Working Paper
Series

National Centre of Competence in Research
Financial Valuation and Risk Management

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 638

Measuring Economic Uncertainty and its Impact on the Stock Market

Michal Dzielinski

First version: May 2009
Current version: October 2010

This research has been carried out within the NCCR FINRISK project on "Behavioural and Evolutionary Finance"

Die Nationalen Forschungsschwerpunkte (NFS) sind ein Förderinstrument des Schweizerischen Nationalfonds.
Les Pôles de recherche nationaux (PRN) sont un instrument d'encouragement du Fonds national suisse.
The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation.

Schweizerischer Nationalfonds
Fonds national suisse
Swiss National Science Foundation

Measuring economic uncertainty and its impact on the stock market*

Michal Dzielinski†

First draft: 05/09
Current draft: 10/10

Abstract

The issue of economic uncertainty is becoming interesting, as it is being more widely recognized that not only investors' expectations but also the confidence put in them matter for stock returns. The problem with most existing indicators is that they are based on market data and so measure the effect *ex post*. In this paper I propose a new one, which I believe is well suited to measure the effect *ex ante*. It is based on the frequency of internet searches for the word "economy" as reported by Google Trends. The underlying argument is that uncertainty induces increased information seeking.

FINRISK project: IP AP1
Keywords: AP, EB

*I am especially indebted to Marc Oliver Rieger for making me aware of existence of Google Trends. Comments from prof. Bernard Dumas, Elise Paysan and other participants of the 2009 FINRISK Workshop at Gerzensee were very helpful in the early stages of this research, while those from Urs Schweri, Sven Christian Steude and Elisa Wu greatly contributed to its current version. The supervision provided by prof. Thorsten Hens was very valuable throughout. I am also thankful to Amid Goyal for making his data library available. The financial support from LGT and Science is gratefully acknowledged. Part of this research has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK). The NCCR FINRISK is a research instrument of the Swiss National Science Foundation. As always, I honor my parents for bringing me to where I am now.

†Swiss Banking Institute, University of Zurich, Plattenstrasse 32, 8032 Zurich, Switzerland, dzielinski@isb.uzh.ch

1 Introduction

This paper makes the case for the use of internet search data in modeling economic uncertainty. In the first stage I show that the proposed measure appropriately captures uncertainty among stock market participants. The departing point is that more uncertain investors seek information more intensively and that this can be represented by the volume of internet searches. This is backed by theory derived from information sciences as well as by empirically found co-movement of an indicator constructed from search volumes reported by Google Trends and measures of uncertainty suggested elsewhere. The second stage is focused on showing the relevance of the proposed measure, by documenting a negative and statistically significant impact on stock returns. The indicator based on internet search volume delivers well both on a weekly and monthly basis. Its impact is not endogenously generated by past returns and also not driven by the financial crisis, which happened in the second half of the analyzed sample. This performance is more convincing than that of other measures found in the literature or the industry. For practical purposes, the comparatively high, weekly frequency is an additional advantage.

The conceptual background for my approach is the stylized model by Wilson (1999), postulating the resolution of uncertainty as a principal motivation for information seeking at any of its three stages: problem identification, problem definition and problem resolution. The empirical evidence behind this approach has been summarized in the abovementioned paper as well as Kuhlthau (1993). Taken onto financial ground it has compelling implications: the greater the uncertainty among investors, the more intensive should their search for information be. Therefore, if one could measure information seeking among investors in a timely fashion, it would hopefully allow gauging changes in their uncertainty before any actions conditional on the level of uncertainty are taken. Considering that the stock market is a fast moving environment, high frequency of any such measurement is very important. An even more fundamental problem arises from the fact that information seeking is an activity performed at the individual level and at the user's discretion, thus inherently difficult to observe outside a controlled environment (e.g. an experiment), which in turn greatly limits the scale of any possible measurement. I argue that what seems like squaring the circle, designing a valid measure of information seeking available on a mass scale, can actually be solved thanks to the immense expansion of internet in the last decade.

This paper has overlaps with several other areas of the finance literature. On one hand the benefits of Google Trends have already been exploited for predicting sales (Choi and Varian (2009a)), jobless claims (Choi and Varian (2009b)), flu outbreaks (Dukic et al. (2009)), private investor demand and IPO returns (Da et al. (2009)) as well as modeling volatility asymmetry (Dzielinski et al. (forthcoming)). On the other hand are the studies addressing the importance of uncertainty for stock returns. Boguth and Kuehn (2009), Bollerslev et al. (2009) and Tedongap (2007) all propose models of uncertainty concerning consumption growth and also derive empirical measures, which are shown to be relevant for stock returns. Otherwise, the issue of investor uncertainty (or its flipside, confidence) has attracted attention within the financial industry with indicators published by companies like Barron's or State Street. The main contribution of this paper lies in applying insights from both sides to design a measure of uncertainty, which is exogenous yet correlated with the market and thus able to deliver useful and timely signals. It also extends the applicability of Google Trends, which has already gained some recognition as an orthogonal data source for financial modeling.

