Asset Prices and Google's Search Data

Bodo Herzog[*]

ESB Business School, Reutlingen University

Reutlingen Research Institute (RRI), Reutlingen University

Abstract

This paper investigates the novel relationship of asset price determination via Google data and trading volume. To capture this relation, we construct a model and secondly estimate panel and time series regressions. We use weekly data from 2004 to 2010 for 30 international banks. Our study is the first which differentiate between Google's search volume and Google's search clicks. We find that asset prices are positively related to the growth rate of Google's search, trading volume and the level of Google search clicks. Moreover, this finding is in line with our theoretical model. Secondly, we find that the absolute level of Google's search volume and Google's search clicks behave differently regarding asset price dynamics. Google's search volume, which measures long-run searches, is negatively related to asset prices and Google's search click is positively related. Thus, Google's search data offer a new insight into two different aspects: On the one hand it is a natural way of measuring attention to assets, and on the other hand it provides a timely and unbiased measurement of asset price dynamics especially during turbulent times. We conclude that Google's data contain important information for the identification of asset bubbles.

Key words: Google Search, Asset Price Determination, Trading Volume, Asset Bubbles

JEL classification: G12, E65, C58

[*]Corresponding author: Dr. Bodo Herzog, Professor of Economics. Address: Alteburgstr. 150, 72762 Reutlingen, Germany. Phone: +49 7121 271 6031, FAX: +49 7121 271 6022. Email address: Bodo.Herzog@Reutlingen-University.de.

Preprint submitted to 25th Australasian Finance & Banking Conference 2012

1. Introduction

There is a vast theoretical literature about the determination of asset prices and the critical importance of attention (Grossmann and Stiglizt 1980; Allen 1990). Basically, in theory we assume that economic agents can gather the relevant information and then use it to make proper investment decisions. However, recent empirical and theoretical work by Cohen and Lou (2010) and Duffie (2010), and the knowledge of the theory of bounded rationality, emphasize that information processing is even more complex and takes time (Kahneman 1973). Consequently, the goal to investigate how attention affects asset prices is a difficult endeavor. In addition, until today there is almost no direct data for the measurement of investors' attention.

In a recent paper, Da, Engelberg and Gao (2010) suggest a direct measure for attention – apart from the existing proxies such as trading volume (Barber and Odeau 2008; Hou, Peng, and Xiong 2008), news and headlines or extreme stock returns (Li, Mahani and Sandhya, 2011) – via Google search. Following this idea, we evaluate the quality of Google's search data and trading volume as a measurement for asset price determination. In contrast to the paper by Da, Engelberg and Gao (2010), we focus on banks and in particular on the cross-country effects. Moreover, our study is the first which differentiates between Google's search volume and Google search clicks. In light of these facts, this paper contributes in several ways to the pre-embryonic empirical literature about the usage of Google data.

Today, more and more financial trading processes at stock exchanges are done electronically and via high frequency trading. The impact of these processes is illustrated by the recent flash crash of May 6th 2010 when electronic trades caused the Dow Jones Industrial Average to collapse by seven hundred points in just minutes. To take a closer look at these developments, we analyze the attention and determination in stock prices via the internet using the information of Google data. In doing so, we limit the attention to three leading banking institutions in 11 different OECD countries. This way we make sure to include enough cross-country data and balance the size of the financial sector across countries. The limitation to three banks per country may produce a selection bias, however, the selected banks always cover a considerable market share in each country, and they are global players, which is necessary for sufficient Google search per day. Consequently, we are able to analyze how attention, measured as the search behavior, may influence the price of financial assets.

In this paper, we suggest that the use of aggregate Google search data is an important step towards the identification of information based stock price fundamentals. One of the first economists, Choi and Varian (2009), proposed that search data has the potential to forecast a variety of important economic variables. This in particular should be strengthened in a so-called mobile age, where almost everyone has online access and thus the ability to search online. Hence, search data is a valuable and unique chance to identify new determinants of market dynamics. We use Google search data, because Google is currently dominating the market for search engines and because it offers data that is publicly available. However, there are two sources that slightly differ from one another: At first, Google Trends (www.google.com/trends) or later on referred to as Google search volume. At second Google Insights (http://www.google.com/insights/search/#) or later called Google search clicks. Data from Google Trends contain the search volume for different search expressions and measures the number of searches by its time-series average. Google calls this index a relative number index, and it measures how many times a certain term is searched for, compared it to its long-run average. It is important to know that the numbers are not absolute search traffic but scaled to the average search, which is one. To give an example, let's take a look at Google search data that is rated 1.9. This can be interpreted as a search traffic that is 1.9 times higher than the long-run average over the period from 2004 to 2010. However, Google Insights data offer a measurement of instant Google search clicks. Therefore, we use both measurements and examine the difference of Google Trends, i.e. a long-run measure, and Google Insights, i.e. an instant measure, of attention. We study panel and time series of weekly Google search data. The data sample ranges of January 2004 to 2010. The sample period is constraint because Google data is only available from January 2004 and so there is no possibility for longer time series investigations.

The contributions in our paper are: first we construct an asset pricing model to include the idea of an attention variable measured via Google. The model is based on a standard Lucas asset-pricing approach (Lucas 1978). Even the recent paper in that field has not established such a systematic approach. Secondly, we empirically estimate the effects and compare the two different Google indicators. Thirdly, we evaluate the role of Google data in order to measure attention and its role in asset price determination. Finally, we study the Google data in respect of the identification of asset bubbles.

