Norwegian School of Economics
Bergen, Spring 2017

The Wisdom of Crowds: A Case Study in Internet Search Trends

By Emil Gustafsson

Supervisor: Tommy Stamland

"[Master thesis, MSC, Finance]"

NORWEGIAN SCHOOL OF ECONOMICS

Contents

1. Introduction

I use meta-data from google search trends to either explain or better understand real economic phenomena. Primarily I create indicators based on search trends intended to correlate with economic factors. First, I determine unemployment as a strong candidate explained variable, as the link between internet searches and unemployment seems intuitively strong. I develop indicators from related searches that track unemployment levels qualitatively.

Finding some success with unemployment, I use the same approach to explain consumption patterns. While the indicator I create is less accurate than the previous one, it also shows features that indicate a relationship between search patterns and consumption. Next, I look at the relationship between market returns, specifically the S&P 500, and market sentiment derived from search trends. This is a field with considerable interest already. I look at work done previously, and lessen the burden of finding search terms with good fits, by copying the work done by creating the FEARS index (Zhi Da, 2013). I use my own method to create a reverse market sentiment indicator, and observe a leading trend.

Another area of interest is how search trends can reveal underlying predispositions and intentions. Specifically, I want to see if there is a link between the relative emigration intensity of a state and certain search patterns. The way I determine emigration intensity is to use reports by United Van Lines on export shipments from each state (United Van Lines, 2016). I use research done by the US Census to develop an indicator to capture emigration propensity (US Census, 2017). Consistent with this research I find that searches for “new job” and “divorce lawyer” correlate positively with emigration intensity. I also find that searches for “suicide” correlate positively, and that “surgery” correlates negatively.

Finally, I look at a potential link between volatility of search interest for a firm, and the implied probability distribution of future stock prices derived from options on that firm’s stock. I theorize that increased search interest for financially related terms about a firm means that uncertainty about future returns has increased. I find daily historical implied volatility data, and try to square this with an indicator. The results are lacking, as stronger numerical methods and data crunching capacity is needed.

2. Theory

2.1 Previous Applications

2.1.1 Predicting Influenza Outbreaks

One of the first instances where internet search trends were used to help explain or predict real world events was the Google Flu Project (Dugas, et al., 2012). In this project researchers used meta-data regarding search behaviors to predict outbreaks of influenza. They link increased activity for specific search terms to likelyhood of an incrase in visits to the doctor with influenza related symptoms. The model they create is able to predict geographically outbreaks of flu accurately within the certain paramaters. While further scrutiny of the model laid bare shortcomings, especially when dealing with unforseen events such as the H1Z1 flu (Butler, 2013). However the idea of linking search behavior with real life events still seems to have merit.

2.1.2 Predicting Stock Market Returns

The “Elephant in the Room” in terms of application of data potentially specifying the sentiment of people on a macro level is to try to predict stock market returns. An early instance of this is research where they found the search term that most negatively correlates with the weekly return of the Dow Jones (Preis, Moat, & Stanley, 2013). This term turns out to be “debt.” The researchers then built a simple trading strategy around this where when searches for “debt” was increasing, they would sell their portfolio, and when searches were going down they would go long the index. They then backtested this strategy from 2004 (when google trends start), to 2011. The resulting portfolio had a 325 point gain over the period, which is fairly decent for such a simple idea.

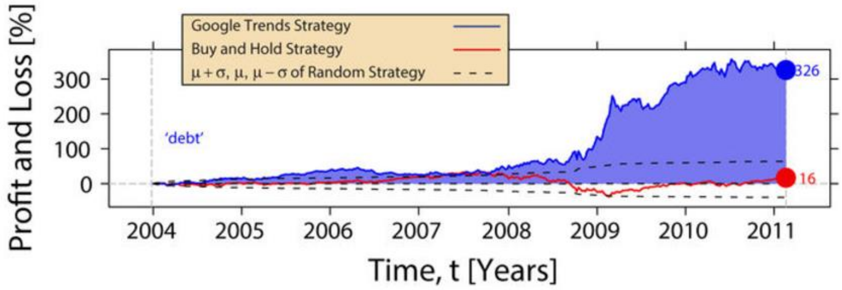


Figure 1: Google Trends Trading Strategy

2.2 How Information is leaked

The primary mode I imagine information can be captured through search trends is by individuals observing the real world around them, and querying the internet for help or information. The search trends can either be leading, coinciding or lagging in terms of the actual event.

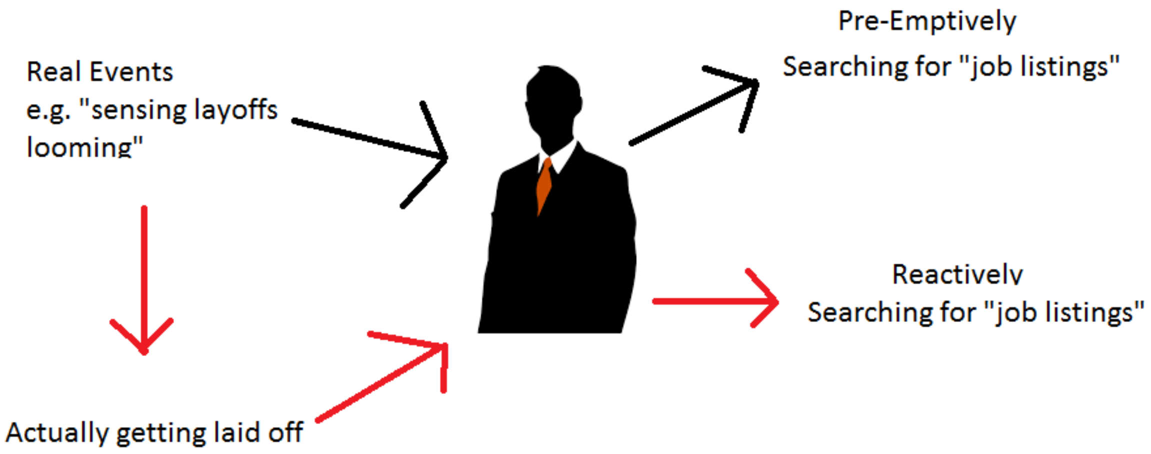


Figure 2: How events may affect search patterns

2.3 Ordinary Least Squares Regression

In order to find the best fit for data, the ordinary least squares method, or OLS, is a common method used (Wooldridge, 2014). I utilize this method in order to find the best weighting on search terms, so that the sum of square residuals between the indicators I create and the factor I am trying to explain is the lowest possible. I also utilize OLS regression to determine which indicators best explain certain phenomena. I also use r-squared in this context to measure the explanatory power of my indicators.

2.4 Kurtosis

An important concept in finance is kurtosis. Kurtosis is a descriptor from statistics that tells us the relative spread of a probability distribution (Hull, 2011). This is important, as an assumption often made in finance is that stock market returns are normally distributed, and models are built with this in mind. So when we are looking forward, and making predictions, we may think of the probability of future events as normally distributed. However, it is key that we understand that not all normal distributions are created equally. Some have a relatively high probability of extreme events “fat tails,” while some are the opposite and feature more of a peak around the centre.

Implied volatility of an option is a measure of the markets belief surrounding future prices of the underlying. High implied volatility means that the market prices in a high likelihood of extreme values of the underlying in the future. Such as an uncertain stock, like a start-up company or a firm that is changing its corporate strategy around. Meanwhile blue-chip firms like Microsoft are unlikely to take on extreme values in the future, and feature commensurately a very low implied volatility in their options. When trying to use google trends to predict future events it can be important to see it in the light of other ways to measure market predictions.

3. Methodology

3.1 Approaching the Google trends data

3.1.1 The data itself

Google has a continuously updating database of search trends worldwide accessible by anyone. They likely limit how easily these data can be extracted, and I am left having to manually download each time series meticulously. The data itself is given in relative search volume, indexed from 0-100. The peak in any given period is given a value of 100, and all other points are given numbers relative to that. The actual volume of searches is not available. However, knowing changes over time is sufficient for the arguments this paper tries to make.

3.1.2 Search Term Selection

Initially I want to find sets of terms that intuitively will all correlate in the same way to the macroeconomic factor I am trying to explain. While I attempt to find an elegant way to gather large quantities of relevant search terms, I ultimately resort to simple economic intuition. While this presents a weakness, it also makes any successful results stronger as a relatively speaking “barebones” approach was able to get results.

3.1.3 Seasonality

A common feature of time series data is seasonality. However, it is important to understand why this seasonality occurs in search data before we start treating it. When individuals make google searches, they are revealing an interest for a topic. The nature of this interest is not directly disclosed, however, and we must infer it through other means. Take for instance searches containing the words “going back to school.” Either searches are likely to mean the end of summer break, or someone might be uncertain of their job security. I utilize a simple method to remove seasonality where each month gets an index calculated by:

$$Index_i = \frac{\frac{\sum_{i=1}^{n} a\ months\ searches}{n}}{average\ of\ all\ months}$$

In the adjusted time series, I divide the search value of each month by the index for that month. Months with values consistently above average are deflated, and months with consistently below average values are inflated (appendix i & ii). The intention is for the adjusted series to reveal more clearly any underlying pattern that the seasonality could obscure. Take for instance searches including the terms “going back to school.”

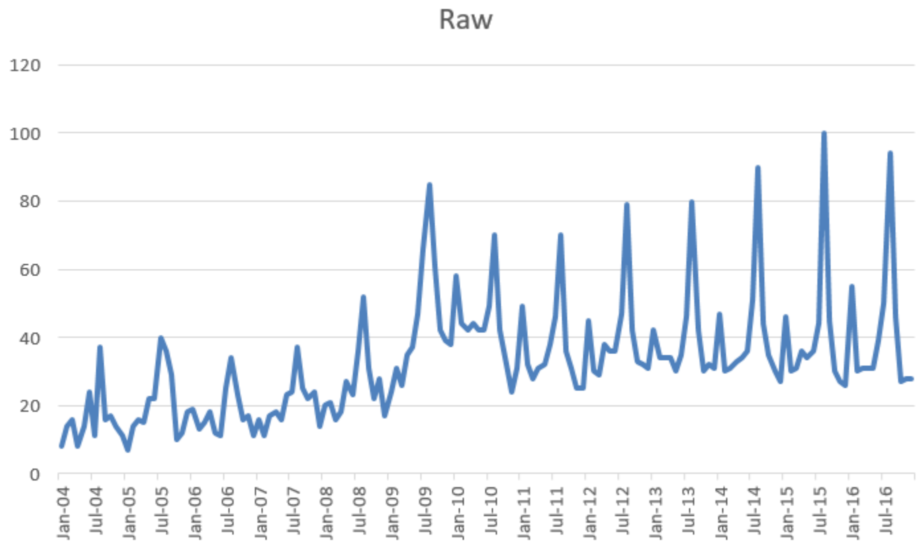


Figure 3: raw "going back to school" searches

This figure clearly illustrates the seasonality of these search terms. While it appears to have increased around the year 2009, the overall trend and peak of the series remains somewhat ambiguous. However when we look at the adjusted time series below.

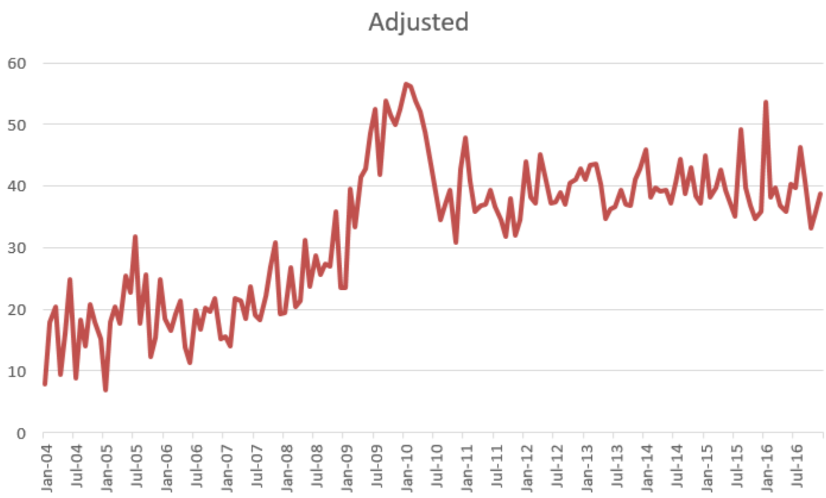


Figure 4: adjusted "going back to school" searches

The data much more clearly displays the peak and continued increase in interest since the recession of 2008 hit.