In the next section I present the details of constructing the uncertainty measure from internet searches. Section 3 contains the results of the empirical analysis based on US data. Section 4 deals with robustness checks. Section 5 extends the results to other stock markets. The final section concludes.

2 Internet searches and economic uncertainty

Internet penetration in developed countries now exceeds 75% of households according to www.internetworldstats.com and its role for various aspects of life but especially information exchange and retrieval is unquestionable. As unusual as it may seem, going online even from the privacy of one's own home leaves a much larger footprint, than many of the old-fashioned activities prevalent before. For instance, buying a newspaper from a newsstand does not generally involve recording any personal information, however accessing the online edition of the same newspaper actually does. Similarly, whenever a user performs and internet search, the information on its content and origin is, or at least can be, recorded by the provider of the search engine. If done systematically and on a large scale, such data would be ideally suited for tracking information seeking activities in the real world environment. Two obstacles were preventing such solutions from emerging. The first one was simply the massive data volumes generated in the process that were even impossible to store, let alone analyze statistically. This eventually became feasible as computational power continued to grow exponentially. The second one was more subtle and had to do with the fragmentation of the internet, including the search engine segment with a multitude of providers with roughly the same market share. This changed when Google redefined the playing field in the late 90's, thanks to its superior search algorithms. By June 2000 it was the biggest search engine, as measured by the number of indexed pages, and soon after also the most popular. In recent years it has consistently accounted for an estimated 70% of global search traffic. These developments made the concentration of a large enough chunk of internet search data in one place possible and indeed Google released them to the public with the launch of Google Trends in 2006. Roughly speaking it is a tool designed for tracking the relative popularity of any given search term over time. The dataset goes back to 2004 and is updated weekly. It is also scaled by the total search traffic, so as to conceal the actual number of Google users, and presented in the form of a search volume index (SVI). In 2008 a sister application was launched under the name Google Insights for Search, which includes a very useful extension, allowing filtering the results by category to determine e.g. what users searching for "apple" were actually interested in (two of the over 20 categories are "Computers & Electronics" and "Food & Drink"). Based on these two sources I construct a measure to capture the information seeking of investors and thus their degree of uncertainty.

The key issue is how to filter out the relevant content of internet searches, which can concern, broadly speaking, everything. In Google Trends this is done by selecting appropriate keywords. One thing to avoid is the hindsight bias. Today we have the comfort of knowing that recession, oil price and subprime were topics of great interest to investors over the past few years (unreported results actually show that an indicator based on a combination of those does a pretty good job) but back in 2005 the choice would have probably been different. The problem is to find a keyword that is general and time-invariant, while maintaining sufficient relevance. The simplest yet reasonable answer is the word "economy" itself. It definitely is sufficiently broad to encompass all possible sources of economic uncertainty, including the three mentioned above and also to be time independent. On the other hand it appears specific enough to contain noise that is either relatively small or at least constant over time because it is not used to describe any other concept, whose popularity might be correlated with the one of interest (this might be the case for instance with the word "depression"). If the assumption of constant noise is correct, the trend component should represent the relevant part of information seeking. I isolate it by dividing the current value by the value of the corresponding week one year ago, which also helps deal with the observed seasonality. Its dynamics is depicted in Figure 1b, it is relatively stable until mid 2007 and increases sharply afterwards. The peak coincides with the bankruptcy of Lehman Brothers, after which a gradual return to pre-crisis levels takes place. This kind of dynamics does make intuitive sense.

To confirm this intuition I compare the year-to-year series obtained from Google Trends with the SVI for "economy" obtained from the Google Insights "Finance & Insurance >> Investment" category and find them to be very similar both visually (see Figure 1c) and statistically. The correlation between is high and positive and the means over the examined time period almost identical, though the Google Insights SVI is more volatile. Why not use the Google Insights directly in the analysis then? For one, it is somewhat less convenient to handle due to the fact that each time series is scaled to its peak (fixed at 100), so in ongoing applications rescaling would be necessary whenever a new peak appeared. More importantly, I discover that the Google Trends year-to-year series actually significantly leads the Google Insights series with a quite large R-squared, as documented in Table 1. The likely reason is the slight shift in the weekly time window used for computing the SVI values, which is Monday-Sunday for GT and Sunday-Saturday for GI. Therefore, the Sunday, which still contributes to the current week for GT is already part of the next week for GI. Therefore, while acknowledging the important corroboration provided by Google Insights, I stick to the Google Trends year-to-year SVI in the further analysis.