In general, we find the following results: firstly, in our model, we identify the stock price determinants, such as the present value of asset returns, trading volume as well as news indicators due to stochastic uncertainty. Secondly, in the econometric part, we find that the change of Google search volume, the trading volume, the stock index, and the level of Google search is positively related to asset prices, although a certain part of the variation remains unexplained. Thirdly, we demonstrate that the theoretical model is in line with the empirical findings. Interestingly enough Google Trends, which measures the long-run search performance, is negatively related to the asset price. This indicates that a higher search trend, in comparison to the long-run average, only has a positive impact to asset prices if simultaneously the instant search is above a certain threshold. In contrast to Da, Engelberg and Gao (2010), the two different sources of Google's search data, enables us to analyze also turbulent asset pricing periods such as bubbles.

Therefore, we use the unique Google data to test several hypothesis of asset price determination. Empirically we also test the theoretical hypothesis of the so-called price momentum effect. A paper by Daniel, Hirshleifer, and Subrahmanyam (1998) shows that investors often overreact to private signals, due to behavioral reasons. A high attention is a necessary condition for overreaction, due to correlation with private information. Consequently, more attention – here measured by Google search – should lead to an overreaction and thus stronger price movements. However, contrary to this literature are work by Hong and Stein (1999), as well as Bodie, Kane and Marcus (2008). They argue that movements in stock prices cannot be foreseen and therefore they are unpredictable. This unpredictable behavior is noted in literature as a random walk (Cooper 1982) and is linked to the famous efficiency market hypothesis (Fama 1970, 1998). This phenomenon states that assets, traded on the market, immediately include the known information in their prices and that only unexpected events will have an influence on the asset price. We examine both effects with the help of Google data and find new empirical evidence for a price momentum effect.

In addition, we check whether Google has an added value for investors. A lot of research has been done in the field of data and text mining techniques. Fung, Yu and Wai (2002) develop a system which enables investors to use the content of news articles in order to forecast the behavior of the stock market. Niederhoffer (1971), Tetlock, Saar-Tsechansky and Macskassy (2007), Calvet and Fisher (2007), Berkman et al. (2009), and recently Li, Mahani and

Sandhya (2011) find that news release as a measure of attention is also related to stock returns. Nonetheless, our study stands out on this literature due to the inclusion of Google data as a timely, objective and novel measurement of attention.

A second issue we are interested in is the role of trading volume. A paper by Brailsford (1994) identifies an asymmetric relationship between price changes and trading volume. Hence, we want to see whether this correlation is valid for our sample and time period. In summary, Choi and Varian (2009) argue that Google contain valuable information for economic or financial forecasting. We follow this hypothesis, however, we use Google search as a measure of attention and therefore apply it to the asset price determination.

The remainder of the paper is organized as follows: Section 2 describes the data. The theoretical and econometric model is illustrated in Section 3. In Section 4, we discuss the empirical results, and Section 5 concludes the paper.

2. Data

The Google data is publicly available and obtained from the Google.com homepage by creating a Google account. From 2004 onwards, the search index is only available on a weekly basis. In other words, we are only granted access to daily data of the last 90 days. The focus of our paper lies upon banking stocks because of the financial crisis. We have decided that the attention of banking stocks is of particular interest and allows the identification of cross-country differences. Consequently, we collect a new data set across the G7 member states and include banks from every country. On top of G7 countries, we include the same number of financial institutions from Switzerland, Spain and China. Therefore, in total, we have 11 countries. The reason we include Spain in our sample is that, thanks to their strictly regulated financial market, banks there were not affected as heavily by the financial crisis of 2007 to 2009. Due to the fact that the Chinese banks were not listed on a stock exchange before September 30, 2007, our data is limited in this case too. You can find a complete list of the relevant banks (Table 1A), as well as a summary of the descriptive statistics (Table 2A) in the appendix.

By gathering time-series data from the stock exchange, we list each bank with the name of its home country. The daily information about the Bank of America, for example, was extracted from the New York Stock Exchange. In order to be consistent with Google Trends, it was necessary to include the daily data on a weekly basis, starting from January 1, 2004 until March 15, 2010 (Table 1A – Appendix). Thereafter, we have double-checked the received data in order to ensure consistency. Besides that, we include the overall performance of the stock market in terms of an index in each of the countries, respectively. Therefore, we utilize the same sources to obtain these daily numbers. Finally, it is important to note that we always use the daily closing price. All in all, by combining the time-series dimension with a maximal value of t = 324 (number of weeks) per bank and the cross-sectional dimension with n = 30, we also create a panel data file (Table 2A – Appendix).

Similarly to Daniel et al. (2010), we left out some company fundamentals and the state of the overall economy in our regression and accept a possible omitted variable bias. Due to multi-collinearity caused by the relatively high correlation between stock price and the index regressors, we make sure to calculate standard errors adjusted for autocorrelation and heterogeneity and apply advanced econometric techniques, such as ARCH and VAR models.

3. The Model

The theoretical foundation of our empirical approach is based on an extension of a Lucas asset-pricing model (Lucas 1978). Yet, in the standard Lucas model we incorporate two new elements. First we explicitly model the dynamics of the dividend process via a stochastic differential equation. A second innovation is the modeling of attention on the dividend process through the consideration of Google search on either positive or negative news. The standard Lucas model is extremely useful when studying empirical issues, such as stock price determination (Campbell 1986, Mehra Prescott 1985, Mankiw 1986). Our extended model contributes to the literature twofold: firstly, it is a step toward a better theoretical understanding of Google search data and asset price determination. Secondly, in contrast to the existing Google search literature, it is the first consistent theoretical foundation of Google data within a theoretical and empirical paper.