Another feature of removing seasonality is increasing the visibility of outlier months, where search interest was unduly high. I present one such example below, where the time series' for "NFL" searches are displayed.

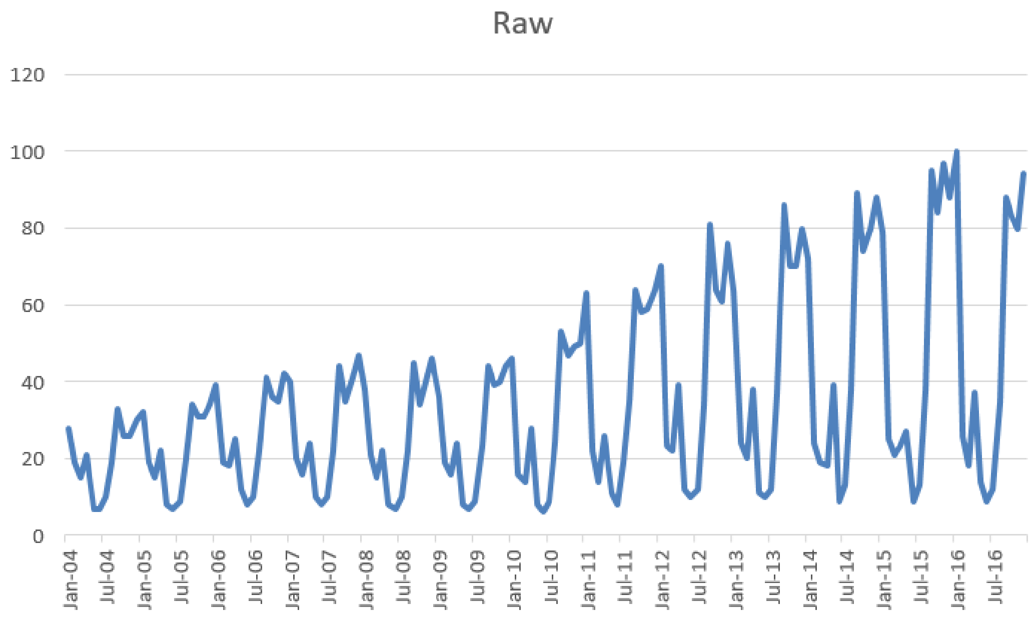


Figure 5: raw "NFL" searches

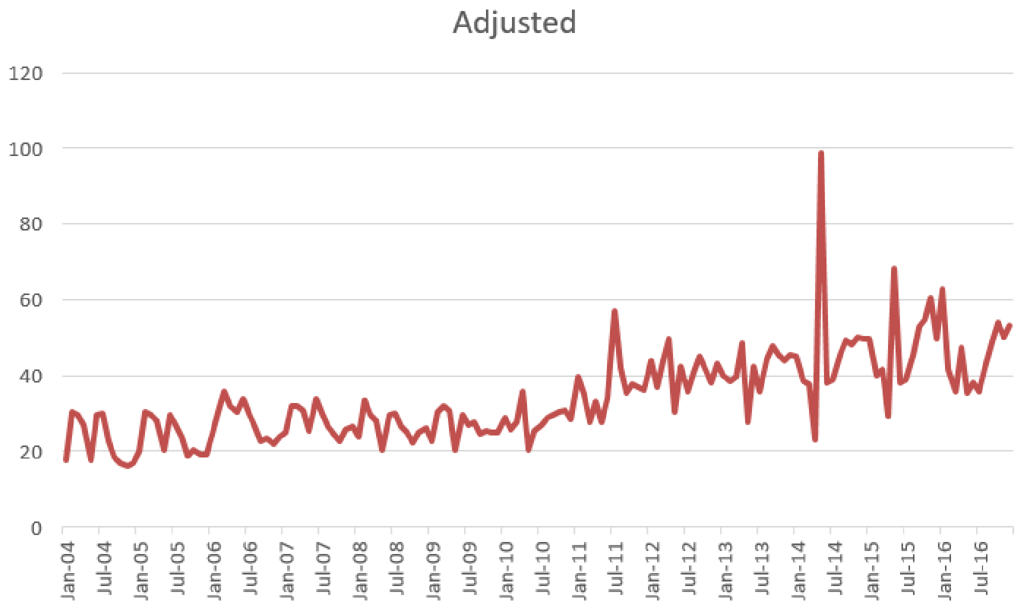


Figure 6: adjusted "NFL" searches

What becomes evident with the adjusted series is a clear spike in the month of May 2014. The reason for this spike is the draft of that year featured the first openly gay football player, Michael Sam (Kang, 2014). The media overexposure of this event turned it in to a "big deal," however it is only in the adjusted series this is evident.

3.1.4 Trend

Many of the time-series of search interest also display a clear upward trend. As internet usage increases over time, so do googles searches. I made several different attempts to correct for this, but was ultimately unsuccessful in finding a comprehensive solution. My first approach was to find the aggregate number of google searches in the period 2004-2016, and then use this to deflate the data to a common standard. However, the issue with this is that it does not account for heterogeneity in the trends across various search terms. For example, the convention of posing search queries as questions, such as "how to write a cv," likely has a higher underlying trend than "job listings" as finding job listing online was commonly known even in 2004. While I attempted to de-trend the data using overall search volume increases, none of these attempts were ultimately fruitful.

3.2 Work With the Search Trends

3.2.1 Creating Macro Indicators

The thesis of this paper is that search interest on the internet leaks non-trivial information that can potentially precede actual events of economic significance. The first way I try to concretize this is by creating indicators. I create these indicators initially by using a simple weighted average of search terms that I deem likely to correlate in some way to the factor I am trying to explain. The most pressing weakness to this approach is the way I select search terms. One method I considered was to select terms by doing a large volume of regressions of individual terms and select based on individual correlation with the thing I am trying to explain, similar to the approach used in developing the FEARS index, an indicator created to gauge market sentiment (Zhi Da, 2013). However, two problems arose when considering this approach. First, finding an exhaustive list of candidate search terms that all related to the factor I was trying to explain proved very difficult. Secondly, it is somewhat beyond the technical scope of the author to parse thousands of search terms directly from google, having to rely on manually downloading the time series. It therefore became necessary to find a different approach.

Ultimately, what proved most successful was to put together the around 20 of the most intuitively relevant search terms in a weighted average, and then performing OLS with the factor I am working with. What constituted good intuition in this case would for instance be saying that it is likely that a person sensing he might get laid off work would search for “how to collect unemployment,” this either signalling or confirming the layoff.

$$Indicator_t = \frac{\sum_{i=1}^{n} Search\ Interest_{i,t}}{n} \qquad \text{Equation (2)}$$

The goal is to find the indicator with the lowest R-squared, which signals a good degree of explanatory power. However, trying to maximize this metric simply through trial-and-error with different search terms proves slow and inefficient. I determine that one way to improve the quality of the indicator was to maximize the R-squared by allowing for different weights on the search terms. I accomplish this in Excel by creating a cell where I write the sum of all squared deviations (Appendix iii), and then using the solver function to minimize the value of that cell by changing the weights given to each search term in the indicator.

$$\min_{w_i} \sum_{t=1}^{T} (predicted\ variable_t - indicator_t * normalizing\ factor)^2 \ s.t. \sum_{i=1}^{n} w_i = 1$$

Equation (3)

This yielded improvements in the indicators fit with the predicted variables. Ultimately, the indicators derived from the aforementioned process are the ones used in the next step.

3.2.2 Working with the Macro Indicators

Now that the indicators are developed, the next step is to test their explanatory, and potentially predictive, power over the explained factors, namely unemployment, consumption, and market return. The method to do this is by simple OLS regression. I use the indicators in various ways to test their aptitude.

3.3 Predicting Inter-State Moving Patterns

The next idea to test the underlying thesis is to see if I am able to create an indicator for each state in the US, which in some way relate to state-to-state migration patterns. I base this on research done to find the root causes of inter-state migration (US Census, 2017). I therefore first create state specific indicators that are based on search terms that I theorize indicate a propensity to move, such as “divorce,” “depression,” and “lawsuit.” The idea is then to find ways to see if there is a correlation between this indicator and moving patterns across states, in particular changes in relative emigration from each state. However, this presents a problem, as the US only release statistics on net migration, and I need to isolate only the state specific emigration. One way around this is to use data from moving companies, as you can use their outbound shipment numbers to have an approximation for relative outflow from a state.

My approach is then to create an instrument describing relative intensity of outflow from a state by dividing outbound shipments from a state on the state’s population.

$$State\ Emigration\ Intensity = \frac{Outbound\ shipments\ from\ state}{Population\ of\ state} \qquad \text{Equation (4)}$$

I then try to find confirmation for the idea that a high level of certain searches in a given state correlates with a high level of emigration intensity. Moreover, I also try to see if the index has more explanatory power over changes in emigration intensity, rather than the intensity itself.

3.4 News and Implied Volatility

The next idea I try to substantiate is whether there is a link between spikes in search interest for a specific firms financially related factors and the implied volatility of the firm’s options. What I imagine to be the channel for this is that as investors and the public become aware of potentially critical information regarding a firms future prospects, search interest for terms important to a firms valuation spike. Examples of these kinds of searches are “buy Netflix,” “Netflix undervalued,” “Netflix merger,” “Netflix buyout,” and “Reed Hastings health.”

The primary issue with resolving this question is the difficulty in finding appropriate volatility data that fits the necessary characteristics of coming from options with the same duration and delta even as time progresses. It turns out to be quite a demanding task, ending up with me reading off the daily volatility from a graph. This naturally constrained my research in this area to a shorter period of only six months, only using call-options, and only looking at one firm.

A very important aspect to consider in this case is how the search terms are loaded, in the sense are they representing news likely to increase the probability of upward movement in the future, or the opposite. I circumvent this issue by allowing for negative weights on the terms in the optimization of the indicator.

4. The Results

4.1 Job Security Indicator

4.1.1 Background

I theorize the macro-economic variable with the highest chance of correlating with search trends is unemployment. This is because I suspect there exists a strong relationship between worrying about losing your job, or actually losing it, and information seeking online to fix the problem. I will post the actual unemployment rate in the US over the period spanned by this research, which will serve as a visual aid and point of comparison for results that follow (United States Department of Labor, 2017).

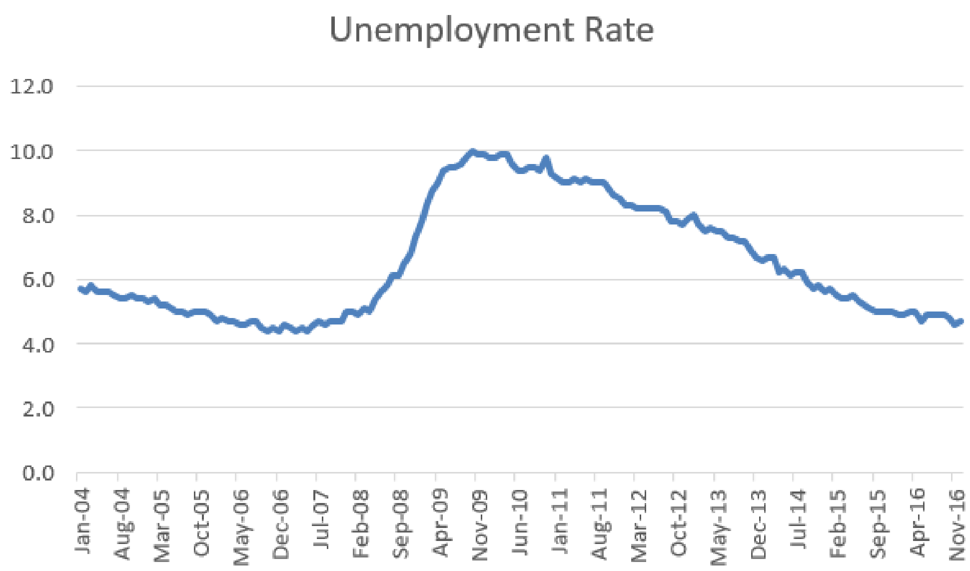


Figure 7: Unemployment rate in the US (%)

4.1.2 Creating a Raw Indicator

Initially I choose search terms thought to correlate with employment. These include, but are not limited to, terms related to seeking re-employment and looking up information regarding government benefits (appendix iv). I then develop an indicator with equal weights for each of the candidate search terms, as described in equation (2). The resulting time series, pictured on the next page, shows some of the tendencies present in the unemployment rate. Most notably the peak of the series falls much later than in the actual unemployment.