There are two other options in the data presentation, which are important to mention here. The first one is the regional filter, based on the *origin* of the query determined from the user's IP address. For the initial analysis I use the S&P 500 as the financial universe and to limit the confounding influence of searches from parts of the world unrelated to the US stock market, I only consider those originating in the US itself. This is admittedly a crude way to deal with the problem, as it implicitly assumes that only domestic investors impact the stock market. However, it is not a significant limitation in practice since the global SVI for "economy" does not fundamentally differ from the US one. The other, more subtle yet more influential, option concerns how the data are scaled, either to a fixed or relative reference frame. The first approach applies the average search traffic in a fixed time period (generally January 2004) as a benchmark value, while otherwise the average for the whole specified time period is used. This aspect is again important in the context of the hindsight bias, since the average for the whole period is something known only at the end and not before. Therefore, I always use fixed scaling.

3 Empirical results

Alternative ways to measure uncertainty are a quite heterogeneous group with respect to the underlying concept and data. First of all, some of them are formulated as measures of confidence, while others as measures of uncertainty. It is meaningful to compare them, because confidence can be seen as the opposite of uncertainty, so the same logic applies to both just with opposite signs.

One such measure can also be obtained on a weekly basis and it is the Barron's Confidence Index. It is defined as the ratio of yields on two bond portfolios, the first one consisting of ten high grade corporate bonds and the second one consisting of ten medium grade corporate bonds, selected by Barron's. The underlying logic is that an increase in the ratio, corresponding to the tightening of the spread between the two yields, indicates rising confidence and is a bullish sign for the stock market.

A broader selection is available on a monthly basis. The Yale School of Management publishes the "One-year confidence index" every month, separately for institutions and individuals, an idea originally developed by Robert Shiller. It is calculated based on a monthly survey of a sample of institutional (in this case investment managers of US pension funds) and wealthy individual investors asking, among others, what change (in percent) they expect in the Dow Jones 30 over the following 12 months. The index is then the percentage of strictly positive responses averaged over a rolling window of six months, with an increase corresponding to higher confidence. A different approach is taken by State Street, based on the vast data on asset flows it gathers as the world's biggest institutional custodian. Roughly speaking, as institutional portfolio holdings shift towards riskier assets, this leads to an increase in the respective confidence index. Finally, Bollerslev et al. (2009) advocate the variance premium as an explicit measure of uncertainty. It is defined as the difference between implied and realized variance, where the first one is based on the VIX and the second calculated from high-frequency returns.

One more measure, suggested to explicitly capture uncertainty, was elaborated by Boguth and Kuehn (2009). The idea is to model volatility of consumption growth as a Markov-type process with a high and low state, and let prior beliefs about the probability of being in the high volatility state correspond to the degree of uncertainty. Since the calculations are based on consumption data, it is only available quarterly. In a broader context the conditional volatility of consumption has also been applied to modeling asset returns by Bansal et al. (2005); Tedongap (2007).

Based on their methodology, those measures can be put into one of two groups. The first group aims to investigate investors' uncertainty directly, by asking them questions (which was probably the only way to do it before Google Trends) and so I call it survey-based. Both Shiller's measures fall in to that group. All other measures employ some kind of proxy based on financial activity: bond yields, volatilities, portfolio flows and therefore I call them market-based. The survey-based measures deserve praise for their attempt at directness, common critique concerns honesty of the answers and limited scale (sample size). On the other hand, the market-based measures are based on actions (this argument is especially underscored by State Street) and have a large scale but they measure already the financial impact of uncertainty and so are not very timely. The SVI has the advantages of both, while avoiding their drawbacks: it is direct yet based on actions rather than statements, so honesty is not an issue and the scale is virtually unlimited.

This claim is substantiated by directly comparing the empirical performance of the SVI and the other measures. The sample consists of either weekly or monthly observations for the years 2005 - 2009. This rather short time period is forced by the relative novelty of Google Trends. For this reason the quarterly consumption based measure had to be dropped from the comparison. Historical data for the Shiller's one-year confidence index and the

State Street confidence index were obtained directly from the websites of their respective providers. The variance premium was calculated for the desired time period using monthly VIX values from Datastream and 5-minute returns on the SPY (the most liquid exchange traded fund tracking the S&P 500) from TAQ. To replicate the Barron's index I used weekly data on the Moody's Aaa and Baa bond yields from the Reserve Bank of St. Louis. In the linear regressions weekly and monthly excess returns on the S&P 500, used as the dependent and control variables, are based on index data from Datastream and the Treasury bill rates provided by the Reserve Bank of St. Louis.