Suppose there are n risky assets in the economy. They generate a stochastic return equal to δ_{it} per period. The assets are the only source of income and p_{it} is the price of asset i in period t. Hence, p_t and δ_t are $n \times 1$ vectors of prices and dividends at time t. The economy consists of identical and infinitely living individuals which maximize the expected utility such as

$$\mathbb{E}\left[\sum_{t=0}^{\infty}(1+\theta)^{-t}U(c_t)|0\right] \tag{1}$$

In any period the individual receives dividends on the quantity x_{it} of each asset that she or he holds between period t and $t+1$. Hence, x_t be a $n \times 1$ vector. The budget constraint is

$$c_t + p_t' x_t = (p_t + \delta_t)' x_{t-1} \,. \tag{2}$$

The right-hand side illustrates the income of the portfolio selected at time *t-1* including dividends. The left-hand side is equal to consumption plus the new value or investment of the portfolio chosen at time *t*. Hence, the first-order conditions of that optimization problem are

$$p_{it}U'(c_t) = (1+\theta)^{-1}\mathbb{E}[U'(c_{t+1})(p_{it+1} + \delta_{it+1})|t], \quad for \ i = 1, \dots, n. \tag{3}$$

In market equilibrium, the quantities of each asset demand must be equal to the exogenous supply. Without loss of generality, suppose there is only one unit of each asset, i.e. $x_{it} = 1$ for all *i* and *t*. Consequently, consumption must be equal to output, which is the sum of dividends according to the budget constraint (2): $c_t = \sum \delta_{it}$. In addition, equation (3) is a recursive relation which determines the price of assets as a function of exogenous variables. The forward solution of equation (3) is straightforward[1] and result in

$$p_{it} = \mathbb{E}\left[\sum_{j=0}^{\infty}(1+\theta)^{-j}\left(\frac{U'(c_{t+j})}{U'(c_t)}\right)\delta_{it+j}|t\right]. \tag{4}$$

According to the solution, the asset price is equal to the expected present discounted value[2] of dividends. To derive testable results it is necessary to make further assumptions about the utility function and distribution of dividends. First, assume risk neutral individuals so that the first derivative of the utility function *U'(c)* stay constant. Second, suppose the dividend process is following a stochastic differential equation, such as

$$\frac{d(\delta_t)}{\delta_t} = (f_t + a_t)dt + \sigma dW_t \tag{5}$$

[1] We use the standard assumption of no bubbles or no-Ponzi-schemes.

[2] The discount rate is the marginal rate of substitution between consumption at time *t+j* and consumption at time *t*.

where *f* represents the stock fundamentals, *a* is the measurement of attention, σ is the volatility and *dW* is a Wiener process. Furthermore, the attention variable a_t is modeled as follows

$$a_t = \begin{cases} SV_t & for\ \ good\ news \\ -SV_t & for\ bad\ news \end{cases} \qquad with\ \ SV_t > 0 \quad \forall t \tag{6}$$

for which SV_t is the Google search, under the assumption that both good and bad news generate higher search volume. We assume that good news increases the attention to buy an asset and bad news implies large negative attention to sell. Respectively, good news increases and bad news decreases the future stream of dividends. Furthermore, we implicitly assume that financial statements and news announcements result in a higher Google search. Using these assumptions we are able to rewrite equation (4) as

$$p_{it} = \mathbb{E}\left[\sum_{j=0}^{\infty}(1+\theta)^{-j}\left(\int\left(f_{t+j}+a_{j+t}\right)\delta_{it+j}d(t+j)+\int\sigma\delta_{it+j}dW_{t+j}\right)\ \ |t\right]. \tag{7}$$

Equation (7) enables us to analyze how the asset price behaves over time. In short, the price is equal to the present discounted value of expected dividends, discounted at a constant rate, which is the subjective discount rate of individuals. Asset prices are determined by movements in expected dividends and consist of two parts: (a) the drift term $\boldsymbol{(f_{t+j}+a_{j+t})\delta_{it+j}}$ based on the asset fundamentals and attention, and (b) the diffusion term $\boldsymbol{\sigma\delta_{it+j}}$ based on volatility. Higher dividends, as a result of more Google search due to good news, add attention and should increase the stock price. But higher dividends also mean higher consumption, and thus lower marginal utility – other factors being unchanged. Therefore, dividends are valued less when attention and consumption is high. This theoretical result is also testable in the empirical section.

4. Estimation Results and Discussion

We first examine the relationship between the change in Google search volume (GSV) and other proxies, including Google Trends, stock index, asset price and stock returns. The results, which are reported in Table 1, contain different models with cross-section fixed-effects and weekly fixed-effects. The standard errors are cluster by banks. The tests of fixed-effects are reported in Table 3A of the appendix.

Comparing the four regressions, we discover that the econometric models are robust and almost all coefficients are statistically significant. We confirm that the change in instant Google search clicks is positively related to both Google Trends, which measures long-run search performance, and the asset price. This suggests that high asset prices result in more attention and search for banking stocks respectively.

Table 1: The Change in GSV an Alternative Measurement of Attention

Variables	(1)	(2)	(3)	(4)
Intercept	0.768***	0.052	0.198	0.197
	(0.126)	(0.033)	(0.266)	(0.265)
log(Google Trends)	0.136***	0.115***	0.162***	0.162***
	(0.010)	(0.009)	(0.011)	(0.011)
log(Index)	-0.116***	-0.010***	-0.041	-0.040
	(0.016)	(0.003)	(0.030)	(0.030)
log(Price)	0.052***	0.003***	0.031***	0.030***
	(0.007)	(0.001)	(0.008)	(0.008)
log(Return)	-0.012***	-0.007***	-0.008***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Change of Trading Volume				0.0001***
				(0.0001)
Corss-section Fixed Effects	YES	NO	YES	YES
Period Fixed Effects	NO	YES	YES	YES
Observations	3857	3857	3857	3857
Adjusted R^2	0.052	0.224	0.238	0.166
S.E. of regression	0.145	0.137	0.136	0.136
F-statistic	6.324***	3.269***	3.209***	3.239***
Durbin-Watson stat	2.142	2.183	2.150	2.152

Dependent variable: Change of Google Search Volume. The standard errors clustered by banks are in parentheses. *, ** and *** represents significance at the 10%, 5% and 1% level.