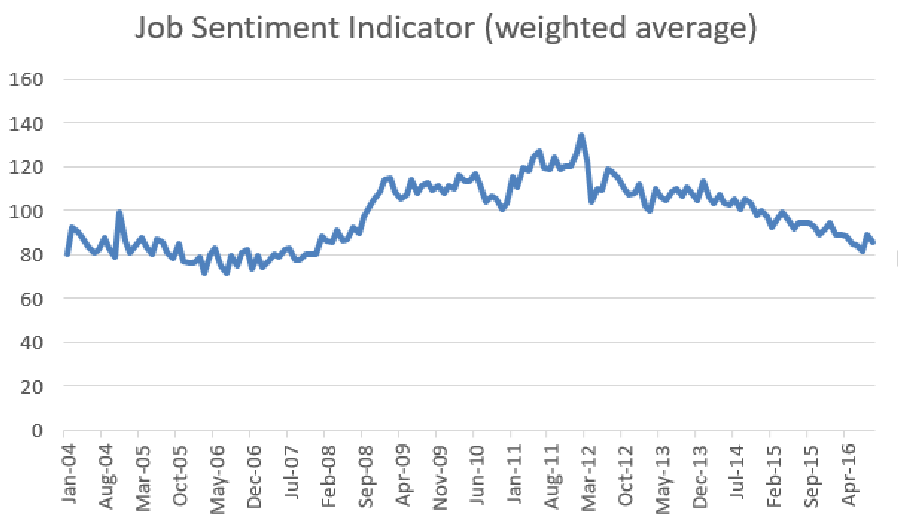


Figure 8: Raw job sentiment indicator

4.1.3 Refining the Indicator

Next on the agenda is to use numerical methods to find a better fit with the data. I now allow different weighting on the search terms, including negative, to see if there exists a better connection between search data and unemployment. Allowing for negative weights is crucial, as some of the search terms likely correlate negatively with unemployment. I now deploy equation (3) by way of Excel's Solver™ function. This results in a new weighting of the search terms composing the indicator (appendix iv). Assessing whether it makes intuitive sense how all the search terms contribute to the indicator is open to scrutiny. However, it seems reasonable that terms such as "career" and "interview questions" contribute negatively to the unemployment indicator, and terms like "unemployment" and "how to find a job" contribute positively. The time series now reads like this:

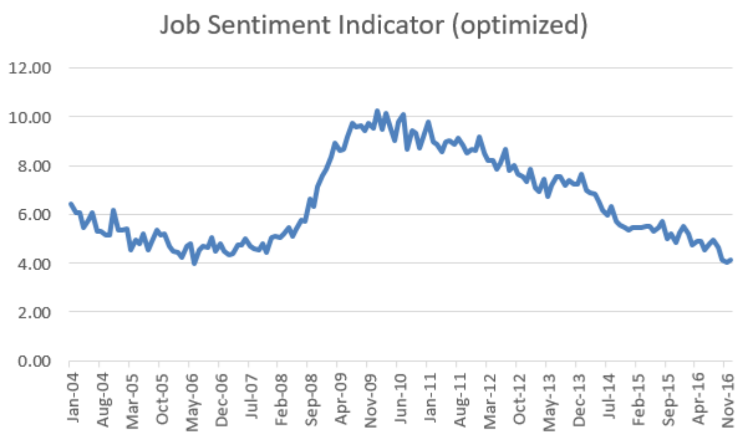


Figure 9: Optimized job sentiment indicator

Compared to actual unemployment this new optimized indicator seems to reveal there could potentially exist a very strong link between patterns in search metadata and real world events, considering the use of relatively few search terms. However, while I submit that finding this link to a certain degree strengthen the underlying argument that this thesis tries to make, using a similar methodology to also "predict" future events is needed. I use the word "predict" in the sense that I want to uncover a proper weighting of the search terms using training data, which I define as the period january 2004 to june 2007. I then use the weights developed by the training data with search data as time progresses. I accomplish this by using equation (3) again, however this time I only sum the deviations from years in the "training data." The weighting is then used on the search data, and produces this time series:

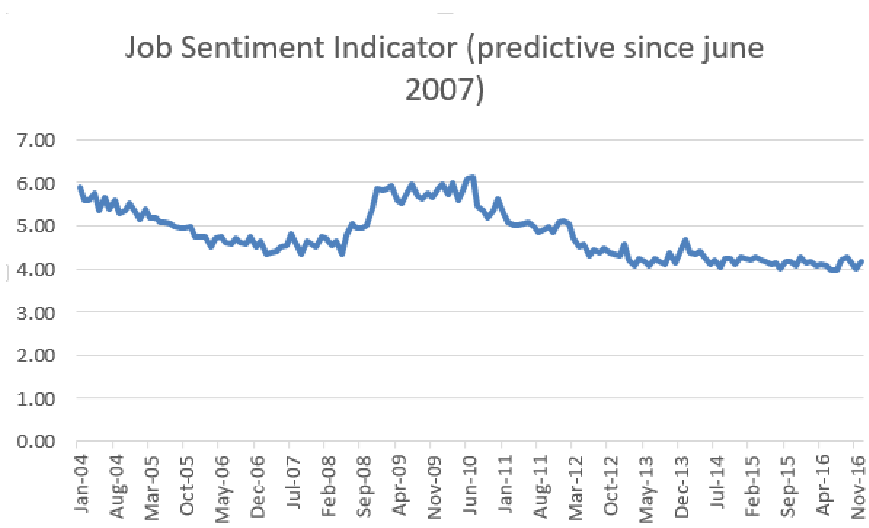


Figure 10: Predictively weighted search terms job sentiment indicator

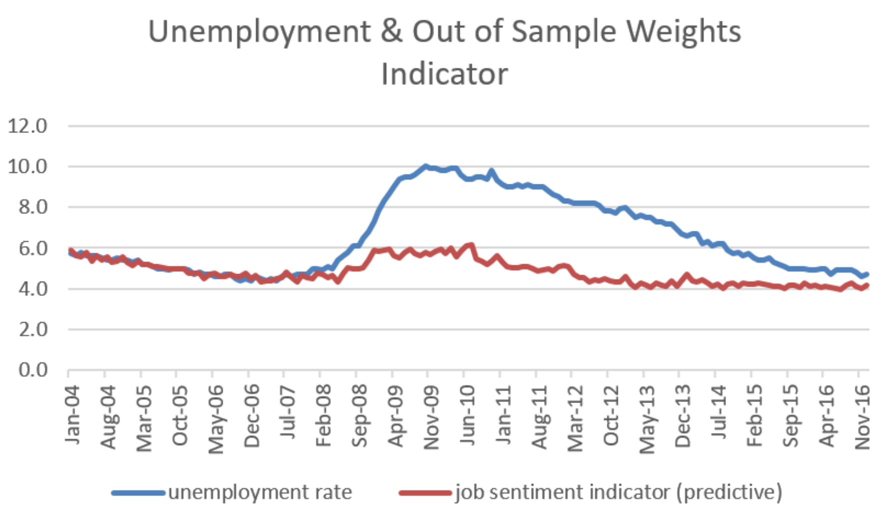


Figure 11: Unemployment and predictively weighted indicator

The key aspects of the indicator time series are retained, even with this limiting factor. Crucially that spike in unemployment caused by the recession, and subesquent tapering off, is covered here. However, the amplitude of the unemployment is less well described. An important note is that data from beyond june 2007 is used in removing seasonality from the search trends. However, I consider this not to be of serious enough consequence to ruin the logic of the exercise. The reason this paper will not cover predictive weights with predictive seasonality adjustment is a matter of the limited improvement that would cause, compared to the sizeable effort it would require. Crucially, however, this does provide evidence that the relationship between the search term and the macro economic variable remains over time.

4.1.4 Concluding Remarks on Unemployment Indicator

I present the summary statistics for the three indicators used as the explanatory variable for unemployment in the following table.

Table 1: Unemployment Indicator Regression

Regression Results Unemployment	Intercept	Coef. on Indicator	Std. Err.	t-value	R2
Simple Average	-3.77552	0.1062486	0.0051439	20.66	0.7386
Optimized	0.0188484	0.9978212	0.0159912	62.40	0.9627
Predictive	-0.9858045	1.570951	0.2174809	7.22	0.2568

It is cause for some concern that the significance level, and r-squared, of the much less well-designed simple average indicator outperform the predictive one. However, I stand by my assessment of the predictive variable as a success, as it clearly tracks the same overall pattern that unemployment actually did. The main reason the predictive weights yielded less “explanatory power” in terms of r-squared was the amplitude of the unemployment spike was quantitatively larger than the indicator, but not qualitatively in the wrong direction. My final thoughts are still that my work is merely a proof of concept, of the idea that search trend metadata contain information about peoples employment prospects and job security. What is needed to make this data useful in a consequential way is more sophisticated numerical methods. By employing machine-learning techniques to handle much larger quantities of data, and allowing for continuous updating of the weighting of search terms, it is certain patterns will emerge more clearly. The focus in the future should be to develop algorithms that predict month-to-month changes in unemployment. Another source of information that can give indications about future short-term levels of unemployment could have significant implications across many fields.

4.2 Consumption Indicator

4.2.1 Background

There existing a link between people's consumption patterns, and the searches they are likely to put into google seems intuitive. With the advent of the internet consumers look for information about purchases, as well as ordering goods and services online. I therefore consider this a strong candidate macro variable for this project. Again, I put as a reference point data on consumption by American consumers at the start of this chapter (Federal Reserve Bank of St. Louis, 2017).

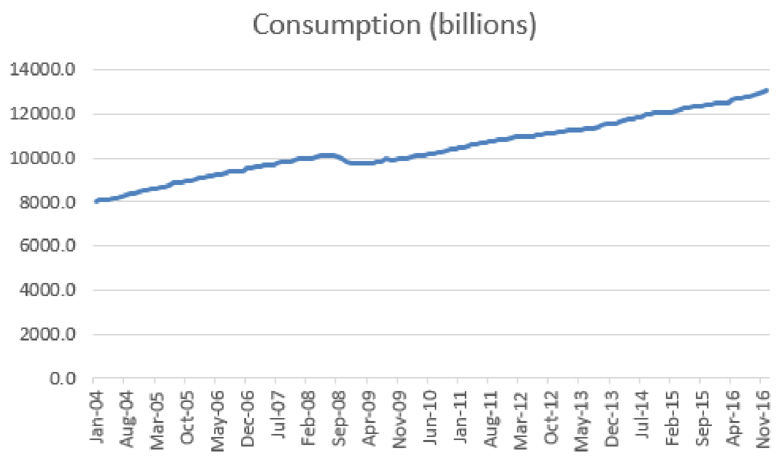


Figure 12: Consumption in the US in billions

4.2.2 Creation of the Indicator

Again, I use economic intuition to determine search terms that I hypothesize correlate with consumption patterns. I focus on searches where individuals might be looking for deals, attempting to free up cash, or are looking for an excuse to make a frivolous purchase (appendix v). Originally, I want to create an indicator where all the terms correlate in the same direction with the underlying pattern. However, this becomes difficult, and allowing for negative weights when optimizing the indicator becomes necessary.

