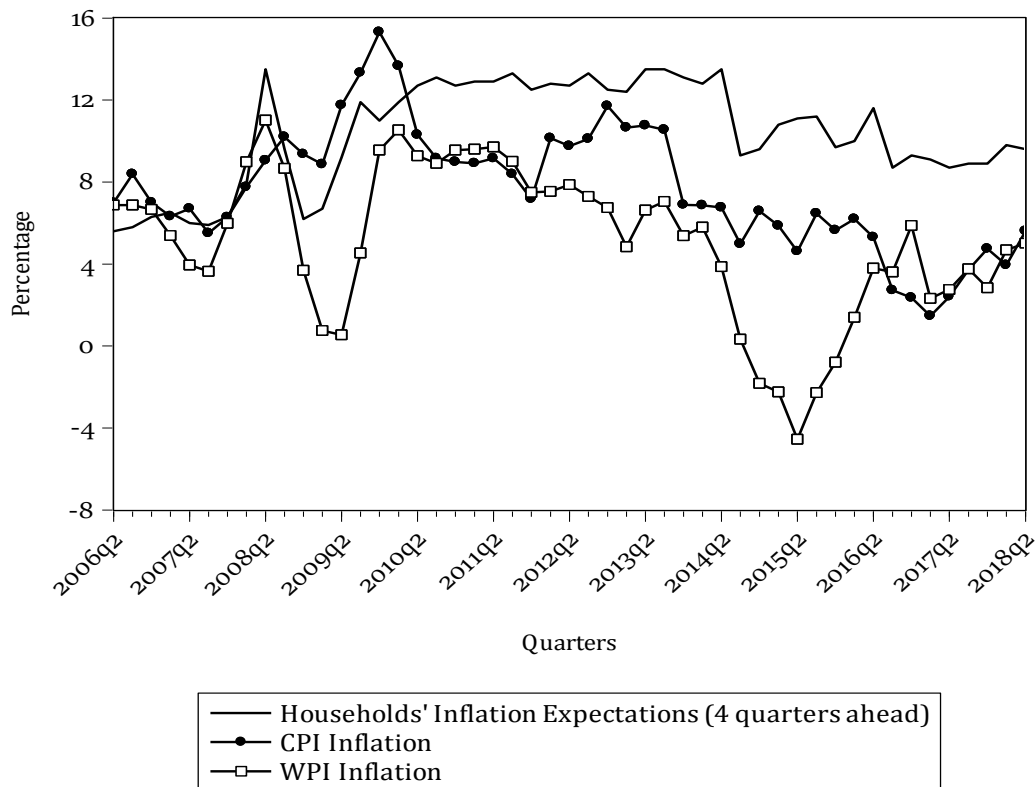


**Figure 2: Mean 4-Quarters-Ahead Households' Inflation Expectations and Actual Inflation in India**
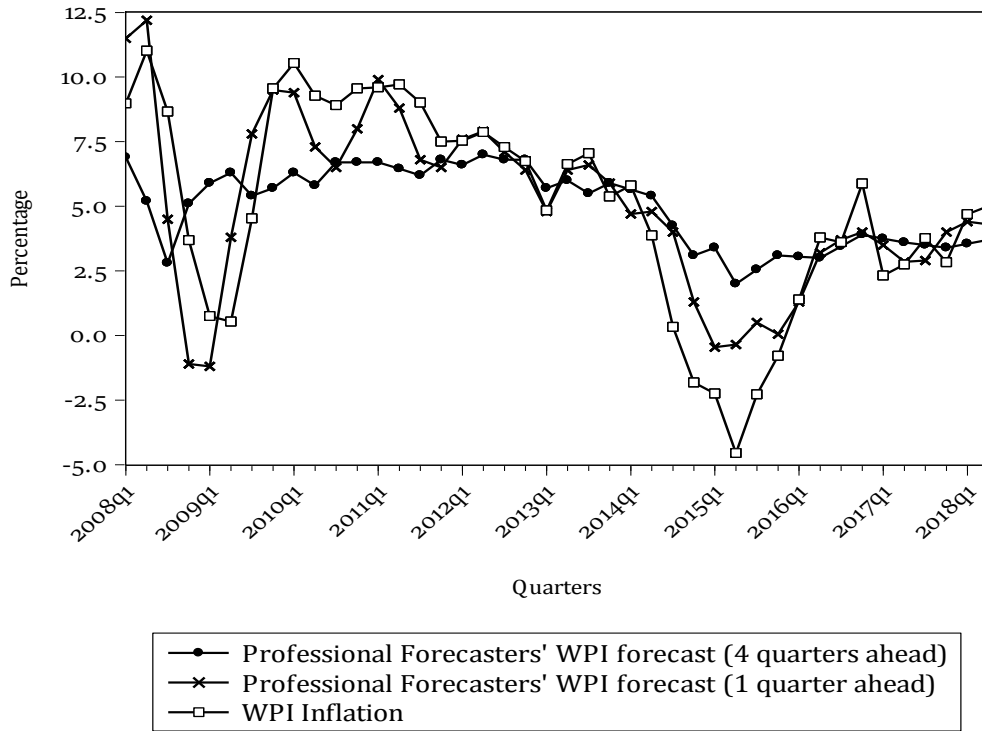
**Figure 3: Survey Professional Forecast of WPI Inflation and Actual WPI Inflation in India**