Moreover, by using the timeframe available of Google data, this relationship helps us recognize an asset bubble or price developments which are not based on changes in fundamentals in our theoretical model measured by the function '*f*'. Both the stock index and the returns have negative and significant coefficients related to the dependent variable. Hence, in case of high returns or index levels, the growth rate of Google's search volume is small. This explains the tipping point property close to a peak or trough in a market. Consequently, the attention in a banking stock is high during an upturn or downturn due to relative high opportunities for profits or losses. But the opportunity of high profits is over in a peak. Moreover, the change in trading volume and Google search is statistically significant positively related. As a consequence, the change in Google's search and trading volume are potential substitutes.

Next, we estimate the change in Google's search, together with the first- and second-order lag variables. This procedure helps to identify possible forecasting effects of Google data. The result reported in Table 2 show that Google search volume (GSV) and trading volume have the same positive and statistically significant signs. As expected of a leading variable, the lag variables are both negatively related. The higher the search traffic the lower the change of Google search, because the attraction in terms of profits is continuously diminishing.

Table 2: The Change in GSV, Panel Estimates

Variables	(1)	(2)	(3)
Intercept	-0.001	-0.001	-0.001
	(0.001)	(0.003)	(0.003)
log(Google Trends)	1.312***	1.133***	1.137***
	(0.009)	(0.005)	(0.005)
log(Google Trends, -1)	-1.311***	-1.133***	-1.127***
	(0.009)	(0.005)	(0.006)
log(Google Trends, -2)		-	-0.011***
			(0.005)
log(Trading Volume)		0.001*	0.002**
		(0.001)	(0.001)
log(Trading Volume; -1)		-0.001*	-0.002**
		(0.001)	(0.001)
Period Fixed Effects	YES	YES	YES
Observations	8921	8455	8426
Adjusted R^2	0.731	0.868	0.868
S.E. of regression	0.089	0.048	0.048
F-statistic	75.76***	172.3***	172.0***
Durbin-Watson stat	2.544	2.574	2.575

Dependent variable: Change of Google Search Volume. The standard errors clustered by banks are in parentheses. *, ** and *** represents significance at the 10%, 5% and 1% level.

The next empirical model explains the stock price movements as a natural logarithm in relation to the other variables, in particular the trading volume, as suggested in the theoretical model. The estimation results are robust and again almost all coefficients are statistically significant (Table 3). Nevertheless, we cannot ignore the observation that asset prices can be explained by Google search variables. Interestingly enough, there is a significant positive relation between the asset price and instant Google search clicks, i.e. search clicks can be related to its rise in prices. Thus a sustainable asset price upturn is recognized with high Google search clicks but still low Google search volume in relationship to the long-run average. As soon as the Google Trends variable (long-run search) is above the average and Google search clicks are further growing, the asset price moves into a bubble or follows a typical unsustainable herd behavior. This effect is demonstrated by the statistically significant negative impact on stock prices via Google Trends. In other words, high attention measured by instant Google clicks has a positive impact on the price. But in case of an asset price boom, which is measured as Google search above its long-run average, the Google Trends variable predicts a kind of the turning point. Consequently, in this case, the asset price declines and displays therefore a negative coefficient. Therefore, we are able to identify the situation in which the asset price upturn is over and the turning point is close. Based on our model, we find that a high average search volume implies higher uncertainty in the market environment – usually close to the asset price peak – and, therefore, a negative price effect.

In line with our model, high search clicks affect the model parameter '*a*' and this is theoretically and empirically positive related to the stock price. Hence, Google search clicks – measured by Google Insights – is a better measurement of attention than the deviation of Google searches based on the long-run search average. Another explanation for the aforementioned finding is the higher transparency and information available in instantaneous Google clicks. Thus, Google trends measures a growing bubble and is a possible indication of information congestion (Shiller 1990). Next, we focus on the impact of stock prices and trading volume.

Table 3: Stock Price in log in Relation to Measurements of Attention, Panel

Variables	(1)	(2)	(3)
Intercept	5.771***	1.893	1.627
	(0.315)	(1.614)	(1.819)
log(Google Search Volume)	-0.176***	-0.015**	-0.015**
(by Google Trends)	(0.011)	(0.006)	(0.006)
log(Google Search Clicks)	0.215*	0.001**	0.001*
(by Google Insights)	(0.119)	(0.001)	(0.001)
log(Trading Volume)	-0.152***	-0.013***	-0.013***
	(0.005)	(0.001)	(0.001)
AR(1)		0.999***	0.942***
		(0.001)	(0.010)
AR(2)			0.057***
			(0.010)
Corss-section Fixed Effects	YES	-	-
Period Fixed Effects	YES	-	-
Observations	8361	8326	8293
Adjusted R^2	0.977	0.998	0.998
S.E. of regression	0.269	0.065	0.065
F-statistic	1.026***	156129***	1246976***
Durbin-Watson stat	0.119	2.119	2.014

Dependent variable: Log of Stock Price. The standard errors clustered by banks are in parentheses. *, ** and *** represents significance at the 10%, 5% and 1% level.

As expected, and in line with literature, stock prices and trading volume are in an asymmetric relationship with one another. This means that if trading volume, for example, increases by one unit, stock prices are affected negatively due to the estimated negative coefficient in Table 3. Even though the trading volume coefficients are small, we cannot deny this link. This finding is robust for all models. The cause for this negative relationship could be the fact that trading volume is a leading variable and measures price contagion. If today all investors purchase or sell a certain stock, it automatically results in a higher trading volume. Therefore, trading volume signals a busy or congested market with declining profit margins. Accordingly, a stock becomes less attractive in the heat of a boom or during a crisis. In other words, the opportunity to make money, measured as a change in prices, declines respectively. All findings support the hypothesis that Google search data are a relevant indicator for investors' decisions. In model (2) and model (3) we include an autoregressive term to capture the time series property in the data. The first- and second-order autoregressive terms are both statistically significant. For a more detailed analysis we also calculate a vector autoregressive model (VAR) and the corresponding impulse responding functions. The results of the VAR model support our findings. The impulse response functions illustrate the positive price reaction after a standard shock noted by Google search (Table 4A and Figure 4A - Appendix). Once again this demonstrates the importance of Google variables in determining the stock price albeit in the classical Lucas asset-pricing model the attention factor was not taken into consideration.