4.2.3 The Raw Weighted Average Indicator

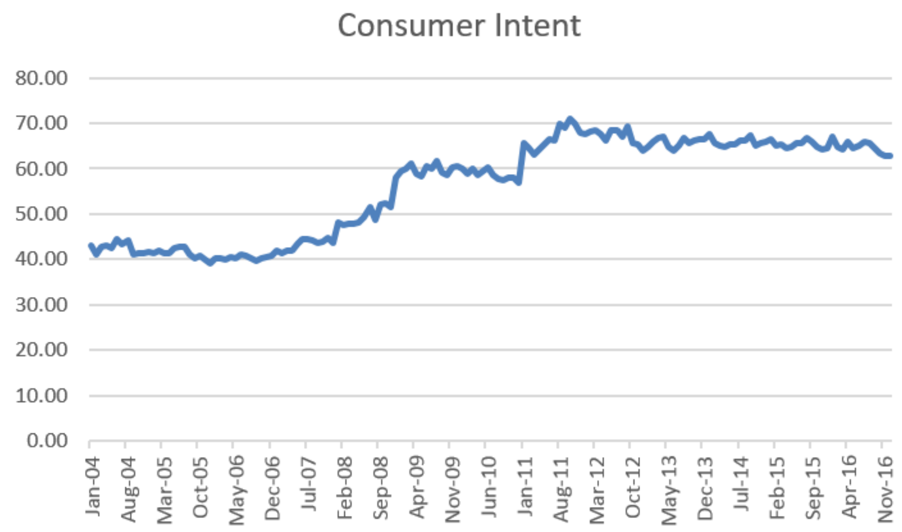


Figure 13: Unmodified average consumption indicator

Initially the indicator shows promise in that it picks up a disturbance in consumption patterns around the 2008 recession. However, it seems to remain flat outside of this period of distress, which distinguishes the raw consumption indicator from the unemployment indicator. I now use equation (3) to optimize the weights on each search term to find the best fit between the indicator and consumption.

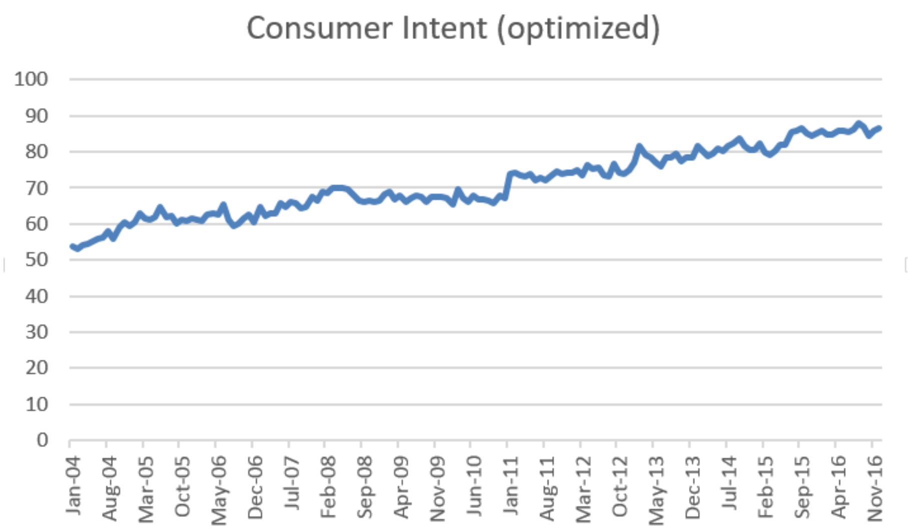


Figure 14: Consumer intent with optimized weighting on terms

The weights that lead to this time series can be seen in appendix (v). The optimized indicator does not track consumption as clearly as in the case with unemployment. But it does provide a reasonable link between search patterns and consumption. What will again be of key interest is to optimize the weights of the indicator using training data. I again define training data as everything up to, and including, june 2007. This allows us to see how well the indicator picks up the drop in consumption during this period.

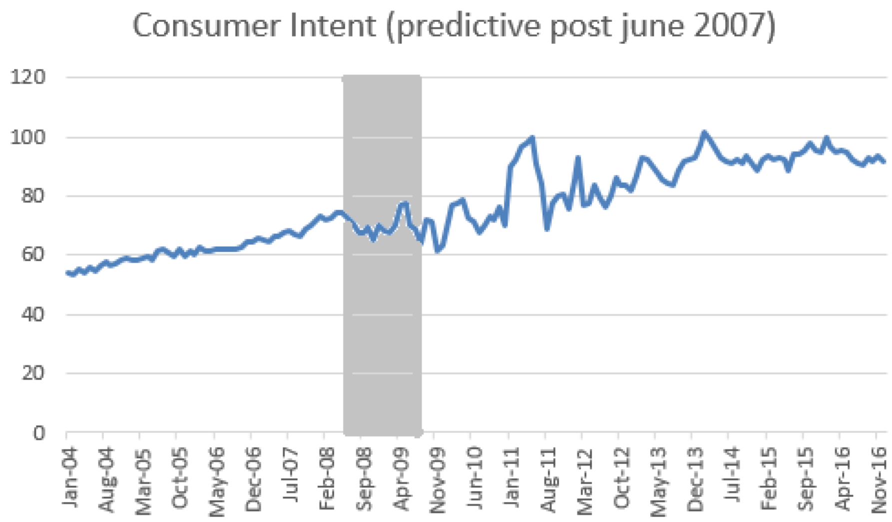


Figure 15: Consumer intent with term weights from training data

This particular weighting leads to a much more volatile time series outside of the training data. It becomes clear how a stronger numerical approach is needed. However, from a less strict perspective, this indicator contains some of the important aspects I am looking for. The recession period Q3-2008 to Q2-2009 (highlighted) is trending toward flat for the indicator, and then upwards since then. While the quantity of search terms, and quality of the numerical approach could be improved, I still consider this an indication in the direction of the hypothesis of the paper, that useful information leaks through search metadata.

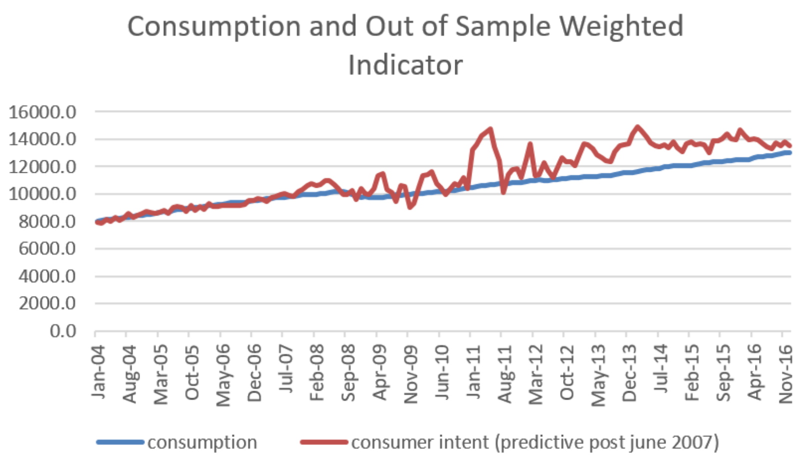


Figure 16: Predictively weighted indicator and actual consumption

Looking at the predictive weigthed indicator superimposed on consumption also reveals the trends of the two series seem very similar.

4.2.4 Regression Data

Table 2: Regression results consumption indicator

Regression Results	Intercept	Coef. on Indicator	Std. Err.	t-value	R2
Simple Average	4819.623	100.4184	5.578535	18.00	0.6778
Optimized	290.9493	143.3024	2.408948	59.49	0.9583
Predictive	3735.798	87.7752	2.973656	29.52	0.8498

This time the predictive indicator seems to fare better than the in the case of unemployment. I attribute this to the failure of the unemployment indicator to properly account for the amplitude of the unemployment increase. The consumption indicator performed much more erratically. Relying on one metric to determine quality is insufficient, as it is clear consumption is a more problemtatic relationship for me to pin down.

4.3 Market Return Indicator

4.3.1 Background

The next relationship I try to determine is between aggregate economic sentiment leaked through search terms, and stock market returns. There already exists work on this subject matter; specifically I find the development of the FEARS index by Da, Engelberg, and Gao interesting (Zhi Da, 2013). The FEARS index is a construction of negatively loaded economic terms, as decided by the Harvard IV-4 Dictionary, which are chosen by significance degree in terms of t-value. They find that negative terms are most useful in identifying sentiment.

For the scope of this paper, I will construct my own reverse market sentiment indicator based on the search terms used in the FEARS index, and make a determination as to whether there is inherent leading, lagging or coinciding bias in the indicator I create (appendix vi).

4.3.2 Creating a Reverse Market Sentiment Indicator

Because work with market sentiment has incredible value if it can be shown to correctly predict future returns, I start by working towards maximizing my indicator with this in mind.

First, I try creating an IF function in Excel to give a numerical value of “1” to a correct prediction by the indicator, and then set the weights in the indicator by maximizing the number of correct predictions. However, I was not able to do this, as Solver™ is incompatible with non-continuous functions. This limitation greatly hinders my work in this area.

My second attempt to optimize the predictive power of my indicator is to create a value for each month where if the indicator and the S&P 500 moved in the opposite direction you would get a negative value, and if they moved in the same direction, you would get a positive value. The goal is then to minimize the value of this as the indicator tracks negative sentiment by construction.

$$\min_{w_i} \sum_{t=0}^{n} (indicator_{t+1} - indicator_t) * (s\&p\ 500_{t+1} - s\&p\ 500_t)$$

However, this approach sadly did not yield any useful results. Most likely, this is because Excel will not value marginally correct predictions with this optimization, and only focus on maximizing the value of already big correct predictions. Therefore, I must move to yet another attempt.

My third foray into this problem is less elegant, but retains the key characteristics of this paper’s main argument, that information leaks through search habits. I now use a simple weighted average of the search terms in the FEARS index to create the indicator for market sentiment. Next, I inflate the value of the indicator by multiplying it by 45 for graphing purposes. Finally, I “invert” the time series by subtracting this value from 4500, a value that I arbitrarily decide to make the series end at approximately the same value as the S&P 500, defining it as:

$$indicator_t = 4500 - \left(\frac{\sum searches_t}{30}\right) * 45$$

By graphing the indicator along with the S&P 500, we can more clearly see if this method appears to have any merit.

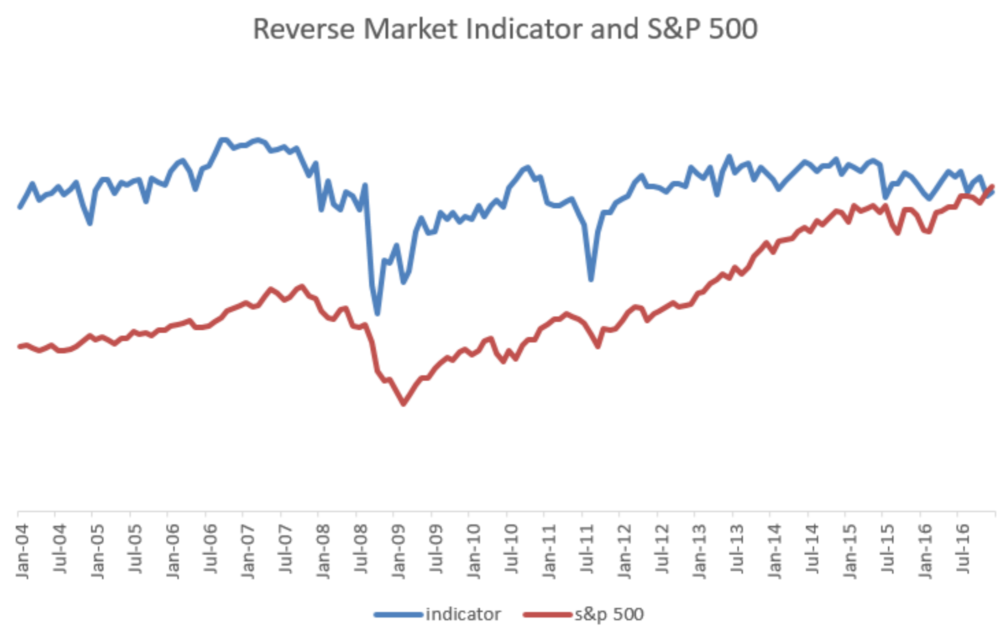


Figure 17: Reverse market indicator and S&P 500

The more straightforward method of averaging search volume of negative sentiment, and giving the two time series the most intuitive economic fit, features little room for forcing the data to fit my conclusion. I consider this successful at illustrating the case for a relationship between search habits and real economic factors.