Table 2 presents summary statistics of all the measures as well as correlations between them. The monthly SVI is aggregated from weekly values by taking simple averages. Whenever a week is split, it is included in the later month. Apart from the obvious smoothing, the overall dynamics does not change much due to that process (see Figure 2). The most volatile measure is by far the variance premium (mostly because of extreme values registered during the peak of the financial crisis) followed by the two SVIs. Other measures fluctuate comparatively little. Concerning the co-movements, there is no surprise at the weekly level with a strong negative correlation between the SVI (measuring uncertainty) and the Barron's Index (measuring confidence). At the monthly level there are some possibly confusing results. For instance the two measure of institutional confidence (SHinst and ST) are virtually uncorrelated with each other (coeff. = -0.17 and insignificant). One way to explain this is the discrepancy between words and deeds mentioned before. However, even assuming both results are honest (as it is maybe better to do without evidence to the contrary) they can still be reconciled. It appears that institutional investors can remain confident about the long term outlook, captured by the Shiller measure, while being uncertain right now, as depicted in current portfolio flows underlying the State Street index. For individual investors, both short and long term would seem to be tied together (no correlation with SHinst, significant and positive with ST) which is actually consistent with them following a simpler, extrapolative market heuristics. In this context, the correlations between the monthly SVI and the three measures just mentioned lend support to saying that the SVI captures uncertainty of both types of investors.

The negative relationship between the variance premium and the SVI, which are both measures of uncertainty, is in fact less surprising than it might seem. The variance premium is constructed to help predict the expected return in equilibrium. The transition from one equilibrium to the next requires some adjustment however. For the expected return to increase, prices have to fall, so the return over this adjustment period will in fact be negative. This is one of the reasons why equilibrium measures perform best at longer time horizons, where the adjustment can be said to take place between the measurements (see also the original results of Bollerslev et al. (2009). On the other hand, the SVI measure is designed to capture the uncertainty of investors at an earlier stage, before it has any impact on the market. This feature is corroborated by the fact that the SVI leads the variance premium with a convincing statistical significance, while the opposite is not true. According to the results in Table 3 an increase today in uncertainty as measured by the SVI will be picked up the variance premium two months after and should lead to higher expected returns. In the meantime however there is negative relationship in place, which means over the course of the first month lower expected returns should be observed. The direction of the influence at different lags is consistent with the adjustment argument. It appears that thanks to being based only on information exogenous to the market, the SVI does not get confused by price adjustments over shorter time horizons and thus can be used in a more timely fashion.

Regression analysis summarized in Table 4 confirms this point of view. The dependent variable is the log excess return on the S&P 500 over the following time period, which can be either one week or one month. In the first stage the only regressors are the respective uncertainty or confidence measures and a (unreported) constant term. The t-statistics are heteroskedasticity consistent following White. The SVI comes out significant and negative at both the weekly and monthly horizon and the corresponding R-squared values are noticeable. There are two other measures yielding statistically significant results, the State Street index and the variance premium. The positive sign is consistent with their declared impact and the degree of significance is comparable to the SVI. In terms of R-squared, the State Street measure is slightly behind the SVI, the one for the variance premium however it is only half as large.

In the second stage these three measures are tested for exogeneity by repeating the regressions and this time controlling for current period market return. Interestingly, this improves the performance of the weekly SVI both in terms of magnitude and significance. On the monthly level the parameter estimates are closer to zero and the t-statistics accordingly lower, suggesting there is some correlation between the explanatory variables. For the variance premium this actually means a loss of significance, which was a likely outcome, given that it is based on realized and implied volatility, which both somehow depend on the market return. The two other measures however retain their significance, which suggest that they contain some exogenous information about the dynamics of the stock market.

4 Robustness

In this section I explore various issues, which could undermine the results or limit their generality. These relate to the impact of the financial crisis and the sensitivity of the measure to the choice of search terms.

4.1 Subperiod analysis

A legitimate concern is to what extent the results could be driven by the financial crisis, especially when it covers roughly half of the sample. To check for this I divide the sample into two approximately equal parts, labeled "pre-crisis" and "crisis". This is reasonable, in the sense of maintaining sufficient sample size, for the weekly data only. The first subsample covers the period January 2005 till the end of June 2007, while the second extend from July 2007 till the end of 2009. The dividing point was chosen more for convenience (subsamples are almost equal in size), yet it is also reasonable: by then the market has fallen by about 10% from its peak and thanks to the announcements made by Bear Sterns about problems in two of its hedge funds "subprime" has become a common concern among investors. In any case the conclusions are not affected by moving the threshold by up to four months in either direction (results not reported).