**Figure 4: Internet Search-based Inflation Expectations, Households Inflation Expectations and Realized Inflation**

**Equation:**

$$E_t \pi_{t+1} = \beta GT_t + \alpha_2 E_{t-1} \pi_t + \mu_t$$

| Google Search Keyword | Method | $\beta$ | $\alpha_2$ | $\bar{R}^2$ | Hansen J test | LM Test | p-DWH |
|---|---|---|---|---|---|---|---|
| Inflation | OLS | 0.86*** (0.05) | 0.49*** (0.19) | 0.65 | -- | 0.02 | -- |
| | 2SLS | 0.87*** (0.05) | 0.47** (0.20) | 0.62 | 0.16 | 0.02 | 0.35 |
| CPI Inflation | OLS | 0.85*** (0.07) | 0.41*** (0.19) | 0.64 | -- | 0.02 | -- |
| | 2SLS | 0.83*** (0.09) | 0.48*** (0.23) | 0.61 | 0.45 | 0.03 | 0.47 |
| WPI Inflation | OLS | 0.74*** (0.08) | 0.74*** (0.22) | 0.65 | -- | 0.08 | -- |
| | 2SLS | 0.75*** (0.08) | 0.72*** (0.24) | 0.61 | 0.09 | 0.05 | 0.48 |
| Core Inflation | OLS | 0.84*** (0.07) | 0.47** (0.23) | 0.65 | -- | 0.02 | |
| | 2SLS | 0.83*** (0.09) | 0.52* (0.27) | 0.62 | 0.26 | 0.01 | 0.97 |
| Headline Inflation | OLS | 0.86*** (0.06) | 0.41** (0.18) | 0.64 | -- | 0.01 | -- |
| | 2SLS | 0.82*** (0.08) | 0.57** (0.25) | 0.59 | 0.83 | 0.02 | 0.21 |

*NOTE:* ***, ** and * stand for 1percent, 5 percent and 10 percent levels of significance respectively. Sample period is 2006:Q2-2018:Q2; instrumental variables in the 2SLS estimations include first two lags of each of the regressors in the model; the standard errors are based on white heteroskedasticity-robust; *p*-values of serial correlation tests up to order two with Breusch–Godfrey LM tests for the regression is mentioned; *p* values of Durbin–Wu–Hausman test (endogeneity test with the null of consistency of the OLS estimator). The p-value Hansen J test with null of over identification is reported.