4.3.3 Testing for Temporal Fit

Next, I use linear regression between the two series, in a bid to determine whether the indicator is inherently predictive in nature. The reason I do not look more deeply into the day-to-day, week-to-week returns is a consequence of limited ability of the author to parse the bigger quantity of data that would require. However, I still consider comparing the econometric fit of my relatively simple indicator to be of value.

Table 3: Regression Results Market Sentiment

Regression Results Market Sentiment					
	Intercept	**Coef. on Indicator**	**Std. Err.**	**t-value**	**R2**
sentiment lagging 1 month	-399.23	0.8282	0.1309	6.33	0.2073
sentiment coinciding	-573.52	0.9091	0.1292	7.04	0.2432
sentiment leading 1 month	-704.4	0.9667	0.1244	7.77	0.2845
sentiment leading 2 months	-708.59	0.9692	0.1243	7.80	0.2869

What the results seem to indicate is that the sentiment indicator fits the S&P 500 index better when it is a leading indicator. Both the t-value and the r-squared improve as I regress the sentiment with a lead on the S&P 500. While the need for stronger numerical methods become ever more pressing, I take solace in still finding results that imply a relationship between searches and real trends.

4.4 Inter-state Moving Patterns

4.4.1 Background

As internet search meta-data can seemingly capture the sentiment of people by indirectly capturing their current preoccupations, the questions regarding what use this information has continues to intrigue me. I want to see if there is a link between the search derived sentiment is specific geographic areas, and relative levels of emigration.

4.4.2 Linking Emigration with Search Data

First, I find a metric to describe how emigration intensity develops over time. I accomplish this by first finding the population level of all 50 states, plus The District of Columbia (appendix vii). The US Census Bureau reports these numbers easily accessible online (US Census, 2017). However, as it pertains to migration across states they only report net numbers. For my research interest, it is important to filter out only the emigration experienced by a state. The way I accomplish this is by using data reported by the United Van Lines Movers Study. While the full dataset is not reported in full, I am able to get consistent data from 2006 to 2011 (United Van Lines, 2016).

The data I have been able to procure details the number of outbound package shipments per state (appendix viii). If we first assume even representation by United Van Lines across America. Furthermore, if we also assume that no activity variation in any state happens for reasons other than aggregate changes in emigration in that state, we can use this data to describe emigration intensity. I accomplish this by first adding the population data from each state from the year 2006 to 2011 into excel (appendix vii). I then make a similar sheet with the outbound shipments from each state (appendix viii). Finally, I divide outbound shipments by population over all the years covered, as described in Equation (4). In addition, I multiply this number by a common factor to inflate it to a number I can more readily interpret (appendix ix).

This quickly becomes an overwhelming amount of information and the need to pare back on the sheer numerical workload becomes unavoidable. I decide to look for states with distinctly different moving patterns, to see if I can use those to find consistent trends in search data. If searches are stable in a state with stable emigration, and volatile in a state with volatile emigration, an argument can be made in favour of a link between the two. I take the two states with the most volatile emigration intensity, and the two least volatile. These are Arizona and North Dakota (most volatile), and Iowa and West Virginia (least volatile).

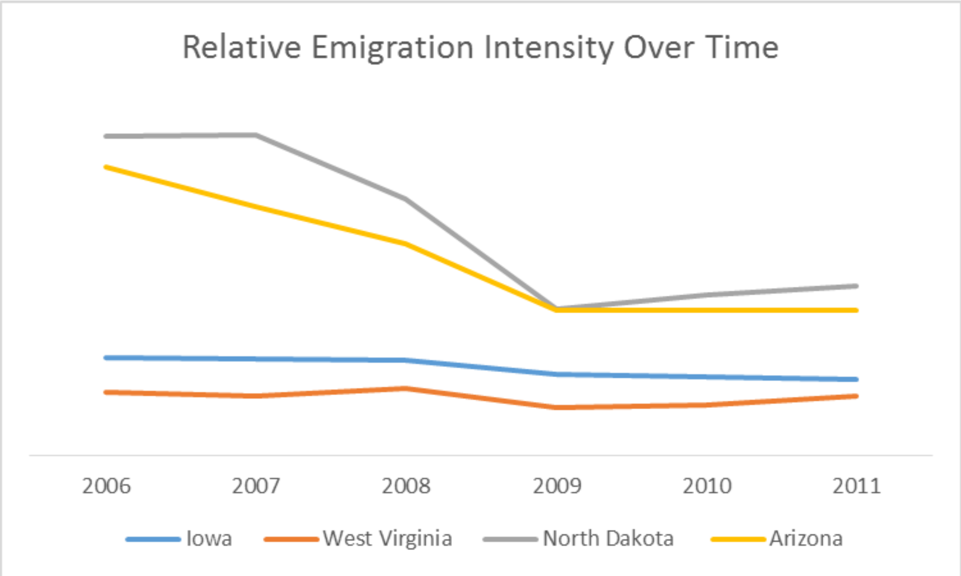


Figure 18: Emigration intensity

No state experienced a rise in relative emigration intensity in the period I cover.

4.4.3 Explaining Inter State Emigration

Next, I look at research regarding why people move, and specifically why Americans move across states. The US Census in 2014 produced an article specifying reasons people gave for moving across state lines (Ihrke, 2014). The primary findings are that within state moves are usually precipitated by housing reasons, while across state moves are job related. Other reasons listed are changes in marital status, retirement, wanting neighbourhood with less crime, natural disaster, change of climate, health reasons, or to attend college.

I again create indicators as described in equation (2) for what I suspect may be causes related to moving. I specifically leave out searches that specifically indicate a move may be imminent, such as “rent moving van” or “hire moving company.” I do this because I am more interested in what may be underlying reasons why people make big moves, rather than try to make an accurate predictive indicator for it. I put the yearly search volumes for terms I think could be related, and arrange them in Excel (appendix x). I dub this the “Moving Inter State Explained by Relative Rates of Yearning Index,” or “MISERY Index” for short.

With the data on board, my first approach is to try using the same weighting on each search term for the indicator to all four states while minimizing the squared residuals of the regression

$$Emigration\ Intensity_i = MISERY\ Index_i + u_i$$

However, the results from this are unsatisfactory. Most likely the limitation of relatively few search terms, with 23, and having to limit myself to 6 years due to data constraints, make it very difficult to find a sensible weighting that uniformly apply to all four states. I should also add that ideas such as using the last 6 months of the previous year and first 6 months of the current year to explain moves were not tested, but might produce better results, as logically not all search increases and moves would coincide in the same year. Also having some of terms lagging and some coinciding is also not tested, but is an idea with merit.

The next best idea I have that is feasible is to take optimize the weights of the search terms for all four states individually using equation (3) (appendix xi & xii). The product of this can be seen in appendix (x). I now simply scan to see which terms contribute in the same direction with the indicator in all four states (either positively or negatively). It makes sense that terms consistently pushing in phase across various different states are likely correlated with across state moves. Four terms among my candidate search terms all had the same sign in the optimized indicator, and these were:

Table 4: Consistently signed terms

Term	Sign of Contribution
“divorce lawyer”	+
“new job”	+
“suicide”	+
“surgery”	-

It is consistent with the reports from the Census Bureau that changes in marital status and career changes precipitate moving across state lines. Furthermore, I think it is of interest that “suicide,” which could be related to either experiencing a loved one taking their life, or contemplating it yourself, seems to maybe have a correlation with big moves. What I think this might be related to is a desire to “start over” when life is at its most depressing, and get a fresh start. While I am unable to find academic studies to support this claim, resources that are more informal echo this sentiment (Lord, 2015). Finally, searches for “surgery,” which probably means a person is having surgery either themselves, or someone they care about. It seems a reasonable link that this would cause someone to be less likely to move, as one is rendered immobile and financially strained by surgery, and is more likely to gravitate toward family.

4.4.4 Concluding on Emigration

While I am overall heavily constrained by the sheer amount of data when working with all 50 states over a long period of time, some intuitive results could be found. I think relating searches to how people behave, and perhaps why, is a meaningful question to ask. Many of our search queries happen for reasons we may not always be conciously aware of, and doing a meta-analysis where these trends can reveal patterns in real life is of considerable interest.

4.5 Implied Volatility, Kurtosis, and Firm Specific Search Interest

4.5.1 Explaining Implied Probability Distributions Using Search Volatility

The final topic I cover in this paper is the relation between implied volatility of a firms stock, and volatility of search interest for that company. Implied volatility, as derived from options on a firms stock, says something about market beliefs about the stock. Specifically, it is the best representation of market predictions regarding the probability distribution of future stock prices. As uncertainty regarding future prices increases, so does implied volatility and the price of options (Hull, 2011). Whereas high implied volatility implies a platykurtic distribution with high probability of tail events, low implied volatility implies a leptokurtic distribution with low probability of tail events.

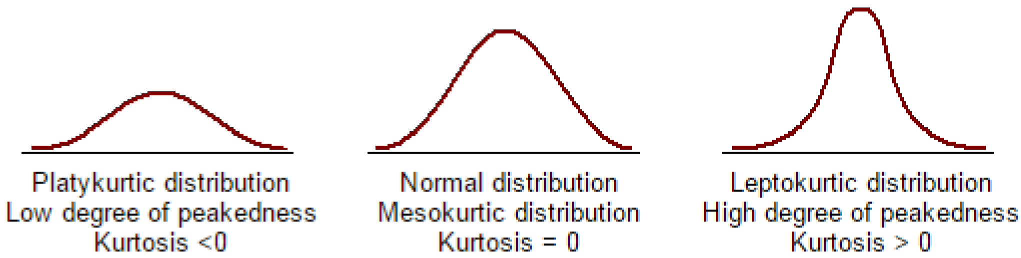


Figure 19: Kurtosis and Probability Distributions

4.5.2 Data and Indicator

I want to test the hypothesis that there is a link between search interest for a firm, specifically financially related searches, and implied volatility. I accomplish this by again first creating an indicator for the search terms I deem likely to in some way correlate to implied volatility (appendix xiii). Next, I have to find historical data on implied volatility. I find data on Yahoo in the period 2014 to 2017 (Quandl, 2017). I limit myself to call options for the first 6 months of 2015. I also do not consider skewness, and assume symmetry in the distribution for this exercise. The gaps in the data correspond to holidays and weekends without any trading happening.

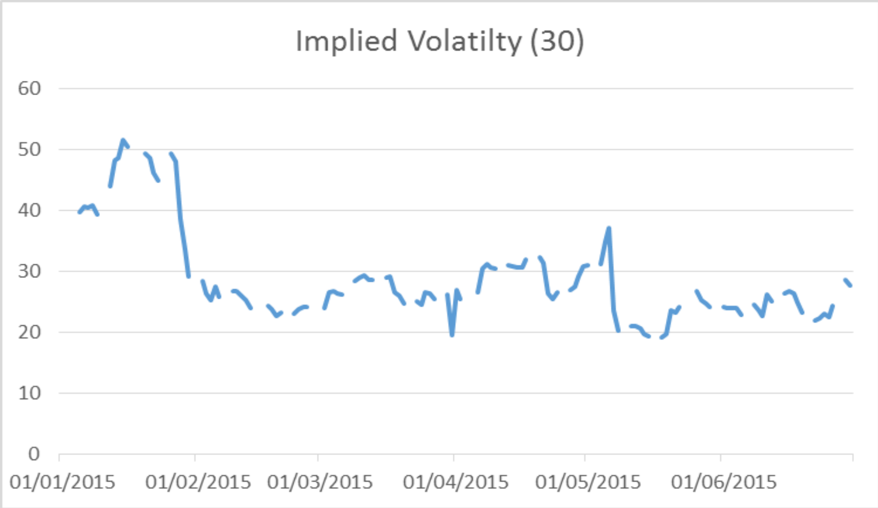


Figure 20: Implied volatility Yahoo (%)

Next, I optimize the indicator to fit the time series of the implied volatility with equation (3).