The first column of Table 5 shows that the uncertainty was on average growing during the crisis and was much more volatile than before. Its impact on returns is consistently negative and statistically significant before as well as during the crisis. The magnitude of the impact is roughly 50% bigger during the crisis, which says that not only was there more uncertainty but investors were also more sensitive to it. Interestingly, the control variable, which is as previously the current market return, comes out insignificant during the crisis, while being significant before. This speaks for the fact that in uncertain, volatile times past returns cannot be relied upon to predict future returns. This lends further support to using exogenous indicators, such as uncertainty.

4.2 Wording of the search term

Table 5 also shows the effects of the exact choice of words used to retrieve search volume data from Google Trends. Even though the word economy was chosen carefully as to not have multiple confusing meanings (or also synonyms) there is still the problem of words, which are grammatically close but in terms of meaning not necessarily so. It is important to check those cases, because if the Google Trends classifier cannot distinguish between them, then the original results cannot be trusted.

To this end I select three alternative search terms, which are all grammatically close to "economy". Comparing the content they reflect, one is deemed virtually identical ("economic"), one is somewhat related ("economics") and one is obvious noise ("fuel economy"). The last alternative is possible because Google Trends allows expressions consisting of several words to be treated as a single search term. One would expect a similar impact for the first case and no significant impact for the last one, so it is a two sided test. The comparison is performed separately for the whole sample as well as the two subsamples, before and during the financial crisis. The results are generally favorable for using Google Trends, if not without reservations. It is reassuring to note that the term "economic" performs very similarly to the original one across all time periods. What is also apparent for the whole sample is a decrease in statistical significance and explanatory power as the meaning moves away from the original, though the "noise case" is still significant at the 10% level. The subsample analysis reveals that this holds only for pre-crisis period. The likely reason is that in this period the original measure has very little variability itself, so it might be harder to distinguish from noise. During the crisis, when there is more search activity, this problem is reduced. The interesting conclusion is that the signals become more reliable as the economic environment becomes more turbulent, contrary to most conventional indicators. This makes Google Trends all the more useful.

5 Regional comparison

The global presence of Google makes a regional comparison compelling. Following the structure of the regional filter in Google Trends it is based on search volume for countries and the most popular indices of the respective stock exchanges. The problem of considering only domestic investors is likely to become more severe for smaller, open exchanges and a better idea for future applications could be to take an average of the SVIs from different countries weighted by the presence investors from those countries have on that stock exchange. However, domestic investors should also account for part of the variation in returns, so as a first validity check for using internet searches to measure uncertainty in different countries this crude approach is sufficient.

Another issue, specific to the regional analysis, is language. The Google Trends interface does include a breakdown of the search volume by languages but this is misleading as it only relates to the language version of the Google site where the query was initiated and not the language of the query itself. For instance the number for German when retrieving search volumes for "economy" for Germany only means that some users in Germany went to google.de (rather than google.com) and searched for "economy" from there. Comparing search volumes reveals that German users searching for "economy" are just a small fraction of those using the German equivalent term "Wirtschaft". Therefore, to maintain a claim on representativeness I use the local language equivalent (also for Japan).

Language still seems to matter, even after those adjustments, as results in Table 6 show. Across the four English speaking countries the picture is very similar for the whole sample as well as the crisis period. Only for the UK and Canada before the crisis the results are not significant, which is in line with the lower reliability of the Google Trends measure overall in this period. This is stark contrast to the case of Germany and Japan, where no significant relationship can be established regardless of the time period. In the current setup there are several possible reasons, either domestic investors have no impact in those markets (not very likely) or they do not use internet for information (possible but also unlikely) or they search for something else than the direct translation of the word "economy". In any case, this means that Google Trends is not necessarily easily extendable across languages without some feeling of the vocabulary.

6 Conclusions

The results of this study successfully establish a novel indicator of economic uncertainty. It is based on the search volume index (SVI) for the word "economy" provided by Google Trends, which monitors all activity flowing through the Google search engine. As such it is shaped directly by the behavior of agents and so has arguably good signalling properties. The underlying intuition is that a higher level of uncertainty increases the demand for information, which in a modern society should be visible in internet search data. Consistent with this intuition an indicator derived for the US increases after the beginning of the subprime crisis, peaks around the collapse of Lehman Brothers and decreases since. It is also negatively correlated with some widely accepted measures of investor confidence. In-sample regressions reveal a significant negative relationship with future aggregate market returns, which holds when controlling for simultaneous returns and as such is not purely endogenous, a property in general not given for other uncertainty / confidence measures. The results are robust to changing the time period as well as the problem of the open texture of any natural language. On the other hand, some limitations of the indicator also emerge over the course of the analysis. In periods of generally low uncertainty the actual signal might hard to distinguish from noise. Also, the results are readily generalizable to other English speaking countries but not to other languages. None of this however really undermines using the indicator for what it was originally intended, identifying periods of stress, when high uncertainty can lead to large (and negative!) price reactions.