The biggest challenge of the rational asset pricing theory is the so-called price momentum effect (Jedgadeesh and Titman 1993). However, with the help of the Google's search variable, we measure the attention in relation to the stock price. Table 3 displays first empirical evidence in favor of the theoretically derived price momentum effect too.

Finally, we analyze how Google search variables and trading volume are related to asset prices (Table 4). The results are surprisingly robust and offer new insights to asset price determination. The negative relationship of asset prices and trading volume as well as Google search volume is as expected. One reason for this might be the fact that trading volume or Google search indicates busy markets. Consequently, both trading volume and Google search are potential leading indicators for busy market environments or even asset bubbles.

Table 4: Log-Change in Stock Price, Newey-West

Variables	(1)
Intercept	-0.001***
	(0.001)
Δ(log(Google Search Volume))	-0.014*
	(0.008)
Δ(log(Index))	1.247***
	(0.043)
Δ(log(Trading Volume))	-0.006***
	(0.001)
Observations	7519
Adjusted R^2	0.394
S.E. of regression	0.052
F-statistic	1632***
Durbin-Watson stat	2.183

Dependent variable: Δ(log(Stock Price). The t-values are calculated with the Newey-West formula. *, ** and *** represents significance at the 10%, 5% and 1% level.

In general, all findings show evidence that the empirical results are in line with the theoretical literature. Since Google's statistics have an impact on stock price and trading volume, bankers and investors may want to consider this variable for future investment decisions. In summary, Google variables are a relevant indicator for asset price determination, as well as nowcasting. Any news announcement released by the public relations department could raise investors' attention and, later, be measured by Google's search. As a result, depending on the magnitude and sign of the coefficient, as well as the positive or negative character of news the stock price will be affected. In order to be prepared for it or eventually avoid such trends, companies should monitor Google's search data. This could help companies to ensure a more sustained growth process of asset prices without major break-ins.

5. Conclusion

Throughout this paper, we have made several contributions to a new field of research about Google's search data. We applied the idea to asset price determination. We are the first to build an asset price model which includes Google variables. By using new Google data we collected ourselves, we are able to evaluate this relationship. To our knowledge, this paper is the first which compares both Google sources – Insights and Trends – and therefore offer a unique contribution to literature. Finally, we find the expected coefficient signs and evidence which strongly speaks for the empirical model. For instance, an increase in Google search and/or trading volume temporarily pushes up asset prices. This finding supports the theoretical idea of a price momentum effect which, until now, has more or less been an unproven theoretical proposition. Furthermore, we prove the asymmetric relationship between stock price and trading volume and, in doing so, reaffirm the recent respective theoretical literature.

To sum it up, according to our study, Google statistics are a useful source. It is obvious that more research needs to be done on the predictive power of Google data. Although this paper is one of the first scientific evaluations of Google data, there is room for further research. We suggest an extension of the model to more macroeconomic fundamentals, analyst forecasts or investors' opinion. Consequently, this paper has the potential to initiate more empirical research in macroeconomics using the unique information of Google variables. The usage of Google offers new insights with substantial benefits to the businesses worldwide.

References

Allen, F. (1990), The market for information and the origin of financial intermediation, *Journal of Financial Intermediation*, Vol. 1, Issue 1, p. 3-30.

Andersen, T.G., T. Bollerslev, F.X. Diebold, C. Vega (2004), Real-Time Price Discovery in Stock, Bond and Foreign Exchange Markets, *Working Paper*, 28.06.2004, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.58.1418&rep=rep1&type=pdf (15.03.2010, 17:02).

Bachetta, P. and E.v. Wincoop (2010), Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle, *American Economic Review,* Vol. 100, Issue 3, p. 870-904.

Ball, R., P. Brown (1968), An empirical evaluation of accounting income numbers, *Journal of Accounting Research*, Volume 6, No.2, Pages 159-178.

Barber, B.M. and T. Odeau (2008), All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *Review of Financial Studies*, 21(2), p. 785-818.

Berkman, H., V. Dimitrov, P.C. Jain, P.D. Koch, S. Tice (2009), Sell on the news : Differences of opinion, short-sales constraints, and returns around earnings announcements, *Journal of Financial Economics*, Vol. 92, Issue 3, p. 376 – 399.

Berry, T.D., K.M. Howe (1994), Public Information Arrival, *The Journal of Finance*, Volume 49, No. 4, p. 1331-1346.

Bessembinder, H., K. Chan, P.J. Seguin (1996), An empirical examination of information, differences of opinion, and trading activity, *Journal of Financial Economics*, Vol. 40, Issue 1, p. 105-134.

Bodie, Z., A. Kane, A.J. Marcus (2008), *Essentials of Investments*. – 7th Edition, New York: McGraw-Hill Companies, ISBN: 978-0-07-340517-9.

Boyd, J.H., R. Jagannathan, J. Hu (2001), The Stock Market's Reaction to Unemployment News: Why Bad News is Usually Good for Stocks, *Working Paper*, http://www.nber.org/papers/w8092.pdf (1.04.2010, 11:05).

Brailsford, T.J. (1994), The empirical relationship between trading volume, returns and volatility, Research Paper 94-01, *Working Paper*, http://www.trade2win.com/boards/attachments/day-trading-scalping/24067-see-now-voluvola.pdf (19.03.2010, 15:44).