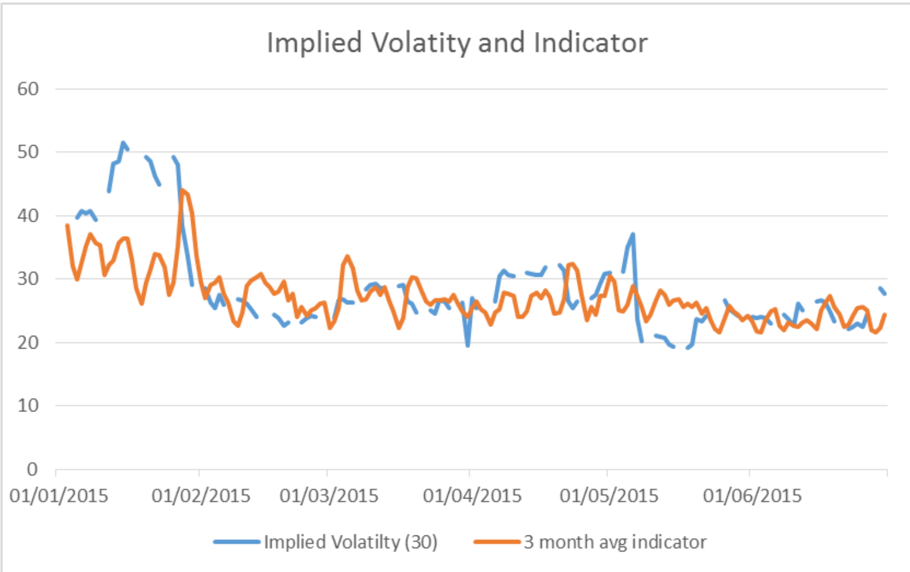


Figure 21: Implied volatility and 3 month avg indicator

4.5.3 Predicting Implied Volatility

What is of main interest to me here is not necessarily finding the best fit, rather trying to see what kind of predictive power a good indicator might provide. The nature of the options market for at-the-money options for a large corporation like Yahoo is likely to be liquid, and very responsive to changes in market information. Therefore, in an ideal world I would have continuous data during the trading day, which would correspond to continuous data from google trends.

Meanwhile, in reality, I do the best with what is available and define predictive success as the indicator and implied volatility moving in the same direction each trading day. I make this calculation in excel with an “if” formula that yields a “1” if they move with the same sign, and “0” if they move with opposite signs (appendix xiv & xv). I also do not concern myself with optimizing for “predictive weights” in the indicator, as my method is lacking enough as it is. The resulting ratio of correct to total predictions is 68/122, or around 55.7%.

The method I employed here is not by any means perfect. Even with using backwards looking weighting in the indicator I was only able to produce an indicator that barely predicts the correct movement of the implied volatility half of the time. However, I still consider the idea something that may have merit. While I am currently unable to do the extensive numerical work needed to properly test the notion, I think it would be a worthwhile effort in the future.

5. Conclusion

The main argument I try to make is that macro-level search behaviour can potentially reveal more about real events than is currently understood. The basic computational methods I employ, combined with simple economic intuition, is able to produce patterns between search data and real events. By creating indicators intended to track, or predict, real world economic patterns there exists opportunity to create better understanding of the factors, but also connections previously unknown.

The indicators for unemployment, consumption, and market returns show real promise as candidate factors that could be well explained, and potentially predicted, by search trends. The features most lacking in this paper is a comprehensive list of potentially relevant search terms, and allowing search terms to interact with different time lags. For future work, I also imagine looking at terms that on their own may not be related to a factor, but in combination could be of significance, interaction terms. For instance, an increase in searches for airplane tickets, without an increase in other pre-vacation related searches, could signal economic strength as presumably more people can suddenly afford long-distance travel.

Looking at statewide emigration intensity provides a slightly different approach to how the search data is used. Instead of trying to determine the best possible prediction of a factor, I try to find consistency in how the search terms correlate with emigration. It can be of interest to see how search terms, and the collective consciousness, is affected by events. On the other hand, it is also a possibility that certain moods in a population, inferable by search trends, can predict phenomena such as emigration intensity.

Trying to capture the markets opinion on the probability distribution of future stock prices is a more complex task. The “loading” of many search terms is either neutral ambiguous. Therefore, it made sense to not predict the stock price itself, but rather the uncertainty. My numerical methods, and ability to parse data, prove woefully inadequate for this task. However, I remain optimistic that with the proper methodology this idea could prove itself worthy.

6. References

Butler, D. (2013, February 14). When Google got Flu Wrong. *London Nature,* s. 155.

Dugas, A. F., Hsieh, Y.-H., Levin, S. R., Pines, J. M., Mareiniss, D. P., Mohareb, A., . . . Rothman, R. E. (2012). *Google Flu Trends: Correlation With Emergency Department Influenza Rates and Crowding Metrics.* Oxford Academic.

Federal Reserve Bank of St. Louis. (2017, Januar 18). *FRED Economic Data*. Hentet fra Personal Consumption Expenditures: https://fred.stlouisfed.org/series/PCE

Hull, J. C. (2011). *Options, Futures, and Other Derivatives.* Prentice Hall.

Ihrke, D. (2014). *Reasons for Moving: 2012 to 2013.* United States Census Bureau.

Kang, J. C. (2014, September 3). *The New Yorker*. Hentet fra http://www.newyorker.com/news/sporting-scene/cowboys-sign-michael-sam

Lord, E. (2015, March 11). *Bustle*. Hentet fra Bustle.com: https://www.bustle.com/articles/64591-13-signs-its-time-for-you-to-move-away-to-a-new-city-because-there-are

Preis, T., Moat, H. S., & Stanley, H. E. (2013, April 25). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Nature*.

Quandl. (2017, May 18). *Quandl*. Hentet fra Alternative Data: https://www.quandl.com/

United States Department of Labor. (2017, Januar 18). *Bureau of Labor Statistics*. Hentet fra Labor Force Statistics: https://data.bls.gov/timeseries/LNS14000000

United Van Lines. (2016, October 2016). *United Van Lines*. Hentet fra Movers Study: https://www.unitedvanlines.com/movers-study

US Census. (2017, Februar 15). *United States Census Bureau*. Hentet fra Population and Housing Estimates: https://www.census.gov/programs-surveys/popest.html

Wooldridge, J. M. (2014). *Introductory Econometrics: A Modern Approach.* CLI.

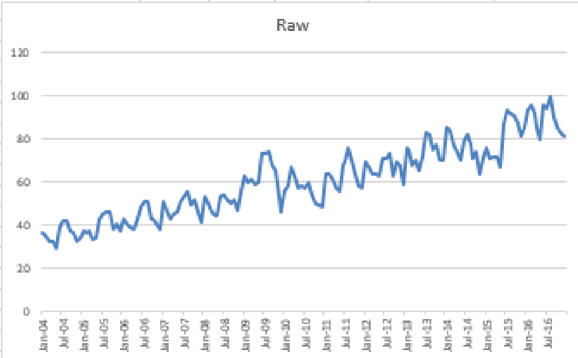
Zhi Da, J. E. (2013). *The Sum of All FEARS: Investor Sentiment and Asset Prices.* New South Wales.

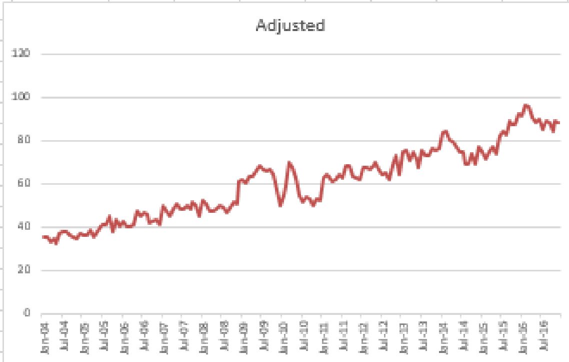
Appendix

Appendix (i) Removing Seasonality in Excel

date	social security office: (US/	Adjusted
Jan-04	36	35.29
Feb-04	35	35.14
Mar-04	32	33.15
Apr-04	32	34.18
May-04	29	31.93
Jun-04	39	36.59
Jul-04	42	38.08
Aug-04	42	37.48
Sep-04	37	36.04
Oct-04	36	35.69
Nov-04	32	34.41
Dec-04	34	36.97
Jan-05	37	36.27
Feb-05	36	36.15
Mar-05	37	38.32
Apr-05	33	35.25
May-05	34	37.44
Jun-05	43	40.34
Jul-05	45	40.80
Aug-05	46	41.05
Sep-05	46	44.81
Oct-05	38	37.67
Nov-05	40	43.02
Dec-05	37	40.23
Jan-06	43	42.15
Feb-06	40	40.16
Mar-06	39	40.40
Apr-06	38	40.59
May-06	43	47.35
Jun-06	48	45.03
Jul-06	51	46.25
Aug-06	51	45.51
Sep-06	43	41.89
Oct-06	43	42.63
Nov-06	40	43.02
Dec-06	38	41.32
Jan-07	51	49.99
Feb-07	47	47.19
Mar-07	43	44.54
Apr-07	45	48.07

	Monthly average	Monthly index
Jan	61.69	1.02
Feb	60.23	1.00
Mar	58.38	0.97
Apr	56.62	0.94
May	54.92	0.91
Jun	64.46	1.07
Jul	66.69	1.10
Aug	67.77	1.12
Sep	62.08	1.03
Oct	61.00	1.01
Nov	56.23	0.93
Dec	55.62	0.92
	60.47	1.00





Appendix (ii) Removing Seasonality in Excel (formula)