The analysis itself also has some limitations. There were only 6 years of weekly observations available, which mandates the use of but the simplest estimation techniques, especially at the monthly frequency. Therefore, little could be said about the internal dynamics of the indicator and optimal noise filtering. Future research should address those issues as the problem of data volume naturally diminishes with time.

References

R. Bansal, V. Khatchatrian, and A. Yaron. Interpretable asset markets? *European Economic Review*, 49:531–560, 2005.

O. Boguth and L.A. Kuehn. Consumption volatility risk. In *EFA Bergen Meetings Paper*, 2009.

T. Bollerslev, G. Tauchen, and H. Zhou. Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492, 2009.

H. Choi and H. Varian. Predicting the present with google trends. *Google Inc.*, 2009a.

H. Choi and H. Varian. Predicting initial claims for unemployment benefits. *Google Inc.*, 2009b.

Z. Da, J. Engelberg, and P. Gao. In search of attention. *SSRN Working Paper*, 2009.

V. Dukic, H. Lopes, and N. Polson. Tracking flu epidemics using google flu trends and particle learning. *SSRN Working Paper*, 2009.

M. Dzielinski, M.O. Rieger, and T. Talpsepp. Volatility asymmetry, news and private investors. In *News Analytics in Finance Handbook*. John Wiley & Sons, Chichester, forthcoming.

C. Kuhlthau. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4):339–355, 1993.

R. Tedongap. Consumption volatility and the cross-section of stock returns. *SSRN working paper*, 2007.

H. White. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48(4):817–838, 1980.

T.D. Wilson. Models in information behaviour research. *Journal of Documentation*, 55(3): 249–270, 1999.

Table 1: Google Trends vs. Google Insights

Panel 1 presents the descriptive statistics for the two series and their correlation. In Panel 2 each row represents the equation for the respective variable with the constant and lagged parameters given in columns. SVI_{GT} is the year-to-year Google Trends index of the search volume for "economy" in the US. SVI_{GI} is the Google Insights index of the search volume for "economy" in the US, restricted to the "Finance&Insurance >> Investment" category. The sample period is Jan 2005 to Dec 2009.

Panel 1: descriptive statistics			
	Mean	Std. dev.	Correlation
SVI_{GT}	1.02	0.29	0.76
SVI_{GI}	1.24	0.78	

Panel 2: VAR(2)						
	Const. (t-stat)	$\text{SVI}_{GI,t-1}$ (t-stat	$\text{SVI}_{GT,t-1}$ (t-stat)	$\text{SVI}_{GI,t-2}$ (t-stat)	$\text{SVI}_{GT,t-2}$ (t-stat)	adj.-R^2(%)
SVI_{GT}	0.06	0.02	0.74***	-0.02	0.21**	85.0
	(1.89)	(1.19)	(12.0)	(-1.59)	(2.95)	
SVI_{GI}	-0.75***	0.27***	1.91***	0.01	-0.31	69.0
	(-6.19)	(4.39)	(8.04)	(0.24)	(-1.15)	

Table 2: Comparison with other uncertainty/confidence measures

VP - variance premium
BR - Barron's Confidence Index
SH - Shiller's one-year confidence index, separately for institutions and individuals
ST - State Streets Investor Confidence Index
SVI_{GT} - is the year-to-year Google Trends search volume index for "economy" in the US
The sample period is Jan 2005 to Dec 2009.
* - correlation significant at the 0.1 level

	Weekly		Monthly				
	SVI_{GT}	BR	SVI_{GT}	VP	SH_{inst}	SH_{ind}	ST
Mean	1.02	81.6	1.02	6.7	79.4	79.6	105.4
Std. dev.	0.29	7.3	0.28	41.6	4.1	5.0	8.9
Corr.	1	-0.43*	1	-0.43*	0.20	-0.66*	-0.62*
		1		1	-0.10	0.01	0.32*
					1	-0.01	-0.17
						1	0.46*
							1

Table 3: Relationship between the SVI and VP measures

Each row represents the equation for the respective variable with the lagged parameters given in columns. VP is the variance premium. SVI is the year-to-year index of the search volume for "economy" in the US. The sample period is Jan 2005 to Dec 2009.