**Table 1A: Inflation Expectations and the Internet as a Source for "All Categories Web Search"**

**Equation:**

$$E_t \pi_{t+1} = \beta GT_t + \alpha_2 E_{t-1} \pi_t + \mu_t$$

| Google Search Keyword | Method | β | $\alpha_2$ | $\bar{R}^2$ | Hansen J test | LM Test | p-DWH |
|---|---|---|---|---|---|---|---|
| Inflation | OLS | 0.89*** (0.04) | 0.42** (0.18) | 0.64 | -- | 0.02 | -- |
| | 2SLS | 0.90*** (0.04) | 0.39*** (0.19) | 0.61 | 0.15 | 0.01 | 0.38 |
| CPI Inflation | OLS | 0.90*** (0.05) | 0.32** (0.16) | 0.63 | -- | 0.01 | -- |
| | 2SLS | 0.87*** (0.07) | 0.43* (0.22) | 0.57 | 0.14 | 0.02 | 0.27 |
| WPI Inflation | OLS | 0.87*** (0.07) | 0.39* (0.22) | 0.58 | -- | 0.01 | -- |
| | 2SLS | 0.79*** (0.08) | 0.64*** (0.22) | 0.51 | 0.31 | 0.01 | 0.34 |
| Core Inflation | OLS | 0.83*** (0.06) | 0.54*** (0.19) | 0.39 | -- | 0.06 | |
| | 2SLS | 0.84*** (0.08) | 0.50 (0.31) | 0.46 | 0.42 | 0.10 | 0.19 |
| Headline Inflation | OLS | 0.87*** (0.05) | 0.47*** (0.17) | 0.45 | -- | 0.01 | -- |
| | 2SLS | 0.78*** (0.07) | 0.80** (0.32) | 0.32 | 0.28 | 0.21 | 0.10 |

*NOTE:* ***, ** and * stand for 1percent, 5 percent and 10 percent levels of significance respectively. Sample period is 2006:Q2-2018:Q2; instrumental variables in the 2SLS estimations include first two lags of each of the regressors in the model; the standard errors are based on white heteroskedasticity-robust; *p*-values of serial correlation tests up to order two with Breusch–Godfrey LM tests for the regression is mentioned; *p* values of Durbin–Wu–Hausman test (endogeneity test with the null of consistency of the OLS estimator). The p-value Hansen J test with null of over identification is reported.

**Table 1B: Inflation Expectations and the Internet as a Source for Source for "Business and Industrial Web Search"**

**Equation:**

$$E_t \pi_{t+1} = \beta GT_t + \alpha_2 E_{t-1} \pi_t + \mu_t$$

| Google Search Keyword | Method | β | $\alpha_2$ | $\bar{R}^2$ | Hansen J test | LM Test | p-DWH |
|---|---|---|---|---|---|---|---|
| Inflation | OLS | 0.90***<br>(0.04) | 0.40**<br>(0.18) | 0.63 | -- | 0.02 | -- |
| | 2SLS | 0.90***<br>(0.04) | 0.39**<br>(0.19) | 0.60 | 0.07 | 0.02 | 0.77 |
| CPI Inflation | OLS | 0.90***<br>(0.05) | 0.36***<br>(0.17) | 0.63 | -- | 0.01 | -- |
| | 2SLS | 0.87***<br>(0.06) | 0.46***<br>(0.21) | 0.60 | 0.39 | 0.02 | 0.31 |
| WPI Inflation | OLS | 0.89***<br>(0.05) | 0.37**<br>(0.18) | 0.63 | -- | 0.01 | -- |
| | 2SLS | 0.86***<br>(0.07) | 0.49*<br>(0.26) | 0.57 | 0.30 | 0.01 | 0.51 |
| Core Inflation | OLS | 0.88***<br>(0.05) | 0.39***<br>(0.18) | 0.55 | -- | 0.01 | -- |
| | 2SLS | 0.81***<br>(0.05) | 0.64***<br>(0.17) | 0.47 | 0.37 | 0.01 | 0.33 |
| Headline Inflation | OLS | 0.90***<br>(0.05) | 0.34***<br>(0.17) | 0.55 | -- | 0.01 | -- |
| | 2SLS | 0.84***<br>(0.04) | 0.57***<br>(0.15) | 0.46 | 0.24 | 0.02 | 0.36 |

*NOTE:* \*\*\*, \*\* and \* stand for 1percent, 5 percent and 10 percent levels of significance respectively. Sample period is 2006:Q2-2018:Q2; instrumental variables in the 2SLS estimations include first two lags of each of the regressors in the model; the standard errors are based on white heteroskedasticity-robust; p-values of serial correlation tests up to order two with Breusch–Godfrey LM tests for the regression is mentioned; *p* values of Durbin–Wu–Hausman test (endogeneity test with the null of consistency of the OLS estimator). The p-value Hansen J test with null of over identification is reported.