Campbell, J. (1986), Bond and Stock Returns in a Simple Exchange Model, *Quarterly Journal of Economics,* Vol. 101, p. 785-804.

Calvet, L.E., A.J. Fisher (2007), Multifrequency news and stock returns, *Journal of Financial Economics*, Vol. 86, Issue 1, p. 178-212.

Chan, W.S. (2003), Stock price reaction to news and no-news: Drift and reversal after headlines, *Journal of Financial Economics*, Vol. 70, p. 223–260.

Choi, H. and H. Varian (2009), Predicting the Present with Google Trends, *Working Paper*, www.google.com/googleblogs/pdfs/google_predicting_the_present.pdf (14.03.2010, 09:37).

Cooper, J.C.P. (1982), World Stock Markets: Some random walk tests, *Applied Economics*, Vol. 14, Issue 5, p. 515-531.

Cohen, L. and D. Lou (2010), Complicated Firms, *Working Paper.*

Cornell, B. and A.C. Shapiro (1986), The reaction of bank stock prices to the international debt crisis, *Journal of Banking and Finance*, Vol. 10, p. 55-73.

Crislip, M. (2008), Google Trends and the Interest in Alternative Medicine, *Working Paper*, http://www.sciencebasedmedicine.org/?p=305 (14.03.2010, 10:04).

Da, Z., J. Engelberg, P. Gao (2010), In Search of Attention, *Journal of Finance*, (forthcoming).

Daniel, K.D., D. Hirshleifer, and A. Subrahmanyam (1998), Investor Psychology and Security Market Under- and Overreaction, *Journal of Finance,* Vol. 53, Issue 5, p. 1839-1885.

Darwin, A. (2009), Using Google Search Trends to Exploit Market Moves, *Working Paper*, http://www.darwinsfinance.com/using-google-search-trends-to-exploit-market-moves/ (14.03.2010, 10:27).

Deutsche Börse (2009, 2010), *Factbook*, http://deutsche-boerse.com/INTERNET/IP/ip_stats.nsf/(KIR+Factbook+Kassamarkt)/407B2EB81C10C693C12577270052CF51/$FILE/Factbook_2009_d.pdf?OpenElement, (19.09.2010, 15:20).

Duffie, D. (2010), Presidential Address: Asset Price Dynamics with Slow-Moving Capital, *Journal of Finance,* Vol. 65, Issue 4, p. 1237-1267.

Duffie, D. and G. Manso (2007), Information Percolation in Large Markets, *American Economic Review,* Vol. 97, Issue 2, p. 203-209.

Easley, D. and M. O'Hara (1987), Price, trade size, and information in securities markets, *Journal of Financial Economics,* Vol. 19, p. 69-90.

Evans, M.D.D. and R.K. Lyons (2008), How is macro news transmitted to exchange rates?, *Journal of Financial Economics*, Vol. 88, Issue 1, p. 26-50.

Fama, E. (1970), Efficient Capital Markets: A Review of Theroy and Empirical Work, *Journal of Finance*, Vol. 25, No 2, p. 383-417.

Fama, E. (1998), Market Efficienc, Long-Term Returns, and Behavioral Finance, *Jounral of Financial Economics*, Vol. 49, p. 283-306.

Fleming, J., C. Kirby and B. Ostdiek (1998), Information and Volatility Linkages in the Stock, Bond and Money Market, *Journal of Financial Economics*, Vol. 49, p. 111-137.

Fung, G.P.C. and J.X. Yu and W. Lam (2002), News sensitive stock trend prediction, in: *Proceedings 6th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, p. 481–493.

Fung, G.P.C., J.X. Yu and H. Lu (2005), The Predicting Power of Textual Information on Financial Markets, in: *IEEE Intelligent Informatics Bulletin*, Vol. 5, No.1.

Gilovich, T., D. Griffin and D. Kahneman (2002), *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press.

Grossmann, S.J. and J.E. Stiglizt (1980), On the Impossibility of Informationally Efficient Markets, *American Economic Review,* Vol. 70, Issue 3, p. 393-408.

Herzog, B. (2010), Leading Indicators for Asset Bubbles in Real and Financial Markets, 30th CIRET Conference, hosted by 'The Conference Board', *Conference Papers and Proceedings*, New York, NY, 2010.

Hong and Stein (1999), A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance*, Vol. 54, Issue 6, p. 2143-2184.

Hou, K., L. Peng and W. Xiong (2008), A tale of two anomalies: The implications of investor attention for price and earnings momentum, *Working Paper*, Ohio State University and Princeton University.

Huang, B.-N. (1995), Do Asian stock market prices follow random walks? Evidence from the variance ratio test, *Applied Financial Economics*, Vol. 5, Issue 4, p. 251-256.

Jain, P.C. (1988), Response of Hourly Stock Prices and Trading Volume to Economic News, *The Journal of Business*, Vol. 61, No. 2, p. 219-23.

Jedgadeesh, N. and S. Titman (1993), Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance,* Vol. 48, p. 65-91.

Kahneman, D. (1973), *Attention and Effort*, Prentice-Hall, NJ.

Karolyi, G.A. and L. Gagnon (2003), Information, Trading Volume and International Stock Market Comovements, *International Finance Review*, Vol. 4, p. 347-377.

Kothari, S.P., J. Lewellen and J.B. Warner (2006), Stock returns, aggregate earnings surprises, and behavioral finance, *Journal of Financial Economics*, Vol. 79, Issue 3, p. 537-568.

Li, X., R.S. Mahani and V. Sandhya (2011), Does Investor Attention Affect Stock Prices?, *Working Paper*, SSRN No. 1748851.

Lucas, R.E. (1978), Asset Prices in an Exchange Economy, *Econometrica*, Vol. 46, Issue 6, p. 1426-1445.