	Const. (t-stat)	VP_{t-1} (t-stat	SVI_{t-1} (t-stat)	VP_{t-2} (t-stat)	SVI_{t-2} (t-stat)	adj.-R^2(%)
VP	26.6	0.05	-170.3***	0.19	149.2***	33.6
	(1.27)	(0.44)	(-5.12)	(1.58)	(4.35)	
SVI	0.1	0.001	0.84***	< -0.001	0.05	74.0
	(1.12)	(1.33)	(6.10)	(-0.67)	(0.38)	

Table 4: Market return regressions

Ret_{t+1} is the next period logged return on the S&P 500 in percentage points. Ret_t is the analogous value for the current period. All other variables are defined as in Table 2 and also in logarithms. The sample period is Jan 2005 to Dec 2009.
t-statistics are computed using heteroscedasticity consistent covariance matrices following White (1980)
significance levels: * - 0.05, ** - 0.01, *** - 0.001

	Monthly (N = 60)					Weekly (N = 260)	
	SVI_{GT} (t-stat)	ST (t-stat)	SH_{inst} (t-stat)	SH_{ind} (t-stat)	VP (t-stat)	SVI_{GT} (t-stat)	BR (t-stat)
Panel 1: with constant only							
Ret_{t+1}	-7.9**	21.1**	-9.9	11.3	0.03**	-3.1**	-0.01
	(-2.56)	(3.08)	(-1.38)	(1.12)	(2.57)	(-2.57)	(-0.12)
adj.-R^2(%)	16.3	13.7	0.0	0.0	7.9	7.1	0.0
Panel 2: controlling for current market return							
Ret_{t+1}						-3.4***	
						(-3.17)	
Ret_{t+1}	-6.1*						
	(-1.82)						
Ret_{t+1}		15.4**					
		(2.00)					
Ret_{t+1}					0.02		
					(0.97)		
adj.-R^2(%)	17.3	16.1			12.3	8.0	

Table 5: Sensitivity to the choice of search term

SVI_{GT} is the log year-to-year Google Trends search volume index for the respective search term in the US . S&P return is the next period log return on the S&P 500 in percentage points.
heteroscedasticity consistent t-statistics in parentheses
significance levels: * - 0.05, ** - 0.01, *** - 0.001

SVI_{GT}	"economy"	"economic"	"economics"	"fuel economy"
Mean	1.02	0.99	0.91	1.15
Std. dev.	0.29	0.35	0.08	0.59
Panel 1: whole sample (Jan 2005 - Dec 2009)				N = 260
S&P return	-3.40***	-2.35***	-3.11**	-0.67*
	(-3.17)	(-2.96)	(-2.50)	(-1.81)
adj.-R^2(%)	8.62	5.51	2.20	1.19
Panel 2: "pre-crisis" subsample (Jan 2005 - Jun 2007)				N = 128
S&P return	-2.40***	-1.43**	-1.23*	-0.67**
	(-3.64)	(-2.23)	(-1.74)	(-2.17)
adj.-R^2(%)	9.57	5.44	4.08	5.34
Panel 3: "crisis" subsample (Jul 2007 - Dec 2009)				N = 131
S&P return	-3.66**	-2.58**	-12.0**	-0.70
	(-2.80)	(-2.61)	(-2.24)	(-1.28)
adj.-R^2(%)	8.47	5.46	5.80	0.00
Controls	S&P current log return			

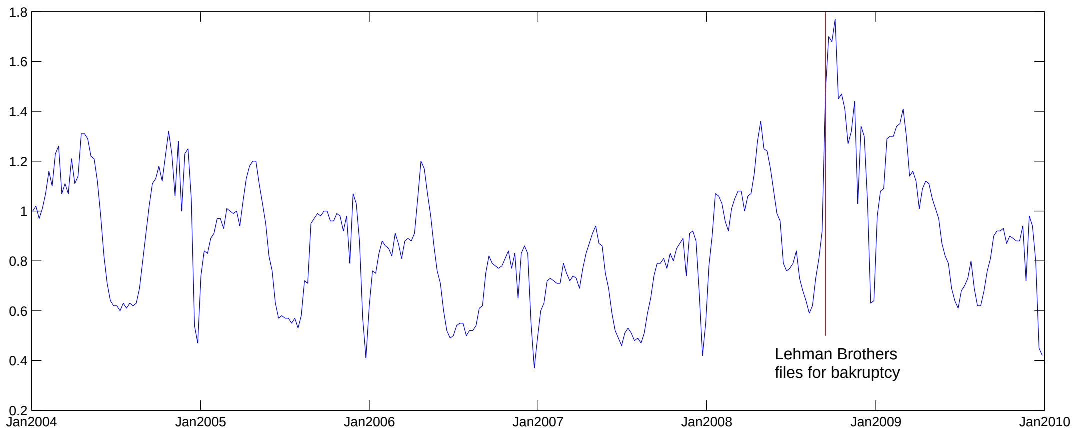
Table 6: Regional comparison

SVI_{GT} is the log year-to-year Google Trends search volume index for "economy" (or its local language translation) in the respective country. S&P return is the next period log return on the S&P 500 in percentage points.
heteroscedasticity consistent t-statistics in parentheses
significance levels: * - 0.05, ** - 0.01, *** - 0.001

