Constructing a Consumer Confidence Index for the US Using Web Search Volume

Nicolás Della Penna and Haifang Huang
August, 2009

(Comments welcome)

Abstract:

We construct consumer sentiment indexes for the US using the popularity trends of six types of Google searches. The resulted indexes highly correlate with the University of Michigan's Index of Consumer Sentiment (ICS) and the Conference Board's Consumer Confidence Index (CCI). In term of forecasting consumer spending, the search-based indexes are at least as powerful as the combined force of ICS and CCI, if not better. For robustness, we subject our indexes to alternative weights of combining the Google information, to alternative measures of consumer spending, and to alternative statistical specifications. The finding is robust.

1. Introduction

About seventy percents of the US GDP is personal consumption. Variations in consumption therefore drive business cycles in the country. Consumption also reflects the population's expectation about the future; consumers' spending thus has forecasting value. As a result, the release of consumption data in the US often affects stock market. A recent example is the weaker-than-expected retail sales reported by the Commerce Department for the July of 2009. The stock market dropped the next day despite the upbeat assessment of the economy by the Federal Reserve just a few days ago. A headline on MSNBC.com declares: "On Wall Street, Shoppers Trump the Fed".[1]

Practitioners have developed various sentiment indexes to keep track of consumer's willingness to spend. In the US the two major indexes are the University of Michigan's Index of Consumer Sentiment (hereafter ICS) and the Conference Board's Consumer Confidence Index (hereafter CCI). They are survey-based measures that intend to gauge consumers' confidence in the economy. The monthly release of these indexes also affects markets. Take for examples the article headings of Bloomberg right after the August 2009 release of the ICS: "Euro Tumbles as US Consumer Confidence Outweighs Region's GDP", and "German Stocks Drop as US Consumer Sentiment Unexpectedly Falls".

We believe variations in Internet search patterns also reflect consumer sentiment. There are two reasons for our view. First, some people use Internet to look for, or do researches on, products they want to purchase; changes in search pattern thus reflect changes in demand. We believe certain demands, say those for luxury goods, are revealing of consumers' purchasing power and confidence. Second, people use Internet to do research on issues that bother them, such as debt burden and energy costs, both of which are thought to be important factors behind consumer's purchasing power.

In this paper we construct consumer-sentiment indexes using the popularity measure of searches on Google provided by "Google Insight for Search". We examine the power of search-based indexes in forecasting consumer spending in the US. We further compare the forecasting power to that of the CCI and the ICS. In the available sample between January 2004 and May 2009, we find Google indexes generally outperform the combined force of CCI and ICS. Given the short sample, we conservatively conclude that our indexes are as good as the competitors. The Google-search indexes, however, have a unique and important advantage: the underlying data are free and are available on weekly base, as opposed to monthly. We thus recommend using Google search as major source of information about consumer confidence, for their accuracy and for the fact that Google indexes can be obtained before the monthly release of CCI and ICS.

The search-based indexes are weighted averages of the popularity of six types of Google searches. We call these types as "components". These six components, we believe, reflect important causes behind, and symptoms of, changes in consumer sentiments. The first component is related to households' debt burden; the second component is related to cost of energy and utility; the third group reflects business conditions. We classify these three into the group that "causes" the changes in sentiment. We then classify the other three components into the group of "symptoms": the reflection of changing sentiment regardless the causes. They are searches that are related to luxury goods, credit products, and insurance.

We need relative weights to combine the six components into indexes of lower dimension. We have two sources of weights. The first one is exogenous to US; the second one is the preferred weight and is derived from US information. The first set of weights comes from an international sample. The weights are the estimated relation between Google searches and retail sales in the sample. We exclude US so that when we construct the index for US, we are doing so "out of sample". We also exclude all information after December 2007. This is to reduce the likelihood that the fit of the resulted index in the US is due to the worldwide nature of the economic crisis in 2008 and 2009. Our preferred weights come from the second source. They are the estimated relation, in a panel of 49 US states, between changes in Google searches and changes in retail-trade employment. We want the weights to contain as much information as possible; for this purpose we use the entire available sample from January 2004 to May 2009. We hope that the first set of weight lends credibility to the information content of Google indexes. We recommend the second set of weights to be used for constructing the sentiment indexes from Google searches.