**Table 1C: Inflation Expectations and the Internet as a Source for "News Category Web Search"**

$$E_t(\pi_{t+1}) = c + \alpha E_{t-1}(\pi_t) + \beta^+ GT\_Favorable_t + \beta^- GT\_Unfavorable_t + \mu_t$$

| | Constant | $\alpha$ | $\beta^+$ | $\beta^-$ | $\bar{R}^2$ | p-auto | Method |
|---|---|---|---|---|---|---|---|
| Households Inflation Expectations | 3.02 (2.74) | 0.75*** (0.11) | -0.96 (0.72) | 0.91** (0.43) | 0.69 | 0.08 | OLS |
| | 5.14 (5.25) | 0.73*** (0.11) | -2.03 (1.92) | 1.47 (0.92) | 0.59 | 0.59 | 2SLS |
| Professional Forecasters' CPI forecast | 1.02 (1.85) | 0.85*** (0.08) | -0.62* (0.32) | 0.68 (0.51) | 0.80 | 0.03 | OLS |
| | 4.71 (5.28) | 0.95*** (0.16) | -1.52 (1.36) | 0.28 (1.14) | 0.79 | 0.02 | 2SLS |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Instrumental variables in the 2SLS estimations include first two lags of each of the regressors in the model; the standard errors are based on white heteroskedasticity-robust; *p*-values of serial correlation tests up to order two with Breusch–Godfrey LM tests for the regression is mentioned; sample period for regression of households' inflation expectations span from 2006Q2 to 2018Q2 and for professionals forecast for CPI inflation 2008Q1 to 2018Q2.

**Table 2: Internet Search Sentiment and Inflation Expectations**

**Equation:**

$$E_t[\pi_{t+1}] = \alpha_0 + \alpha_1 SPF\_WPI + \alpha_2 E_{t-1}[\pi_t] + \alpha_3 P_t[\pi_{t-1}] + \epsilon_t$$

| Model | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\overline{R}^2$ | D-W Stats |
|-------|------------|------------|------------|------------|------------------|-----------|
| 1 | -- | 0.26[*] (0.13) | 0.87[***] (0.07) | -- | 0.43 | 1.81 |
| 2 | -- | 0.07 (0.05) | 0.93[***] (0.05) | -- | 0.33 | 1.79 |
| 3 | 4.60[***] (1.00) | 0.30[***] (0.11) | 0.42[***] (0.13) | -- | 0.62 | 1.44 |
| 4 | -- | 0.43[***] (0.13) | 0.90[***] (0.04) | -0.22[***] (0.06) | 0.54 | 1.99 |
| 5 | 4.06[***] (0.95) | 0.43[***] (0.11) | 0.50[***] (0.11) | -0.17[**] (0.08) | 0.68 | 1.67 |
| 6 | -- | -- | 1.00[***] (0.04) | 0.01 (0.06) | 0.27 | 1.87 |

*NOTE:* \*\*\*, \*\* and \*, stand for 1, 5 and 10 percent significance level respectively. Columns 2-5 report the coefficient values for different versions of equation with standard error at parenthesis. All standard errors are corrected for heteroskedasticity and auto correlation. Columns 6-7 report diagnostics for each model. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. $SPF\_WPI_t$ is mean 3 months ahead professional forecast of WPI at time period t. $P_t[\pi_{t-1}]$ is one period lag of actual WPI inflation. The time period for estimation is 2008:Q1 to 2018:Q2. Following Carroll (2003), only OLS estimates have been considered.