	A	B	C	D	E	F	G	H	I
1	date	social security office: (USA)	Adjusted					Monthly average	Monthly index
2	37987	36	=B2/$I$2				Jan	=GJENNOMSNITT(B2,B14,B26,B38,B50,B62,B74,B86,B98,B110,B122,B134,B146)	=H2/$H$14
3	38018	35	=B3/$I$3				Feb	=GJENNOMSNITT(B3,B15,B27,B39,B51,B63,B75,B87,B99,B111,B123,B135,B147)	=H3/$H$14
4	38047	32	=B4/$I$4				Mar	=GJENNOMSNITT(B4,B16,B28,B40,B52,B64,B76,B88,B100,B112,B124,B136,B148)	=H4/$H$14
5	38078	32	=B5/$I$5				Apr	=GJENNOMSNITT(B5,B17,B29,B41,B53,B65,B77,B89,B101,B113,B125,B137,B149)	=H5/$H$14
6	38108	29	=B6/$I$6				May	=GJENNOMSNITT(B6,B18,B30,B42,B54,B66,B78,B90,B102,B114,B126,B138,B150)	=H6/$H$14
7	38139	39	=B7/$I$7				Jun	=GJENNOMSNITT(B7,B19,B31,B43,B55,B67,B79,B91,B103,B115,B127,B139,B151)	=H7/$H$14
8	38169	42	=B8/$I$8				Jul	=GJENNOMSNITT(B8,B20,B32,B44,B56,B68,B80,B92,B104,B116,B128,B140,B152)	=H8/$H$14
9	38200	42	=B9/$I$9				Aug	=GJENNOMSNITT(B9,B21,B33,B45,B57,B69,B81,B93,B105,B117,B129,B141,B153)	=H9/$H$14
10	38231	37	=B10/$I$10				Sep	=GJENNOMSNITT(B10,B22,B34,B46,B58,B70,B82,B94,B106,B118,B130,B142,B154)	=H10/$H$14
11	38261	36	=B11/$I$11				Oct	=GJENNOMSNITT(B11,B23,B35,B47,B59,B71,B83,B95,B107,B119,B131,B143,B155)	=H11/$H$14
12	38292	32	=B12/$I$12				Nov	=GJENNOMSNITT(B12,B24,B36,B48,B60,B72,B84,B96,B108,B120,B132,B144,B156)	=H12/$H$14
13	38322	34	=B13/$I$13				Dec	=GJENNOMSNITT(B13,B25,B37,B49,B61,B73,B85,B97,B109,B121,B133,B145,B157)	=H13/$H$14
14	38353	37	=B14/$I$2					=GJENNOMSNITT(H2:H13)	=GJENNOMSNITT(I2:I13)
15	38384	36	=B15/$I$3						
16	38412	37	=B16/$I$4						
17	38443	33	=B17/$I$5						
18	38473	34	=B18/$I$6						
19	38504	43	=B19/$I$7						
20	38534	45	=B20/$I$8						
21	38565	46	=B21/$I$9						
22	38596	46	=B22/$I$10						
23	38626	38	=B23/$I$11						
24	38657	40	=B24/$I$12						
25	38687	37	=B25/$I$13						
26	38718	43	=B26/$I$2						
27	38749	40	=B27/$I$3						
28	38777	39	=B28/$I$4						
29	38808	38	=B29/$I$5						
30	38838	43	=B30/$I$6						
31	38869	48	=B31/$I$7						

Appendix (iii) Minimizing Squared Deviations by Changing Weights

			Total sum, minimized by changing the weights on the terms	
date	consumptio	sq deviation from indicator	890122.104	
Jan-04	8037.3	7311.780833		
Feb-04	8072.1	46087.36654		
Mar-04	8121.0	1948.140476		
Apr-04	8141.6	17616.17898		
May-04	8212.9	3461.54954		
Jun-04	8204.6	27432.58394		
Jul-04	8270.7	1315.546954		
Aug-04	8294.4	64543.47265		
Sep-04	8373.0	4666.163004		
Oct-04	8417.9	1816.552792		
Nov-04	8458.4	19414.11015		
Dec-04	8516.5	30470.60798		
Jan-05	8521.2	9284.940884		
Feb-05	8575.7	350.9109628		
Mar-05	8622.5	1107.843892		
Apr-05	8715.9	7914.48886		
May-05	8680.6	8249.569066		
Jun-05	8775.3	74295.04114		
Jul-05	8867.9	56572.12265		
Aug-05	8872.6	12801.50813		
Sep-05	8923.6	29337.85075		
Oct-05	8959.6	33468.40608		
Nov-05	8987.7	26806.18553		
Dec-05	9026.8	1787.752735		
Jan-06	9100.1	44956.16189		
Feb-06	9134.7	19908.86062		
Mar-06	9168.1	6495.962099		
Apr-06	9223.3	23764.99212		
May-06	9254.1	6466.922392		
Jun-06	9283.8	17239.48507		
Jul-06	9360.4	34327.14318		
Aug-06	9368.6	55831.05331		

Appendix (iv) Job Sentiment indicator

Job Sentiment Indicator

Search Term	Equal Weight	Optimized Weights, complete run	Optimized weights, predictive post june 2007
cover letter	0.0455	5.2576	0.0874
how to write a cover lett	0.0455	-12.8408	-6.4961
reeducation	0.0455	-0.0245	0.0081
going back to school	0.0455	1.9245	-0.3140
changing careers	0.0455	-0.3961	0.1048
career	0.0455	-2.0460	1.8195
job listing	0.0455	0.9959	0.6811
losing job	0.0455	-0.2291	-0.4293
work placement	0.0455	-0.1491	0.3473
how to find a job	0.0455	2.7719	0.0817
unemployment	0.0455	5.6397	2.4777
job application	0.0455	0.6398	-1.2744
interview questions	0.0455	-4.7513	0.3272
career test	0.0455	1.9523	0.2191
job security	0.0455	0.1541	1.4533
unemployment benefits	0.0455	-1.5755	-0.5869
food stamp eligibility	0.0455	1.5977	-0.3036
lost job	0.0455	0.8290	0.0430
retraining	0.0455	0.2101	0.2256
part time job	0.0455	-1.4784	-1.4401
cv	0.0455	3.4924	3.6046
layoffs	0.0455	-0.9742	0.3640
sum weights	1	1	1

Appendix (v) Consumption indicator

Consumption Indicator

Search Term	Equal Weight	Optimized Weights, complete run	Optimized weights, predictive post june 2007
buy	0.0500	0.082760523	0.540756381
loan	0.0500	0.223842897	0.089214697
cancel contract	0.0500	0.018997691	-0.084236443
where to buy	0.0500	0.034520756	0.501024262
apply for credit card	0.0500	-0.042752101	-0.02850133
online shopping	0.0500	0.011474016	-0.168922846
buy online	0.0500	-0.233369886	-0.098192212
buy new	0.0500	0.09294342	-0.039451015
best credit card	0.0500	0.074396121	-0.079327823
gift ideas	0.0500	-0.021200286	0.028632966
vacation ideas	0.0500	0.090747172	-0.014226527
coupons	0.0500	-0.020126303	0.280166599
refinance	0.0500	-0.034721833	-0.049212012
cost of	0.0500	0.368820225	0.216982162
can I afford	0.0500	-0.035959133	-0.074793918
down payment	0.0500	-0.045017871	0.038708317
shop	0.0500	0.371412187	0.107102818
cash for gold	0.0500	-0.004857628	-0.389213263
saving money	0.0500	0.036154145	0.109398734
value	0.0500	0.031935887	0.114090438
sum weights	1.00	1.00	1.00

Appendix (vi) Market sentiment indicator

Market Sentiment Indicator (FEARS Index)

Search Term	Equal Weight
gold prices	0.0333
recession	0.0333
gold price	0.0333
depression	0.0333
great depression	0.0333
gold	0.0333
economy	0.0333
price of gold	0.0333
the depression	0.0333
crisis	0.0333
frugal	0.0333
gdp	0.0333
charity	0.0333
bankruptcy	0.0333
unemployment	0.0333
inflation rate	0.0333
bankrupt	0.0333
the great depression	0.0333
car donate	0.0333
capitalization	0.0333
expense	0.0333
donation	0.0333
savings	0.0333
social security card	0.0333
the crisis	0.0333
default	0.0333
benefits	0.0333
unemployed	0.0333
poverty	0.0333
social security office	0.0333
sum weights	1.0000

Appendix (vii) Population in American states

	2006	2007	2008	2009	2010	2011
Alabama	4,629,000	4,673,000	4,718,000	4,758,000	4,785,000	4,801,000
Alaska	675,302	680,300	687,455	698,895	713,985	722,720
Arizona	6,029,000	6,168,000	6,280,000	6,343,000	6,414,000	6,469,000
Arkansas	2,822,000	2,849,000	2,875,000	2,897,000	2,922,000	2,939,000
California	36,020,000	36,250,000	36,600,000	36,960,000	37,350,000	37,700,000
Colorado	4,720,000	4,804,000	4,890,000	4,972,000	5,049,000	5,119,000
Connecticut	3,517,000	3,527,000	3,546,000	3,562,000	3,577,000	3,594,000
Delaware	859,268	871,749	883,874	891,730	899,769	907,619
District of Columbia	570,681	547,404	580,236	592,228	604,453	620,472
Florida	18,170,000	18,370,000	18,530,000	18,650,000	18,840,000	19,110,000
Georgia	9,156,000	9,350,000	9,505,000	9,621,000	9,713,000	9,812,000
Hawaii	1,310,000	1,316,000	1,332,000	1,347,000	1,364,000	1,378,000
Idaho	1,469,000	1,505,000	1,534,000	1,554,000	1,571,000	1,584,000
Illinois	12,640,000	12,700,000	12,750,000	12,800,000	12,840,000	12,860,000
Indiana	6,333,000	6,380,000	6,425,000	6,459,000	6,491,000	6,517,000
Iowa	2,983,000	2,999,000	3,017,000	3,033,000	3,050,000	3,065,000
Kansas	2,763,000	2,784,000	2,808,000	2,833,000	2,859,000	2,870,000
Kentucky	4,219,000	4,257,000	4,290,000	4,317,000	4,346,000	4,368,000
Louisiana	4,303,000	4,376,000	4,436,000	4,492,000	4,544,000	4,575,000
Maine	1,324,000	1,327,000	1,331,000	1,330,000	1,328,000	1,328,000
Maryland	5,627,000	5,653,000	5,685,000	5,730,000	5,786,000	5,844,000
Massachusetts	6,410,000	6,432,000	6,469,000	6,518,000	6,557,000	6,612,000
Michigan	10,040,000	10,000,000	9,947,000	9,902,000	9,878,000	9,877,000
Minnesota	5,164,000	5,207,000	5,247,000	5,281,000	5,311,000	5,348,000
Mississippi	2,905,000	2,928,000	2,948,000	2,959,000	2,970,000	2,978,000
Missouri	5,843,000	5,888,000	5,924,000	5,961,000	5,996,000	6,011,000
Montana	952,692	964,706	976,415	983,982	990,898	997,746
Nebraska	1,773,000	1,783,000	1,796,000	1,813,000	1,830,000	1,842,000
Nevada	2,523,000	2,601,000	2,654,000	2,685,000	2,705,000	2,719,000
New Hampshire	1,308,000	1,313,000	1,316,000	1,316,000	1,317,000	1,318,000
New Jersey	8,662,000	8,678,000	8,711,000	8,756,000	8,802,000	8,843,000
New Mexico	1,962,000	1,990,000	2,011,000	2,037,000	2,066,000	2,078,000
New York	8,251,000	8,310,000	8,347,000	8,392,000	8,192,000	8,287,000
North Carolina	8,917,000	9,118,000	9,309,000	9,450,000	9,562,000	9,651,000
North Dakota	649,422	652,822	657,569	664,968	674,499	685,326
Ohio	11,480,000	11,500,000	11,520,000	11,530,000	11,540,000	11,550,000
Oklahoma	3,594,000	3,634,000	3,669,000	3,718,000	3,762,000	3,787,000
Oregon	3,671,000	3,722,000	3,769,000	3,809,000	3,839,000	3,869,000
Pennsylvania	12,510,000	12,560,000	12,610,000	12,670,000	12,710,000	12,750,000
Rhode Island	1,063,000	1,057,000	1,055,000	1,054,000	1,053,000	1,052,000
South Carolina	4,358,000	4,444,000	4,529,000	4,590,000	4,636,000	4,673,000
South Dakota	783,033	791,623	799,124	807,067	816,463	824,289
Tennessee	6,089,000	6,167,000	6,247,000	6,306,000	6,357,000	6,398,000
Texas	23,360,000	23,830,000	24,310,000	24,800,000	25,260,000	25,650,000
Utah	2,526,000	2,598,000	2,663,000	2,723,000	2,776,000	2,816,000
Vermont	622,892	623,481	624,151	624,817	625,960	626,687
Virginia	7,674,000	7,751,000	7,833,000	7,926,000	8,025,000	8,111,000
Washington	6,371,000	6,462,000	6,562,000	6,667,000	6,744,000	6,823,000
West Virginia	1,828,000	1,834,000	1,840,000	1,848,000	1,854,000	1,855,000
Wisconsin	5,578,000	5,611,000	5,641,000	5,669,000	5,691,000	5,710,000
Wyoming	522,667	534,867	546,043	559,851	564,460	567,768