We are not aware of papers that use Internet searches to measure consumer sentiment. There are papers that relate Google searches to other interests in economic study. Da, Engelberg & Gao (2009) find the aggregate search frequency in Google for stocks is correlated with existing proxies of investor attention, and there is "stronger price momentum among stocks with higher level of [Google Search Volume Index]". Askitas and Zimmermann (2009) finds "strong correlations between keyword searches and unemployment rates using monthly German data…"

Among earlier literature, our paper is related to those that examine the information content of ICS and CCI in forecasting consumer spending and other macro variables. Carroll, Fuhrer, & Wilcox (1994) finds that “lagged values of the ICS … explain … variation in the growth of total real personal consumption expenditures …” Howrey (2001) reaffirms the forecasting power of ICS for personal consumption, and finds that “the ICS is a useful recession indicator variable.” His finding echoes an earlier finding by Matsusak and Sbordone (1995) that the “hypothesis that consumer sentiment does not cause GNP (in the Granger sense) can be rejected.” Bram and Ludvigson (1998) conducted a horse race between the ICS and the Conference Board’s CCI. They find CCI to “have both economically and statistically significant explanatory power for several spending categories …” while the ICS “exhibit weaker forecasting power…” Our paper does not take a position regarding the relative power of ICS and CCI, we compares the search-based indexes to the CCI and ICS together.

The structure of the paper is as follows: In section 2 we explain the nature of the Google data; in section 3 we select and classify the Google data into components of our sentiment indexes; in section 4 we explain how we obtain the weights to combine the components into the indexes. In section 5 we test the predictive power of the resulted indexes, and engage them in a competition against the combined power of CCI and ICS. We conclude in section 6.

2. Google Insights for Search and its measure of popularity of searches

Through “Google Insights for Search”, Google.inc has made available the information on its users’ internet search pattern. According to the company, the data enable the public to “compare search volume patterns across specific regions, categories, time frames and properties.” The data do not really reflect absolute volume; instead they reflect popularity, or “the likelihood of users in a particular area to search for a term on Google on a relative basis” in Google’s words. More specifically, when Google’s data show that the term “hotel” is equally popular in Canada and Fiji, it does not mean the absolute volume of searches is the same. It means “users in both Fiji and Canada are equally likely to search for the term hotel”. Furthermore, even if the absolute search volume for “hotel” is constant over time in Fiji, Fiji’s popularity measure for “hotel” can still fluctuate if Fiji residents’ overall search volume changes.

The data we choose to use in this paper has been further normalized by initial popularity. The value for initial date (usually January 2004) are always zero; the data after January 2004 measure “…percentage of growth [of the popularity measure], with respect to the first date…”.[2] A positive (negative) value for “hotel” in September 2005 means Google users are more (less) likely to search for “hotel” in September 2005 than they were in January 2004; the numerical value reflects the difference as percentage points of the initial popularity measure.[3]

Google has classified its users’ internet searches into more than six hundred “categories” and subcategories. A few examples of the categories include “Shopping > Luxury Goods”, “Society > Legal > Bankruptcy”, “Science > Ecology”, “Recreation > Hobbies > Paintball”, and “Society > Government & Regulatory Bodies > Visa & Immigration”.

We only use data at the level of categories, which means that we do not pay attention to what key words were used in the search: if a search was deemed by Google.inc as belonging to the category "Shopping > Luxury Goods", we treat it as so. The data we use are all from URL: http://www.google.com/insights/search/#.

Google Insight provides data at weekly frequency. This means consumer indexes based on Google searches are timelier than the ICS and CCI. Both of the latter are monthly releases. In an effort of highlighting this unique advantage of Google information, we construct our indexes in a manner that they always come out before the monthly releases of Conference Board's CCI. The release date for CCI is usually in the last week of the current month, the earliest date being the 24^{th} of the month. We choose our cutoff date for our month series so that the Google indexes always become available before the 24^{th} of a month. This means that our indexes are always available before CCI. For example, our April 2009 indexes are aggregated from Google searches between March 22, 2009 (the previous cutoff date) and April 18, 2009. The release date of CCI for that month is April 28. So our index preempts the CCI by 10 days in this case.

3. The six components of our search-based sentiment indexes

Our search-based indexes of consumer sentiment are essentially weighted averages of six components. Each component measures the trends in the popularity of certain category(ies) of Google searches. The six components contain information from totally eleven categories. The selection and classification reflect our hypotheses about the causes behind, and the symptoms of, changes in consumer confidence. We loosely place the six components, and their associated categories, into two groups: "causes" and "symptoms".

3.1. The group of "causes": We use this group to reflect causes behind variations in consumer confidence. There are three components in this group: 1) searches that reflect concerns and/or consequence of households' debt burden; 2) searches that reflect concerns and/or consequence of changes in energy costs; 3) searches that reflect business conditions. We include "debt burden" based on the belief that rising debt burden reduces consumers' purchasing power and confidence. The inclusion of "energy cost" reflects the frequently expressed view in media, and by politicians, that rising gas prices and cost of heating oil are burdens on households.[4] We include the last component, "business conditions", because private businesses are the main source of income for workers, owners, and stockholders. We expect changes in business conditions to affect consumer spending, and we are not alone: both the University of Michigan and Conference Boards use survey respondents' evaluation of business conditions as part of their sentiment indexes.[5]

The following table describes the search categories in this group:

1. "Debt burden" has the following categories
    a. Society > Legal > Bankruptcy
    b. Finance & Insurance > Credit & Lending > Debt Management
2. Energy costs has the following categories
    a. Industries > Energy & Utilities > Oil & Gas

    b. Industries > Energy & Utilities > Alternative Energy
    c. Automotive > Hybrid & Alternative Vehicles
3. “Business condition” has the following categories
    a. Business > Office & Printing Services > Office Supplies
    b. Business > Office & Printing Services > Office Furniture

3.2. The group of “symptoms”: We presume that some consumers use Google to do researches on their targeted purchases; changes in the popularity of searches thus convey information about demand. Take for example a decline in popularity of searches for luxury goods. We view it as a symptom that consumers are suffering from various ills that make them reluctant to spend on discretionary items. We do not have an exhaustive list of “symptoms” in this group. There are only three components: 1) searches related to shopping for luxury goods; 2) searches related to credit products (credit card, auto loan, student loans, etc); 3) searches related to insurance products. Our hypotheses are that searches for luxury goods reflect purchasing power and willingness; searches for credit products reflects demand for credit thus consumers’ confidence to repay; search for insurance reflect heightened sense of risk aversion. We acknowledge there is substantial ambiguity. Take for example the credit products. It is possible that rising searches for credit cards and student loans are result of worsened credit conditions, which forces borrowers to actively seek, and actively research on, credit providers. It is an empirical question how the searches for credit products correlate with consumption or other measures of consumer confidence.

The following table describes the categories in this group:

4. Demand for luxury goods includes categories
    a. Shopping > Luxury Goods
5. Demand for credit includes categories
    a. Finance & Insurance > Credit & Lending > Student Lending
    b. Finance & Insurance > Credit & Lending > Credit Cards
    c. Automotive > Auto Financing
6. Demand for insurance includes categories
    a. Finance & Insurance > Insurance

4. Weights we use to construct Google-based sentiment indexes

We need to combine various components and categories into indexes. The purpose is to reduce dimension while retaining important information. The resulted index is thus easier to present. University of Michigan and the Conference Board use similar strategy: their overall indexes of consumer sentiment/confidence are both built from five survey questions.

We need to aggregate categories into components, and then from components to indexes. For the first step we adopt a simplistic approach: we use simple averages whenever a component has more than one category. So in the end we have six time series. We then take weighted sums of the six. The source of the weights is the main subject in this section.

We choose to estimate the weights from simple regressions that use consumption-related data as dependent variables, and Google searches as explanatory variables. The rationale for using estimation-based weights is necessity. If there was a large empirical literature on the relation between internet searches and consumption behavior, we would have chosen the weights from micro evidences. If there was a well-accepted theory on searches and consumption, we could have the option to calibrate the weights based on theoretical predictions. But the literature is not established yet. We are not aware of existing papers that explore the link between internet searches and consumption. So regression is practically our only choice.

We have the option to regress the time series of US national consumption on the six components (and their lags), and then use the estimated coefficients as weights. But we quickly run into the problem of short time span. The highest frequency for the time series of consumption and retail sales is monthly. The maximum number of months from Google Insight, as we are drafting the paper, is 66 (from January 2004 to June 2009). The short sample raises the concern of over fitting within the sample: the estimated weights can fit the sample well but perform poorly out of sample. We choose not to pursue this route. Instead, we look for larger sample for more robust estimates. For this we look internationally, and at the state level in the US.

The international sample is available for most of the developed countries. Google Insight allows the public to download search data at country level, although not all countries have detail categorization. Also available is the monthly data of retail sales for OECD countries provided by OECD.StatExtracts, which also has the retail data for some major non-OECD countries. We downloaded the biggest set of countries we can get. The total number of countries for which we have all necessary data is 13, including the US, Canada, United Kingdom, Germany, France, Japan, Italy, Netherland, Sweden, Austria, Poland, Russia, and Brazil.

In our effort to focus on out-of-sample properties, we will discard US data, as well as the data for all countries after December 2007. We hope to use this set of weights to show that Google indexes have predictive power in the US, even when the weights are exogenous to US, estimated from non-US samples and before the world wide economic crisis. This set of weights is not our preferred one. Our preferred source of weights is the next one: the sample of 49 US states.

We have Google searches for all 50 US states plus DC. The availability of consumption data is more problematic. There are no public-available monthly data of personal consumption expenditure or retail sales that cover most states. But the Bureau of Labor Statistics does publish monthly data on retail-trade employment except for Texas and Louisiana. We will use retail-trade employment as the proxy for state-level consumption, under the presumption that there is link between the two at the business cycle frequency.