**Table 3: Benchmark Epidemiology Model with WPI Inflation Forecasts of Professional Forecasters**

**Equation:**

$$E_t[\pi_{t+1}] = \alpha_0 + \alpha_1 SPF\_CPI + \alpha_2 E_{t-1}[\pi_t] + \alpha_3 P_t[\pi_{t-1}] + \epsilon_t$$

| Model | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\overline{R}^2$ | D-W Stats |
|-------|-----------|-----------|-----------|-----------|------------------|-----------|
| 1 | -- | 0.33*** (0.11) | 0.76*** (0.09) | -- | 0.45 | 1.79 |
| 2 | -- | 0.14* (0.08) | 0.86*** (0.08) | -- | 0.37 | 1.76 |
| 3 | 2.98** (1.18) | 0.28*** (0.09) | 0.52*** (0.14) | -- | 0.52 | 1.58 |
| 4 | -- | 0.57** (0.27) | 0.77*** (0.07) | -0.25 (0.19) | 0.49 | 1.91 |
| 5 | 2.52* (1.40) | 0.45 (0.28) | 0.57*** (0.14) | -0.17 (0.22) | 0.53 | 1.67 |
| 6 | -- | -- | 0.89*** (0.05) | 0.14** (0.06) | 0.32 | 1.87 |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Columns 2-5 report the coefficient values for different versions of equation with standard error at parenthesis. All standard errors are corrected for heteroskedasticity and auto correlation. Columns 6-7 report diagnostics for each model. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. $SPF\_CPI_t$ is mean 3 months ahead professional forecast of CPI at time period t. $P_t[\pi_{t-1}]$ is one period lag of actual CPI inflation. The time period for estimation is 2008: Q1 to 2018: Q2. Following Carroll (2003), only OLS estimates have been considered.

**Table 4: Benchmark Epidemiology Model with CPI Inflation Forecasts of Professional Forecasters**

**Equation:**

$$E_t \pi_{t+1} = c + \beta\, GT_{inf} Exp_t + \alpha\, E_{t-1}\pi_t + \varepsilon_t$$

| Model | Methodology | Constant | α | β | $\bar{R}^2$ | Hansen J test |
|---|---|---|---|---|---|---|
| All Category Web Search | | | | | | |
| 1 | OLS | -- | 0.88*** (0.05) | 0.16*** (0.06) | 0.63 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15*** (0.06) | 0.60 | 0.15 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10*** (0.05) | 0.63 | -- |
| | 2SLS | -- | 0.90*** (0.05) | 0.10** (0.05) | 0.60 | 0.17 |
| 3 | OLS | 1.88** (0.91) | 0.73*** (0.10) | 0.11*** (0.04) | 0.66 | -- |
| | 2SLS | 2.08** (0.97) | 0.72*** (0.10) | 0.10** (0.04) | 0.64 | 0.59 |
| Business and Industrial Web Search | | | | | | |
| 1 | OLS | -- | 0.88*** (0.05) | 0.16*** (0.06) | 0.63 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15*** (0.06) | 0.60 | 0.15 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.63 | -- |
| | 2SLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.60 | 0.17 |
| 3 | OLS | 1.88** (0.90) | 0.73*** (0.10) | 0.11*** (0.04) | 0.66 | -- |
| | 2SLS | 2.08** (0.97) | 0.72*** (0.10) | 0.10** (0.04) | 0.64 | 0.59 |
| News Web Search | | | | | | |
| 1 | OLS | -- | 0.88*** (0.05) | 0.15*** (0.06) | 0.63 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15*** (0.06) | 0.60 | 0.15 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.63 | -- |
| | 2SLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.60 | 0.15 |
| 3 | OLS | 1.89** (0.90) | 0.73*** (0.10) | 0.10*** (0.04) | 0.66 | -- |
| | 2SLS | 2.08** (0.97) | 0.72*** (0.10) | 0.10** (0.04) | 0.64 | 0.58 |

*NOTE:* *** , ** and * stand for 1, 5 and 10 percent significance level respectively. Instruments used for 2SLS estimates are one lag of inflation CPI, inflation primary commodity, inflation fuel group and each regressor. All standard errors are corrected for heteroskedasticity and auto correlation. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. Sample span 2006Q2-2018Q2.