Lynch, A.W. and R.R. Mendenhall (1997), New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index, *The Journal of Business*, Vol. 70, No. 3, p. 351-383.

Mankiw, N.G. (1986), The Equity Premium and the Concentration of Aggregate Shocks, *NBER Working Paper*, No 1788.

Mehra, R. and E. Prescott (1985), The Equity Premium: A Puzzle, *Journal of Monetary Economics*, Vol. 25, Issue 2, p. 145-162.

Niederhoffer, V. (1971), The Analysis of World Events and Stock Prices, *The Journal of Business*, Vol. 44, No. 2, p. 193-219.

Rigobon, R. and B. Sack (2004), The Impact of Monetary Policy on Asset Prices, *Journal of Monetary Economics*, Vol. 51, Issue 8, p. 1553-1575.

RKI (2011), *E.Coli Data of Germany*, Berlin, Germany.

Shankar, S.G. and J.M. Miller (2006), Market Reaction to Changes in the S&P SmallCap 600 Index, *The Financial Review*, Vol. 41, No. 3, p. 339-360.

Shell, A. (2007), Technology squeezes out real, live traders, http://www.usatoday.com/money/markets/2007-07-11-nyse-traders_N.htm (17.09.2010, 17:52).

Shiller, R.J. (1990), Speculative Prices and Popular Models, *Journal of Economic Perspectives,* Vol. 4, Issue 2, p. 55-65.

Tetlock, P.C., M. Saar-Tsechansky and S. Macskassy (2007), More Than Words : Quantifying Language to Measure Firms' Fundamentals, in: *Proceedings of 9th Annual Texas Finance Festival.*

Vassalou, M. (2003), News related to future GDP growth as a risk factor in equity returns, *Journal of Financial Economics*, Vol. 68, Issue 1, p. 47-73.

Vega, C. (2006), Stock price reaction to public and private information, *Journal of Financial Economics*, Vol. 82, Issue 1, p. 103-133.

Verrecchia, R.E., O. Kim (1991), Trading Volume and Price Reactions to Public Announcements, *Journal of Accounting Research*, Vol. 29, No. 2, p. 302-321.

Winkler G. and B. Herzog (2010), Google Statistics as Leading Indicators: A Cross-Country Study, (under review).

Appendix

Table 1A: Data Overview

List of Banks according to Countries		
Country	**Name of Bank**	**Time-series of Financial Data**
USA	JP MORGAN CHASE & CO.	March 1990 - March 2010
	WELLS FARGO & CO	March 1990 - March 2010
	BANK OF AMERICA	March 1990 - March 2010
Japan	MITSUBISHI UFJ FINL. GP.	April 2001 - March 2010
	MIZUHO FINL. GP.	September 2000 - March 2010
	SUMITOMO MITSUI FINL. GP.	December 2002 - March 2010
Germany	DEUTSCHE BANK	January 1999 - March 2010
	COMMERZBANK	Abril 2000 - March 2010
	DEUTSCHE POSTBANK	June 2004 - March 2010
UK	HSBC HDG.	July 1992 - March 2010
	BARCLAYS	April 1990 - March 2010
	ROYAL BANK OF SCTL. GP.	April 1990 - March 2010
France	BNP PARIBAS	October 1993 - March 2010
	CREDIT AGRICOLE	January 2002 - March 2010
	SOCIETE GENERALE	June 1991 - March 2010
Italy	UNICREDIT	April 1993 - March 2010
	INTESA SANPAOLO	July 1993 - March 2010
	BANCA MONTE DEI PASCHI	June 1999 - March 2010
Canada	ROYAL BANK CANADA	April 1990 - March 2010
	TORONTO-DOMINION BANK	March 1990 - March 2010
	BANK OF NOVA SCOTIA	March 1990 - March 2010
Spain	BANCO SANTANDER	March 1990 - March 2010
	BBV.ARGENTARIA	April 1990 - March 2010
	BANCO POPULAR ESPANOL	March 1990 - March 2010
China	INDUSTRIAL & COML. BANK. OF CHINA	October 2006 - March 2010
	CHINA CONSTRUCTION BANK	September 2007 - March 2010
	BANK OF CHINA	July 2006 - March 2010
Switzerland	CREDIT SUISSE GROUP	April 1990 - March 2010
	UBS	March 2000 - March 2010

Table 2A: Descriptive Statistics

	Price	Return	Index	Google Search Volume (GSV)	Google Insights Search Volume (GISV)
Mean	489.080	0.000341	8515	1.047	16.293
Median	45.465	0.000000	6852	1.010	16.400
Maximum	13800.00	0.648000	40775	8.100	31.930
Minimum	0	-0.789091	458.13	0.000	1.000
Std. Dev.	1697	0.063332	5750	0.453	8.767
Skewness	5.358	-0.026439	2.089	1.832	0.008
Kurtosis	32.858	17.780	10.014	19.380	1.814
Observations	7844	7844	7844	7844	7844