While we are willing to assume that changes in consumption affect employments in the retail sector, we do not know the length of the lag from consumption to employment. They could occur in the same month if the sector's labor market is flexible. If not, the transmission process could take longer time. We therefore use two dependent variables. The first one is the current-month growth in the employment; the other is the sum of growths in the next-six months. The first dependent variable gives us a set of weights that are the most capable of capturing the link between Google searches and immediate changes in retail employment. We call the set the "short-run weights". The second DV, the sum of growth in the next six month, gives us a set of weights that are the most capable of capturing the link in the longer run. We call the set the "long-run weights".

Two set of weights of means we have two indexes from the Google information. We find both indexes helpful in forecast consumption and retail sales in the US. So we decide against further reducing the dimension. We use both indexes to forecast consumer spending. Considering that we have more than ten categories of Google information to start with, we believe that the final dimension of two is an appropriate tradeoff between dimensionality and information contents.

We now focus on the international sample. It is going to gives us weights that are exogenous to the US. The first two columns of Table 1 present the weights from the international sample; the first column has the short-run weights (the estimated coefficients of Google variables when the DV is the current-month growths); the second column has the long-run weights (when the DV is growth over the next six months). Recall that in this sample we exclude US and all information after December 2007. The regression method is fixed-effect panel regression. The dependent variable of the first column is the month-to-month growth in the volume of retail sales. The dependent variable of the second columns is the sum of growth rates in the next six months. The explanatory variables are the same for both columns. They are the constant term plus Google searches. For the Google variables, we include the current values of the six search popularity minus their lagged three-month averages, as well as the lagged averages themselves. This way both the levels and the changes enter the regressions.

The exact regression models are shown here

$$Gr_t = const + \sum_{i=1}^{6} \hat{w}^s_i G_{i,t} + \sum_{i=1}^{6} \hat{w}^s_{i,lag}\left(\frac{G_{i,t-1} + G_{i,t-2} + G_{i,t-3}}{3}\right) + u_{1t}$$

$$\sum_{j=1}^{6} Gr_{t+j} = const + \sum_{i=1}^{6} \hat{w}^l_i G_{i,t} + \sum_{i=1}^{6} \hat{w}^l_{i,lag}\left(\frac{G_{i,t-1} + G_{i,t-2} + G_{i,t-3}}{3}\right) + u_{2t}$$

The term *Gr* is the monthly growth; the term $G_{i,t}$ are the value of the i^{th} component of Google searches at time *t*. The estimated coefficients are denoted as $\hat{w}$ because they will be used as "*w*"eights to construct the index for the US. They have a superscript that is either "*s*" for "short-run" weights, or "*l*" for "long-run" weights. Their first subscript "*i*" indexes the six components. Their second subscript, if exists at all, is "*lag*", and it tells that the coefficients is associated with the lagged-averages.[6]

The short-run weights (i.e., estimated coefficients in the first column) are almost all insignificant. The only significant variable is the change in the popularity of "energy cost". It has the expected negative sign. The situation is better for the long-run weights in the second column (when the DV is growths in the next six months). In this column, two types of internet searches have the statistical significance below 10%, and the expected signs. They are searches related to energy cost and those related to luxury goods. In both cases, the lagged levels and the changes from the lagged level have positive coefficients. We want to take note that these two types of searches also have the "right" signs in the first column, albeit largely insignificant.

Of the two columns in Table 1, the only type of Google searches that has unexpected sign and has statistical significance is demand for credit. We expect the level to attract a positive coefficient but the opposite occurs.

To help understanding quantitative importance of different type of Google searches, the lower portion of Table 1 multiplies estimated weights (coefficients) with the standard deviation of the corresponding variable in the US. The resulted numerical value should be understood as the impact on growth rates from one standard deviation of the variations. It turns out that the dominant forces (we use the sum of volatility-adjusted coefficients as a crude criterion) are energy cost and luxury goods. They happen to be the ones that have the statistical significance and the expected signs. All other types of searches are lagging far behind. This suggests that the US index constructed from the international weights will be primarily driven by searches related to energy costs, and then by search for luxury goods.

We now move on to the second sources of weights. The third and the fourth columns of Table 1 are from the sample of US 49 states (including DC, but without Texas and Louisiana for lack of data). The data cover the entire available period between January 2004 and May 2009. The third column has the short-run weights (i.e., its DV is the growth in retail-trade employment from the last month to the current month). The fourth column has the long-run weights (its DV is the sum of the growths in next six months). The explanatory variables are the same set of Google variables we discussed before.[7]

The state-level sample delivers much stronger result. The adjusted R2 for explaining current-month growth is 19%; that for explaining the sum of growth in next six months is 55%. An overwhelming majority of the weights (i.e., estimated coefficients) are statistically significant; many at below 1%; nearly all have the expected signs. The popularity of searches related to debt burden is associated with lower growths; so is the popularity of energy cost. Searches for office furniture and supply are strongly positive for longer term growth. Searches for luxury goods are strongly positive for growth in current month. Searches for credit products have positive coefficients both in the short run and in the long run; searches for insurance have the opposite signs. There is only one unexpected sign among the twenty-four weights, and it is small judging by volatility-adjusted coefficients.