**Table 5: Epidemiology Model with Internet Search-based Inflation Expectations (Search Keyword: "Inflation")**

**Equation:**

$$E_t \pi_{t+1} = c + \beta \, GT_{infWPI} Exp_t + \alpha \, E_{t-1}\pi_t + \varepsilon_t$$

| Model | Methodology | Constant | $\alpha$ | $\beta$ | $\bar{R}^2$ | Hansen J test |
|---|---|---|---|---|---|---|
| All Category Web Search | | | | | | |
| 1 | OLS | -- | 0.88*** (0.05) | 0.16*** (0.06) | 0.43 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15*** (0.06) | 0.34 | 0.03 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.43 | -- |
| | 2SLS | -- | 0.91*** (0.05) | 0.09* (0.05) | 0.34 | 0.03 |
| 3 | OLS | 3.20*** (0.73) | 0.60*** (0.08) | 0.13*** (0.04) | 0.53 | -- |
| | 2SLS | 3.71*** (0.93) | 0.55*** (0.11) | 0.13** (0.06) | 0.48 | 0.44 |
| Business and Industrial Web Search | | | | | | |
| 1 | OLS | -- | 0.87*** (0.06) | 0.17** (0.07) | 0.44 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15** (0.06) | 0.35 | 0.03 |
| 2 | OLS | -- | 0.93*** (0.05) | 0.07 (0.05) | 0.38 | -- |
| | 2SLS | -- | 0.93*** (0.05) | 0.07 (0.05) | 0.22 | 0.01 |
| 3 | OLS | 3.28*** (0.74) | 0.57*** (0.09) | 0.16*** (0.03) | 0.55 | -- |
| | 2SLS | 3.82*** (0.97) | 0.53*** (0.11) | 0.14*** (0.05) | 0.50 | 0.79 |
| News Web Search | | | | | | |
| 1 | OLS | -- | 0.87*** (0.06) | 0.17** (0.08) | 0.33 | -- |
| | 2SLS | -- | 0.88*** (0.06) | 0.16** (0.07) | 0.32 | 0.08 |
| 2 | OLS | -- | 0.91*** (0.05) | 0.09 (0.06) | 0.12 | -- |
| | 2SLS | -- | 0.92*** (0.05) | 0.08 (0.05) | 0.08 | 0.01 |
| 3 | OLS | 3.77*** (0.98) | 0.53*** (0.11) | 0.15** (0.06) | 0.47 | -- |
| | 2SLS | 3.80*** (1.19) | 0.53*** (0.14) | 0.15* (0.07) | 0.44 | 0.94 |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Instruments used for 2SLS estimates are one lag of inflation CPI, inflation primary commodity, inflation fuel group and each regressor. All standard errors are corrected for heteroskedasticity and auto correlation. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. Sample span 2006Q2-2018Q2.

**Table 6: Epidemiology Model with Internet Search-based Inflation Expectations (Search Keyword: "WPI Inflation")**

**Equation:**

$$E_t \pi_{t+1} = c + \beta\, GT_{infCPI} Exp_t + \alpha\, E_{t-1} \pi_t + \varepsilon_t$$

| Model | Methodology | Constant | α | β | $\bar{R}^2$ | Hansen J test |
|---|---|---|---|---|---|---|
| All Category Web Search | | | | | | |
| 1 | OLS | -- | 0.87*** (0.06) | 0.17*** (0.06) | 0.64 | -- |
| | 2SLS | -- | 0.89*** (0.05) | 0.14** (0.06) | 0.61 | 0.12 |
| 2 | OLS | -- | 0.89*** (0.06) | 0.11* (0.06) | 0.63 | -- |
| | 2SLS | -- | 0.91*** (0.05) | 0.09* (0.05) | 0.60 | 0.14 |
| 3 | OLS | 1.90** (0.89) | 0.71*** (0.10) | 0.13*** (0.04) | 0.67 | -- |
| | 2SLS | 2.16** (0.96) | 0.71*** (0.10) | 0.11*** (0.04) | 0.64 | 0.61 |
| Business and Industrial Web Search | | | | | | |
| 1 | OLS | -- | 0.92*** (0.05) | 0.11* (0.06) | 0.56 | -- |
| | 2SLS | -- | 0.95*** (0.05) | 0.05 (0.05) | 0.63 | 0.04 |
| 2 | OLS | -- | 0.95*** (0.04) | 0.05 (0.04) | 0.55 | -- |
| | 2SLS | -- | 0.96*** (0.04) | 0.04 (0.04) | 0.64 | 0.06 |
| 3 | OLS | 4.16*** (0.97) | 0.46*** (0.12) | 0.21*** (0.05) | 0.68 | -- |
| | 2SLS | 3.59*** (1.19) | 0.54*** (0.16) | 0.18** (0.07) | 0.69 | 0.24 |
| News Web Search | | | | | | |
| 1 | OLS | -- | 0.88*** (0.07) | 0.17* (0.09) | 0.55 | -- |
| | 2SLS | -- | 0.92*** (0.05) | 0.11 (0.07) | 0.56 | 0.01 |
| 2 | OLS | -- | 0.94*** (0.05) | 0.06 (0.05) | 0.52 | -- |
| | 2SLS | -- | 0.96*** (0.04) | 0.04 (0.04) | 0.55 | 0.01 |
| 3 | OLS | 3.97*** (0.62) | 0.47*** (0.08) | 0.23*** (0.04) | 0.70 | -- |
| | 2SLS | 4.55*** (0.85) | 0.39*** (0.11) | 0.27*** (0.05) | 0.69 | 0.30 |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Instruments used for 2SLS estimates are one lag of inflation CPI, inflation primary commodity, inflation fuel group and each regressor. All standard errors are corrected for heteroskedasticity and auto correlation. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. Sample span 2006Q2-2018Q2.