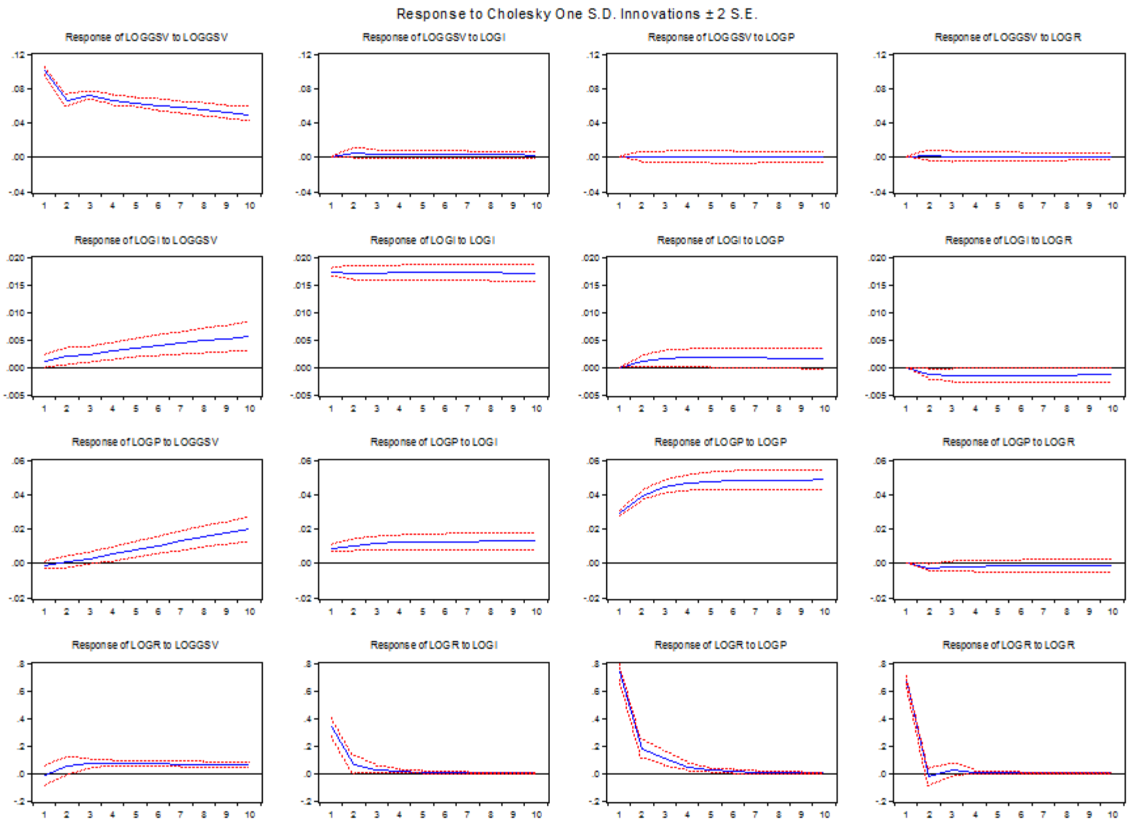
Table 3A: Test of Fixed-Effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	2.209	-29,3514	0.0002
Cross-section Chi-square	6.969	29	0.0000
Period F	2.779	-309,3514	0.0000
Period Chi-square	843.308	309	0.0000
Cross-Section/Period F	2.833	-338,3514	0.0000
Cross-Section/Period Chi-square	929.579	338	0.0000
Note: Test cross-section and period fixed effects			

Table 4A

	LOGGSV	LOGI	LOGP	LOGR
LOGGSV(-1)	0.651758	0.009294	0.022211	0.565821
	(0.03145)	(0.00539)	(0.00953)	(0.32919)
	[20.7217]	[1.72426]	[2.33102]	[1.71885]
LOGGSV(-2)	0.284155	-0.002365	0.002462	0.199280
	(0.03153)	(0.00540)	(0.00955)	(0.32994)
	[9.01362]	[-0.43781]	[0.25780]	[0.60399]
LOGI(-1)	0.241725	0.985536	-0.070869	0.727461
	(0.18828)	(0.03227)	(0.05704)	-197.052
	[1.28387]	[30.5440]	[-1.24248]	[0.36917]
LOGI(-2)	-0.237467	0.013311	0.068535	-0.777628
	(0.18793)	(0.03221)	(0.05693)	-196.688
	[-1.26359]	[0.41330]	[1.20379]	[-0.39536]
LOGP(-1)	-0.044190	0.086800	1.432.518	6.955.889
	(0.13704)	(0.02349)	(0.04152)	-143.430
	[-0.32245]	[3.69582]	[34.5042]	[4.84967]
LOGP(-2)	0.040823	-0.088107	-0.432054	-6.948.909
	(0.13710)	(0.02349)	(0.04153)	-143.484
	[0.29777]	[-3.75008]	[-10.4027]	[-4.84298]
LOGR(-1)	0.002565	-0.002005	-0.003733	-0.038709
	(0.00429)	(0.00074)	(0.00130)	(0.04492)
	[0.59753]	[-2.72608]	[-2.87062]	[-0.86171]
LOGR(-2)	-0.000110	-1.40E-05	0.002182	0.057313
	(0.00304)	(0.00052)	(0.00092)	(0.03179)
	[-0.03606]	[-0.02694]	[2.37127]	[1.80280]
C	-0.020652	0.017417	0.031359	-3.713.076
	(0.05155)	(0.00883)	(0.01562)	(0.53952)
	[-0.40063]	[1.97147]	[2.00805]	[-6.88221]
Observations	908	908	908	908
R-squared	0.901847	0.999331	0.999658	0.130897
Adj. R-squared	0.900973	0.999325	0.999654	0.123163
S.E. equation	0.100986	0.017307	0.030594	1.056.922
F-statistic	1.032.516	167747.8	328003.4	1.692.492
Akaike AIC	-1.737.800	-5.265.606	-4.126.189	2.958.461
Schwarz SC	-1.690.111	-5.217.918	-4.078.500	3.006.150
Standard errors in ( ) & t-statistics in [ ]				

Figure 4A



Response to Cholesky One S.D. Innovations ± 2 S.E.
Response of LOGGSV to LOGGSV
Response of LOGGSV to LOGI
Response of LOGGSV to LOGP
Response of LOGGSV to LOGR
Response of LOGI to LOGGSV
Response of LOGI to LOGI
Response of LOGI to LOGP
Response of LOGI to LOGR
Response of LOGP to LOGGSV
Response of LOGP to LOGI
Response of LOGP to LOGP
Response of LOGP to LOGR
Response of LOGR to LOGGSV
Response of LOGR to LOGI
Response of LOGR to LOGP
Response of LOGR to LOGR