The volatility-adjusted coefficients are presented in the lower portion of Table 1. We use the sum of coefficients associated with level and change as a crude measure of quantitative importance. Based on this criterion, the short-run index and the long-run index have different driving forces. For the short-run index, the most important factor is "luxury goods" and "insurance". All other types of searches follow far behind. For the longer-run index, the most important factors are "energy cost" and "business conditions". All others follow far behind.

5. The search-based consumer sentiment indexes for US at national level

In section 4, we discuss how to derive the relative weights, which are needed to reduce the dimension of Google information, using an international sample and a US sample at the state level. The objective of the paper, however, is to construct consumer sentiment indexes for the US at national level. The only purpose of the international and the state-level samples is to estimate the weights. We now use the estimated weights to construct the national indexes for the US. We call the weights from the international sample (excluding US and excluding information after December 2007) as the "exogenous" weights. We call the weights from the US state-level sample covering all available data as the "endogenous" weights.

5.1. When weights are exogenous

These exogenous weights (column 1 and 2 of Table 1) come from the pre-2008 international sample that does not contain any US information. We discard the US sample so that the derivation of the weights is as exogenous as possible from the US perspective. We further exclude information after December 2007 in order to avoid the high correlation across countries in the economic crisis of 2008 and 2009. The US entered the recession in the last quarter of 2007, but majority of the OECD countries did not until in 2008. So the sample can be properly called the pre-recession sample.

The Google index has two-dimension; the first dimension is constructed from the short-run weights; the second dimension from the long-run weights. Here is the exact formula:

$$S_t = \sum_{i=1}^{6} \hat{w}_i G_{i,t} + \sum_{i=1}^{6} \hat{w}_{i,lag} \left(\frac{G_{i,t-1} + G_{i,t-2} + G_{i,t-3}}{3}\right)$$
$$L_t = \sum_{i=1}^{6} \hat{w}_i G_{i,t} + \sum_{i=1}^{6} \hat{w}_{i,lag} \left(\frac{G_{i,t-1} + G_{i,t-2} + G_{i,t-3}}{3}\right)$$

In the formula, S_t is the time-t value of the index constructed from the short-run weights. The term L_t is from the long-run weights. The first two columns of Table 1 have the exact value of the weights. The variables $G_{i,t}$ are the popularity trends of the i^{th} Google searches in the US at time t.

We will compare the search-based consumer sentiment indexes to University of Michigan's ICS and Conference Board's CCI to see if they following similar trajectory; we will test if the search-based indexes predict changes in the US consumer spending; we will also compare their predictive power to that of ICS and CCI.

The simple bivariate correlation coefficients between the search-based indexes and ICS and CCI are modestly high. Across indexes and weights, the minimum is 0.44, the maximum is 0.81. The average is around 0.6.

Table-2: Correlation between search-based indexes and other sentiment indexes

Google Indexes	CCI	ICS

	Short-run weights	Long-run weights		
CCI	0.44	0.64	1.00	
ICS	0.62	0.81	0.91	1.00

Figure 1 presents the plot the sentiment indexes. They follow similar trajectories through out the entire sample.

Table 3 shows the correlation coefficients between different sentiment indexes and consumption growth. We focus on the *level* of consumer sentiment and the *growth* in consumption, because they are known to exhibit high contemporaneous correlations (for example see Carroll, Fuhrer and Wilcox (1994)).[8] We use two measures of consumption in the US at the national level: the growth in personal consumption expenditure and the growth in retail sales. For both measures, the search-based sentiment indexes, in particular the one constructed using long-run weights, have higher correlation coefficient than either ICS or CCI. The difference is small, though.

Table-3: Correlation between various sentiment indexes and growths in consumer spending

	Growth in retail sales, US		Growth in personal consumption, US	
	Current month	Next-3 months	Current month	Next-3 months
Google Index				
Short-run weights	0.38	0.43	0.33	0.57
Long run weights	0.44	0.55	0.41	0.67
CCI	0.33	0.48	0.35	0.57
ICS	0.38	0.50	0.40	0.59

Figure 2 plots the Google indexes and the growth in consumption and retail sales. The plot shows a close co-movement between the Google indexes and changes in consumer spending. The trajectory of the search-based indexes is dominated by two large declines, one in mid 2005, the other after mid 2007. Both occasions also saw large decline in consumer spending. Recall that the weights are exogenous to US, so the coincidence is not a within-sample fit.