**Table 7: Epidemiology Model with Internet Search-based Inflation Expectations (Search Keyword: "CPI Inflation")**

$$E_t \pi_{t+1} = c + \beta \, GT_{infCore} Exp_t + \alpha \, E_{t-1}\pi_t + \varepsilon_t$$

| Model | Methodology | Constant | α | β | $\bar{R}^2$ | Hansen J test |
|-------|-------------|----------|---|---|-------------|---------------|
| **All Category Web Search** | | | | | | |
| 1 | OLS | -- | 0.89*** (0.04) | 0.14*** (0.05) | 0.63 | -- |
|   | 2SLS | -- | 0.88*** (0.05) | 0.15*** (0.05) | 0.60 | 0.09 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.63 | -- |
|   | 2SLS | -- | 0.90*** (0.05) | 0.10** (0.05) | 0.60 | 0.11 |
| 3 | OLS | 1.88** (0.91) | 0.74*** (0.09) | 0.09** (0.04) | 0.66 | -- |
|   | 2SLS | 2.02** (0.94) | 0.72*** (0.10) | 0.10*** (0.04) | 0.63 | 0.31 |
| **Business and Industrial Web Search** | | | | | | |
| 1 | OLS | -- | 0.86*** (0.09) | 0.18 (0.11) | 0.32 | -- |
|   | 2SLS | -- | 0.85*** (0.09) | 0.19 (0.12) | 0.47 | 0.06 |
| 2 | OLS | -- | 0.81*** (0.06) | 0.09 (0.06) | 0.32 | -- |
|   | 2SLS | -- | 0.90*** (0.06) | 0.10 (0.06) | 0.49 | 0.06 |
| 3 | OLS | 3.99*** (1.22) | 0.48*** (0.16) | 0.20*** (0.07) | 0.46 | -- |
|   | 2SLS | 3.15*** (1.00) | 0.53*** (0.14) | 0.22** (0.10) | 0.54 | 0.39 |
| **News Web Search** | | | | | | |
| 1 | OLS | -- | 0.83*** (0.09) | 0.23** (0.12) | 0.35 | -- |
|   | 2SLS | -- | 0.83*** (0.08) | 0.23** (0.11) | 0.33 | 0.05 |
| 2 | OLS | -- | 0.92*** (0.06) | 0.09 (0.06) | 0.33 | -- |
|   | 2SLS | -- | 0.92*** (0.05) | 0.08 (0.05) | 0.32 | 0.05 |
| 3 | OLS | 3.76*** (0.92) | 0.49*** (0.12) | 0.21*** (0.07) | 0.49 | -- |
|   | 2SLS | 3.86*** (1.23) | 0.47*** (0.16) | 0.23** (0.10) | 0.46 | 0.75 |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Instruments used for 2SLS estimates are one lag of inflation CPI, inflation primary commodity, inflation fuel group and each regressor. All standard errors are corrected for heteroskedasticity and auto correlation. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. Sample span 2006Q2-2018Q2.