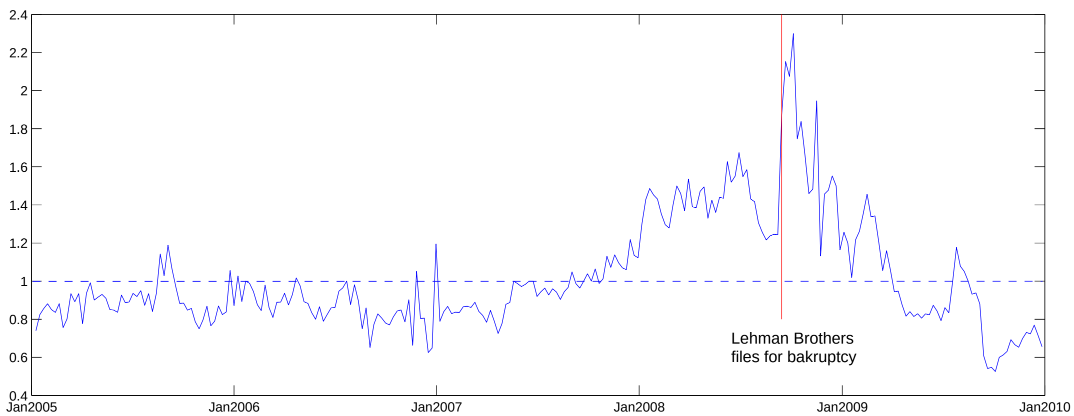
SVI_{GT}	US	UK	Canada	Australia	Germany	Japan
Mean	1.02	0.97	0.99	0.96	0.86	0.97
Std. dev.	0.29	0.14	0.24	0.21	0.12	0.11
Panel 1: whole sample (Jan 2005 - Dec 2009)						N = 260
S&P return	-3.40***	-3.94*	-3.04**	-3.35***	-1.39	-6.16
	(-3.17)	(-1.94)	(-2.13)	(-3.18)	(-0.75)	(-1.59)
adj.-R^2(%)	8.62	4.77	4.64	5.59	0.70	3.32
Panel 2: "pre-crisis" subsample (Jan 2005 - Jun 2007)						N = 128
S&P return	-2.40***	0.1	-0.54	-1.8*	-1.45	1.08
	(-3.64)	(0.06)	(-0.52)	(-1.77)	(-0.87)	(0.39)
adj.-R^2(%)	9.57	1.40	1.93	1.75	2.15	0.0
Panel 3: "crisis" subsample (Jul 2007 - Dec 2009)						N = 131
S&P return	-3.66**	-5.97*	-4.61*	-4.65**	1.96	-9.0
	(-2.80)	(-1.79)	(-1.93)	(-2.71)	(0.63)	(-1.65)
adj.-R^2(%)	8.47	5.8	7.14	6.89	0.0	5.94
Controls	S&P current log return					

Figure 1: Comparing different SVIs.

(a) Search index for "economy", Google Trends, US only, fixed scaling



(b) Search index for "economy", year-to-year (2004 = base), Google Trends, US only, fixed scaling



(c) Search index for "economy", Finance&Insurance >> Investing category, Google Insights, US only, fixed scaling

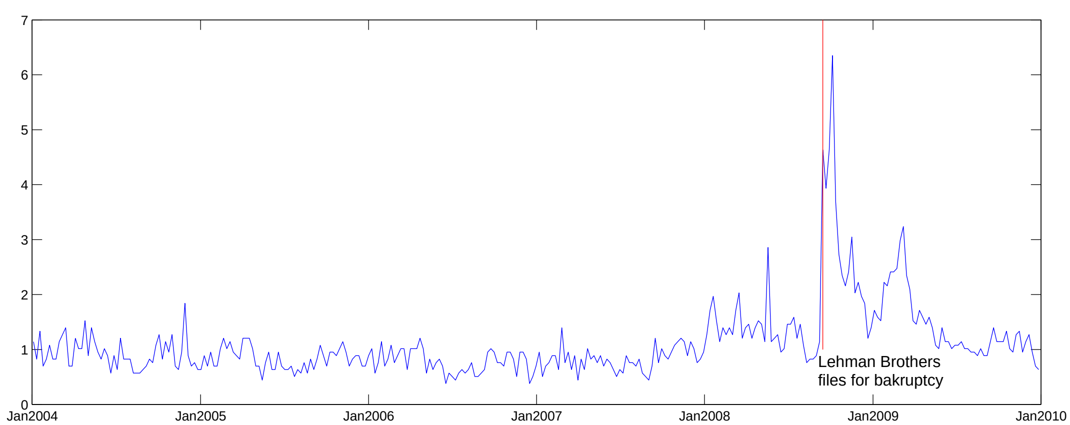


Figure 2: Dynamics of the monthly uncertainty measure.