We now use a simple forecast model to see if the Google indexes forecast changes in consumption. The forecasting models are identical to that in Carroll, Fuhrer and Wilcox (1994), who ask "whether an index of consumer sentiment [the ICS in their case] has any predictive power on its own for future changes in consumption spending, and second, whether it contains information about future changes in consumer spending aside from the information contained in other available indicators." Their forecast models try to explain the growth of households spending using the lagged values of the consumer sentiment indexes. In our case, we have four measures of consumer spending: growth of personal consumption from last month, growth of retail sales from last month, the growth of consumption in the next three month, and the growth of retail sales in the next three month. Notice that our Google data are constructed using information one or two weeks before the end of the current month in an effort to preempt the release of the Conference Board's CCI. So it is still appropriate to use the term "forecast" despite that two of the dependent variables are growths from last month to the current month.

We intend to present a relatively complete picture to help readers judge on the robustness of our findings. Table 4 presents results from twenty-four regressions. To save space only the adjusted R-square and hypothesis testing are shown. Hypothesis tests use heteroscedasticity-and-serial-correlation-robust covariance matrices.

We have twenty-four regressions because we have four dependent variables, and we run six regressions for each DV; the first three use consumer sentiments as the only predictors; the other three also include control variables (six lagged changes in personal disposable income and in personal consumption expenditure). Within the three that have no control variables, the first regression uses the two-dimensional Google indexes and their three lags as predictors. The exact regression is

$$DV_t = const + \beta_{s,0} S_t + \beta_{l,0} L_t + \sum_{i=1}^{3} \beta_{s,i} S_{t-i} + \sum_{i=1}^{3} \beta_{l,i} L_{t-i} + u_t$$

The *DV* in the regression is one of the four dependent variables (growth in personal consumption expenditure, growth in retail sales, growth in personal consumption expenditure in the next three month, and growth of retail sales in the next three months). The term S_t is the Google index constructed using the short-run weights (coefficients in column 1 of Table 1 in the current case). The term L_t is the Google index constructed using the long-run weight (column 2 of Table 1 in the current case).

The second regression uses CCI, ICS and their six lags as the only predictors. We use six lags because, more often than not, the regressions of this type achieve the highest adjusted R2 with six lags. The exact regression is

$$DV_t = const + \beta_{ics,0} ICS_t + \beta_{cci,0} CCI_t + \sum_{i=1}^{6} \beta_{ics,i} ICS_{t-i} + \sum_{i=1}^{6} \beta_{cci,i} CCI_{t-i} + u_t$$

The third regression uses all sentiment indexes: the Google indexes, the CCI, and the ICS. The exact regression is

$$DV_t = const + \beta_{s,0} S_t + \beta_{l,0} L_t + \sum_{i=1}^{3} \beta_{s,i} S_{t-i} + \sum_{i=1}^{3} \beta_{l,i} L_{t-i}$$
$$+ \beta_{ics,0} ICS_t + \beta_{cci,0} CCI_t + \sum_{i=1}^{6} \beta_{ics,i} ICS_{t-i} + \sum_{i=1}^{6} \beta_{cci,i} CCI_{t-i} + u_t$$

Regression four to six are simply the counterparts of regression one to three, but with the control variables added. The control variables include six lags of growth in personal disposable income and that in personal consumption. With this twelve extra variables, the degree of freedom is low, dangerously so for the regressions with all three types of sentiment indexes and their lags.

First let's look at models that use Google indexes as the only predictors. Without the control variables, the adjusted R2 is 26% for explaining the current month growth in personal consumption, 34% for explaining current growth of retail sales, 47% for explaining next-three-month growth of personal consumption, and 37% for that of retail sales. The hypothesis that the Google indexes have no forecasting value is rejected at below 1% for all the four regressions.

The finding is similar when we add to the model the set of control variables (six lags of growth in disposable income and that of personal consumption). The null hypothesis is still rejected at below 1% for all regressions. It is interesting to note that, for forecasting three-month growths, the addition of control variables does not improve the adjusted R2 much. The observation suggests that there is no need for control variables once we have the Google indexes in the model.

After establishing the informational value of the Google indexes, we now compare them to CCI and ICS. The first step is to compare the adjusted R2 of models that uses only Google indexes and those of models that uses only CCI and ICS. The comparison is also in Table 4. Without exception and regardless whether the control variables are present, the models that use Google indexes always have higher adjusted R2 that models that use CCI and ICS. For retail sales, the advantage of Google indexes over CCI and ICS is 0.18 on average in term of difference in adjusted R2. For personal consumption, the average difference is about 0.09 on average. The fact that Google indexes are more powerful in predicting retails sales probably comes from the fact that the weights we use to construct the index are coefficients that tries to explain retail sales. It is encouraging to see that the indexes do well in forecasting personal consumption, too.