**Table 8: Epidemiology Model with Internet Search-based Inflation Expectations (Search Keyword: "Core Inflation")**

**Equation:**

$$E_t \pi_{t+1} = c + \beta\, GT_{infHead}Exp_t + \alpha\, E_{t-1}\pi_t + \varepsilon_t$$

| Model | Methodology | Constant | α | β | $\bar{R}^2$ | Hansen J test |
|---|---|---|---|---|---|---|
| **All Category Web Search** | | | | | | |
| 1 | OLS | -- | 0.88*** (0.05) | 0.15** (0.06) | 0.60 | -- |
| | 2SLS | -- | 0.88*** (0.05) | 0.15** (0.06) | 0.57 | 0.17 |
| 2 | OLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.60 | -- |
| | 2SLS | -- | 0.90*** (0.05) | 0.10* (0.05) | 0.57 | 0.19 |
| 3 | OLS | 2.06** (0.98) | 0.72*** (0.11) | 0.10** (0.05) | 0.63 | -- |
| | 2SLS | 2.17** (1.03) | 0.70*** (0.11) | 0.11** (0.05) | 0.61 | 0.66 |
| **Business and Industrial Web Search** | | | | | | |
| 1 | OLS | -- | 0.94*** (0.07) | 0.07 (0.08) | 0.62 | -- |
| | 2SLS | -- | 0.96*** (0.05) | 0.04 (0.06) | 0.60 | 0.06 |
| 2 | OLS | -- | 0.96*** (0.05) | 0.04 (0.05) | 0.63 | |
| | 2SLS | -- | 0.96*** (0.04) | 0.04 (0.04) | 0.61 | 0.09 |
| 3 | OLS | 4.22*** (1.21) | 0.43** (0.18) | 0.26** (0.10) | 0.71 | -- |
| | 2SLS | 4.54*** (1.17) | 0.38** (0.17) | 0.28*** (0.09) | 0.70 | 0.40 |
| **News Web Search** | | | | | | |
| 1 | OLS | -- | 0.95*** (0.07) | 0.05 (0.09) | 0.60 | -- |
| | 2SLS | -- | 0.96*** (0.06) | 0.04 (0.07) | 0.57 | 0.02 |
| 2 | OLS | -- | 0.94*** (0.07) | 0.06 (0.07) | 0.54 | -- |
| | 2SLS | -- | 0.99*** (0.03) | 0.01 (0.03) | 0.75 | 0.05 |
| 3 | OLS | 4.01*** (1.43) | 0.47** (0.20) | 0.22** (0.11) | 0.68 | -- |
| | 2SLS | 4.38*** (1.41) | 0.41** (0.20) | 0.26*** (0.10) | 0.66 | 0.11 |

*NOTE:* ***, ** and * stand for 1, 5 and 10 percent significance level respectively. Instruments used for 2SLS estimates are one lag of inflation CPI, inflation primary commodity, inflation fuel group and each regressor. All standard errors are corrected for heteroskedasticity and auto correlation. $E_t[\pi_{t+1}]$ is mean 3 months ahead inflation expectations survey of households at time period t. Sample span 2006Q2-2018Q2.

**Table 9: Epidemiology Model with Internet Search-based Inflation Expectations (Search Keyword: "Headline Inflation")**

**Equation:**

$$|GAP_t| = \alpha + \beta(GT_t) + u_t$$

| | $\alpha$ | $\beta$ | $\bar{R}^2$ | D-W Stats |
|---|---|---|---|---|
| Professional Forecasters' CPI forecast | 6.56*** (0.97) | -1.12*** (0.35) | 0.25 | 1.36 |
| Professional Forecasters' WPI forecast | 10.54*** (2.15) | -1.76*** (0.65) | 0.22 | 0.68 |

*NOTE:* ** stand for 5 percent level of significance. Time period for estimation is 2008Q1 to 2018Q2. D-W stats indicate Durbin Watson statistics for autocorrelation. $GAPM$ is the modulus of the difference between the households' inflation expectations and that of the professional forecasters. All standard errors are corrected for heteroskedasticity and auto correlation. $GT_t$ here indicates Google Trends search for the word "Inflation" under all categories.

**Table 10: Internet Search and Inflation Expectations Gap**