Table 4 also presents results from models that have all the sentiment indexes: Google, CCI, and ICS all together. This type of models largely has similar adjusted R2 as those that using the Google indexes only. This suggests that CCI and ICS are largely unnecessary once we know the Google indexes. We conduct a set of hypothesis tests to see if Google indexes bring extra information on top of CCI and ICS, and if CCI and ICS bring extra information on top of the Google indexes. Out of the eights tests (four dependent variables, with and without control variables), six of them reject the null at 10% that the Google indexes have no information in the presence of CCI and ICS; only four rejection are found in the opposite direction. So if anything, the Google indexes win the competition.

A quick summary of Table 4 is as follows: Google indexes have information in forecasting consumer spending; they have more information than the combined force and CCI and ICS; they bring extra information on top of the combined force of CCI and ICS.

5.2. When the weights are derived from US information

This is our preferred set of weights (column 3 and 4 of Table 1). We prefer the weights because they are from the US sample, and we use all available information to derive the weights. To recall, these weights are the estimated relation between Google searches and changes in retail-trade employment from a panel of 49 states between January 2004 and May 2009. We use the changes in retail-trade employment as a proxy for changes in spending for the lack of data. The relation between consumption and retail-trade employment is presumed. If the relation between the two is weaker than we hope, the quality of the weights may be adversely affected. This is a risk we are taking.

The indexes constructed from endogenous weights have very high correlation with ICS and CCI. The correlation coefficients range from 0.85 to 0.95. These coefficients are shown in Table 5.

Table 5: Correlation between the Google indexes and other Indexes of consumer sentiment

	Google Indexes		CCI	ICS
	Short-run weights	Long-run weights		
CCI	0.89	0.89	1.00	
ICS	0.85	0.95	0.91	1.00

We would like to note that the correlation coefficients are substantially higher than those from indexes that are constructed using the weights exogenous to the US. It is also worthwhile to note that the construction of the index do not use any information from either the ICS or the CCI. The key information is changes in retail-trade employment at the state level. The high correlation, average about 0.9, is not by construction. What it suggests is that a) changes in retail-trade employment are related to changes in consumer sentiment; b) Google searches contain information that is similar as those measured by the consumer sentiment survey by the University of Michigan and that by the Conference Board.

Figure 3 shows the trajectory of these sentiment indexes. They all follow similar trajectories through out the entire sample, which is to be expected given the correlation coefficients in Table 5.

Table 6 shows the correlation between the US-weighted Google indexes and measures of consumer spending. Compared with either CCI or ICS, the Google indexes generally have higher correlation coefficients with changes in the spending.

Table 6. Correlation between various sentiment indexes and growths in consumer spending

	Growth in retail sales, US		Growth in personal consumption, US	
	Current month	Next-3 months	Current month	Next-3 months
Google Index				
Short-run weights	0.44	0.54	0.44	0.54
Long run weights	0.41	0.63	0.42	0.70
CCI	0.33	0.48	0.35	0.57
ICS	0.38	0.50	0.40	0.59

Figure 4 plot together the Google sentiment indexes and the changes in consumer spending. Overall they have high co-movements. The early part of 2009 sees consumer spending picking up, but the sentiment indexes remain low. The low sentiment index cast a shadow about recent improvement in consumer spending and is dragging down the stock market as we are drafting this paper.

We now examine whether the Google indexes constructed using the US weights predict consumer spending in the US, and how their forecasting power compared to the ICS and CCI. The lower panel of Table 4 repeats the regressions that are presented in the upper panel of the same table. The only difference is that we now use Google indexes built on the US weights, instead of the international weights that were used in the upper panel.

We look at the three observations we made earlier in section 5.1:

1. Does Google indexes have information to forecast US consumer spending? The answer is yes and it is robust. Statistical tests reject the null that Google searches have no information at below 1% for all dependent variables, with and without control variables in the models. The adjusted R2 ranges from 21% to 49% depending on the identity of the dependent variable and whether control variables are present.
2. Do Google indexes bring more information than the combined force and CCI and ICS? The answer is yes: out of eight comparisons (four DVs, with and without control variables), the Google indexes lose to CCI and ICS only in one case in term of adjusted R2.
3. Do Google indexes bring extra information on top of CCI and ICS? The answer is yes when the model does not include control variables. The answer is largely no (three out of four tests) if the control variables are presents. We would like to note that, two out of the four tests also fail to reject the null that CCI and ICS have no extra information on top of the Google index. Our reading of the results is that, with control variables, Google indexes can substitute CCI and ICS, and vice versa.

We conclude this section by stating that the Google index, constructed using the weights constructed from the US information, do better than CCI and ICS combined under majority of the statistical criterions. But a more conservative reading suggests they are in par with the combined force of CCI and ICS.

6. Conclusion

As Internet becoming simply part of life, people use Internet searches to look for information about issues and things that interest them or concern them. It follows that by keeping track of aggregate search patterns, one can keep an eye on the public's interest and concerns. This paper is one example of using search data for such purposes. We construct search-based indexes of consumer sentiments from the popularity of six types of Google searches: those that are related to debt burdens, those that reflect business conditions, those that are related to energy costs, searches for luxury goods, searches for credit products, and searches for insurance. We do not view our selection as the last word. There is room for refinements and expansions.

We construct the indexes of consumer sentiment for the US as the weighted sum of the popularity trends of the six types of searches. We subject the construction of the indexes to alternative weights. One set of weights is exogenous to the US. It is the estimated relations between Google searches and retail sales in an international panel without US. We also exclude information after December 2007, when the worldwide economy crisis brought significant co-movement across countries. The second set of weights is endogenous to US. It is from a panel of 49 states in the US. The weight measures the link between Google searches and changes in retail-trade employment.

With these weights, we construct the indexes and find they have similar time trajectory as the University of Michigan's Index of Consumer Sentiment (ICS) and the Conference Board's

Consumer Confidence Index (CCI). The correlation coefficient on average is 0.6 when the weight is exogenous, and is close to 0.9 when the weight is endogenous to US.

We find the resulted index forecast the growth in personal consumption and retail sales in the US, either in the case of exogenous weights or the case of endogenous weights. The finding is also robust to different measures of consumer spending. We use four of them: monthly growth in personal consumption expenditure, monthly growth in retail sales, the sum of next-three month growths in the consumption expenditure, and the sum of next-three month growths in the sales. The finding is also robust to alternative models, whether or not the models include control variables or not. We use the six lags of personal disposable income and that of personal consumption expenditure as the control variables.

Finally, in a head-to-head comparison in term of forecasting power, we find that the Google indexes are at least as informative as the ICS and CCI. The observation is robust across a spectrum of dependent variables and models specifications. More often than not, the Google indexes outperform the combined force of CCI and ICS.

We thus conclude that the pattern of Internet searches made available by Google Insight for Search can be used to monitor changes in consumer sentiment. The Google data are available at weekly basis, as opposed to the monthly release of ICS and CCI. Google data thus allow researches to keep a more timely watch on consumer sentiment, and for investors to preempt the monthly release of the ICS and CCI, which can have impact on stock and bonds market.

Reference

Askitas, Nikos & Zimmermann, Klaus F., (2009). "Google Econometrics and Unemployment Forecasting," IZA Discussion Papers 4201, Institute for the Study of Labor (IZA).

Bram, Jason & Ludvigson, Sydney (1998). "Does consumer confidence forecast household expenditure? a sentiment index horse race," Economic Policy Review, Federal Reserve Bank of New York, issue Jun, pages 59-78.

Carroll, Christopher D & Fuhrer, Jeffrey C & Wilcox, David W, (1994).
"Does Consumer Sentiment Forecast Household Spending? If So, Why?," American Economic Review, American Economic Association, vol. 84(5), pages 1397-1408, December.

Da, Zhi & Engelberg, Joseph & Gao, Pengjie, (June 4, 2009). "In Search of Attention", Available at SSRN: http://ssrn.com/abstract=1364209

Howrey, E. Philip (2001). "The Predictive Power of the Index of Consumer Sentiment," Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 32(2001-1), pages 175-216.

Matsusaka, John G & Sbordone, Argia M., (1995). "Consumer Confidence and Economic Fluctuations," Economic Inquiry, Oxford University Press, vol. 33(2), pages 296-318, April.

[1] Associated Press, http://www.msnbc.msn.com/id/32438282/ns/business-stocks_and_economy/; updated 11:59 a.m. MT, Sun., Aug 16, 2009

[2] For more information, please refer to Google Help > Insights for Search Help > Working with insight for Search > Analyzing Data. URL: http://www.google.com/support/insights/bin/topic.py?topic=13975

[3] Google does not clarify their exact formula, but we believe it is along the line x_t = [ln(searches for "x" at time t/total searches at time t) - ln(searches for "x" at time 0/total searches at time 0)] *100%.

[4] The winter of 2007-08 saw more than a few reports of elderly found dead in unheated or under-heated residences. The media generally linked these incidents to high cost of heating. Regarding gas prices, in 2008 two major U.S. presidential hopefuls, John McCain and Hillary Clinton, proposed a gas tax holiday to ease the burden on households in the face of rising gas prices.

[5] The University of Michigan survey asks how the respondents think about the "business conditions in the country as a whole". The Conference Board survey asks the respondents to "rate present general business conditions" as well as the conditions six months from now.

[6] A final note: we seasonally adjust the times series from the Google Insight by regressing them on monthly dummies, within each country, before subtracting the estimated coefficients of the dummies. We also adjust the growth rate in retail sales from OECD.StatExtract with the same method before the regressions.

[7] Again, the Google information is seasonally adjusted within each state using monthly dummies. The BLS has already removed seasonal pattern from the retail-trade employment; we use their data without any further processing.

[8] In fact the growth of sentiment index and the growth of consumption have much smaller correlation coefficients in our data.