Appendix (viii) Outbound Package Shipments

	2006	2007	2008	2009	2010	2011
Alabama	2,326	2,151	2,200	1,691	1,880	1,749
Alaska						
Arizona	7,838	6,911	5,992	4,143	4,194	4,221
Arkansas	1,487	1,303	1,098	797	873	948
California	27,435	23,538	20,380	14,729	14,892	14,758
Colorado	6,336	6,322	5,414	3,642	3,956	4,080
Connecticut	3,013	2,817	2,740	1,971	2,093	2,155
Delaware	594	566	536	385	433	444
District of Columbia	800	793	710	477	516	588
Florida	17,019	15,123	13,470	9,378	8,745	9,067
Georgia	7,967	7,292	6,989	4,849	4,775	5,489
Hawaii						
Idaho	1,266	1,398	1,182	657	624	711
Illinois	9,500	9,599	9,097	6,365	7,153	6,821
Indiana	3,871	3,660	3,642	2,584	2,474	2,133
Iowa	1,316	1,297	1,290	1,101	1,072	1,039
Kansas	2,285	2,223	2,320	1,706	1,883	1,934
Kentucky	2,151	2,168	2,087	1,645	1,680	2,119
Louisiana	2,872	2,244	2,010	1,494	1,332	1,429
Maine	1,040	918	1,082	823	735	747
Maryland	5,440	4,660	4,562	3,404	3,518	3,299
Massachusetts	5,037	4,357	4,253	3,157	3,190	3,129
Michigan	6,816	7,074	6,680	4,846	4,458	3,487
Minnesota	4,250	3,960	3,887	3,264	2,918	2,865
Mississippi	1,448	1,345	1,204	949	962	943
Missouri	4,571	3,983	4,057	2,973	3,847	2,827
Montana	1,020	1,050	866	579	751	886
Nebraska	1,119	1,153	1,229	857	998	921
Nevada	2,278	2,186	1,928	1,303	1,451	1,373
New Hampshire	1,007	850	918	721	715	708
New Jersey	6,244	6,220	5,322	4,239	5,179	4,190
New Mexico	1,767	2,061	1,693	1,104	1,141	1,346
New York	10,354	9,804	8,693	6,223	6,315	6,574
North Carolina	6,106	6,120	6,219	4,590	4,648	5,122
North Dakota	933	945	759	438	487	523
Ohio	8,230	7,771	7,131	5,282	5,575	4,748
Oklahoma	2,211	2,420	2,087	1,381	1,597	1,456
Oregon	3,042	3,213	2,966	1,664	1,635	1,694
Pennsylvania	7,534	7,194	7,678	5,411	5,528	4,662
Rhode Island	722	576	691	422	402	531
South Carolina	2,958	3,044	3,082	2,211	2,111	2,328
South Dakota	391	349	343	273	315	355
Tennessee	3,987	3,776	3,608	2,845	2,876	2,927
Texas	17,224	16,707	14,908	10,622	11,011	10,800
Utah	2,051	1,909	1,801	1,206	1,462	1,424
Vermont	377	320	332	282	242	255
Virginia	8,309	7,477	8,081	6,079	5,804	6,691
Washington	7,651	7,210	7,181	5,004	4,874	4,626
West Virginia	514	485	557	399	413	491
Wisconsin	4,050	3,941	3,549	2,726	2,741	2,594
Wyoming	497	452	458	303	363	417

Appendix (ix) Relative Emigration Intensity

	2006	2007	2008	2009	2010	2011
Alabama	5.02	4.60	4.66	3.55	3.93	3.64
Alaska	0.00	0.00	0.00	0.00	0.00	0.00
Arizona	13.00	11.20	9.54	6.53	6.54	6.52
Arkansas	5.27	4.57	3.82	2.75	2.99	3.23
California	7.62	6.49	5.57	3.99	3.99	3.91
Colorado	13.42	13.16	11.07	7.33	7.84	7.97
Connecticut	8.57	7.99	7.73	5.53	5.85	6.00
Delaware	6.91	6.49	6.06	4.32	4.81	4.89
District of Columbia	14.02	14.49	12.24	8.05	8.54	9.48
Florida	9.37	8.23	7.27	5.03	4.64	4.74
Georgia	8.70	7.80	7.35	5.04	4.92	5.59
Hawaii	0.00	0.00	0.00	0.00	0.00	0.00
Idaho	8.62	9.29	7.71	4.23	3.97	4.49
Illinois	7.52	7.56	7.13	4.97	5.57	5.30
Indiana	6.11	5.74	5.67	4.00	3.81	3.27
Iowa	4.41	4.32	4.28	3.63	3.51	3.39
Kansas	8.27	7.98	8.26	6.02	6.59	6.74
Kentucky	5.10	5.09	4.86	3.81	3.87	4.85
Louisiana	6.67	5.13	4.53	3.33	2.93	3.12
Maine	7.85	6.92	8.13	6.19	5.53	5.63
Maryland	9.67	8.24	8.02	5.94	6.08	5.65
Massachusetts	7.86	6.77	6.57	4.84	4.87	4.73
Michigan	6.79	7.07	6.72	4.89	4.51	3.53
Minnesota	8.23	7.61	7.41	6.18	5.49	5.36
Mississippi	4.98	4.59	4.08	3.21	3.24	3.17
Missouri	7.82	6.76	6.85	4.99	6.42	4.70
Montana	10.71	10.88	8.87	5.88	7.58	8.88
Nebraska	6.31	6.47	6.84	4.73	5.45	5.00
Nevada	9.03	8.40	7.26	4.85	5.36	5.05
New Hampshire	7.70	6.47	6.98	5.48	5.43	5.37
New Jersey	7.21	7.17	6.11	4.84	5.88	4.74
New Mexico	9.01	10.36	8.42	5.42	5.52	6.48
New York	12.55	11.80	10.41	7.42	7.71	7.93
North Carolina	6.85	6.71	6.68	4.86	4.86	5.31
North Dakota	14.37	14.48	11.54	6.59	7.22	7.63
Ohio	7.17	6.76	6.19	4.58	4.83	4.11
Oklahoma	6.15	6.66	5.69	3.71	4.25	3.84
Oregon	8.29	8.63	7.87	4.37	4.26	4.38
Pennsylvania	6.02	5.73	6.09	4.27	4.35	3.66
Rhode Island	6.79	5.45	6.55	4.00	3.82	5.05
South Carolina	6.79	6.85	6.81	4.82	4.55	4.98
South Dakota	4.99	4.41	4.29	3.38	3.86	4.31
Tennessee	6.55	6.12	5.78	4.51	4.52	4.57
Texas	7.37	7.01	6.13	4.28	4.36	4.21
Utah	8.12	7.35	6.76	4.43	5.27	5.06
Vermont	6.05	5.13	5.32	4.51	3.87	4.07
Virginia	10.83	9.65	10.32	7.67	7.23	8.25
Washington	12.01	11.16	10.94	7.51	7.23	6.78
West Virginia	2.81	2.64	3.03	2.16	2.23	2.65
Wisconsin	7.26	7.02	6.29	4.81	4.82	4.54
Wyoming	9.51	8.45	8.39	5.41	6.43	7.34

Appendix (x) MISERY index

The MISERY index	North Dakota	Arizona	Iowa	West-Virginia
new job	0.56	0.14	0.25	0.04
unemployment	-0.21	0.00	-0.22	-0.06
divorce	0.10	-0.11	-0.10	-0.10
divorced	-0.15	0.13	0.07	0.10
retire	0.00	0.19	-0.06	-0.10
rape	0.01	0.17	-0.06	0.02
lawyer	0.24	-0.20	-0.04	-0.03
prison	-0.11	-0.22	0.01	-0.04
drugs	0.00	0.33	0.01	-0.03
surgery	-0.20	-0.15	-0.08	0.00
heart attack	0.09	0.22	0.05	-0.06
health	-0.01	0.24	0.15	0.09
funeral	-0.07	-0.23	0.04	-0.11
death	-0.04	-0.02	0.02	0.09
accident	0.00	-0.32	-0.04	-0.06
custody	0.08	-0.56	0.12	-0.02
obesity	-0.13	0.07	0.10	0.00
debt	0.17	-0.04	0.02	0.13
murder	-0.03	0.19	0.03	0.02
college application	0.48	-0.03	0.02	0.54
divorce lawyer	0.38	0.72	0.65	0.49
retired	-0.22	0.06	0.00	-0.02
suicide	0.07	0.41	0.07	0.12

Appendix (xi) Optimizing weights for individual state emigration

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Total Squares																												
2	0.000204			North Dakota			0.56	-0.21	0.10	-0.15	0.00	0.01	0.24	-0.11	0.00	-0.20	0.09	-0.01	-0.07	-0.04	0.00	0.08	-0.13	0.17	-0.03	0.48	0.38	-0.22	0.07
3		Squared Dev.	Emigration	intensity	12.510115	Terms:	new job	unemployment	divorce	divorced	retire	rape	lawyer	prison	drugs	surgery	heart attack	health	funeral	death	accident	custody	obesity	debt	murder	college application	divorce lawyer	retired	suicide
4		9.66283E-06	14.37		179.773461		292	115	482	124	253	360	345	459	524	414	327	540	321	615	361	299	346	515	585	92	0	200	477
5		1.41819E-05	14.48		181.142699		261	289	467	258	268	570	285	517	572	667	334	586	382	627	361	448	418	484	751	338	130	225	480
6		1.98369E-06	11.54		144.365319		394	344	411	125	109	377	272	502	557	469	307	608	526	662	447	328	359	553	573	125	93	399	494
7		5.41157E-06	6.59		82.4393316		347	687	601	165	396	600	377	469	621	699	456	718	700	727	501	439	330	656	594	145	49	317	527
8		1.24715E-05	7.22		90.3194988		429	874	833	336	225	579	460	641	731	874	555	883	884	938	633	356	491	716	595	189	250	471	612
9		2.10736E-05	7.63		95.4567681		360	658	802	399	520	629	543	631	681	884	408	906	944	843	591	508	402	758	664	183	72	259	608
10																													
11				Arizona			0.14	0.00	-0.11	0.13	0.19	0.17	-0.20	-0.22	0.33	-0.15	0.22	0.24	-0.23	-0.02	-0.32	-0.56	0.07	-0.04	0.19	-0.03	0.72	0.06	0.41
12		Squared Dev.	Emigration	intensity	48.2609851	Terms:	new job	unemployment	divorce	divorced	retire	rape	lawyer	prison	drugs	surgery	heart attack	health	funeral	death	accident	custody	obesity	debt	murder	college application	divorce lawyer	retired	suicide
13		1.39306E-05	13		627.396538		486	226	708	391	580	769	823	880	963	756	512	768	500	797	564	767	682	468	962	423	525	533	804
14		4.26425E-06	11.2		540.520968		477	241	651	345	539	710	792	941	811	717	489	728	574	751	594	836	679	524	934	445	680	531	702
15		1.5632E-05	9.54		460.405844		410	387	633	392	512	721	783	935	821	708	478	724	637	827	637	723	569	644	917	386	570	574	646
16		7.39716E-07	6.53		315.145093		570	717	776	519	487	759	833	923	893	865	684	819	772	966	772	813	588	729	860	517	416	626	697
17		1.29536E-05	6.54		315.623243		669	720	956	560	681	846	905	926	997	1001	608	847	842	974	946	913	771	724	893	682	490	744	732
18		2.04371E-06	6.52		314.663052		579	550	837	654	606	768	978	867	902	865	558	668	676	922	800	901	573	659	799	499	546	649	679
19																													
20				Iowa			0.25	-0.22	-0.10	0.07	-0.06	-0.06	-0.04	0.01	0.01	-0.08	0.05	0.15	0.04	0.02	-0.04	0.12	0.10	0.02	0.03	0.02	0.65	0.00	0.07
21		Squared Dev.	Emigration	intensity	82.9003781	Terms:	new job	unemployment	divorce	divorced	retire	rape	lawyer	prison	drugs	surgery	heart attack	health	funeral	death	accident	custody	obesity	debt	murder	college application	divorce lawyer	retired	suicide
22		1.77442E-06	4.41		365.589335		242	214	734	219	463	741	690	675	709	746	514	723	335	739	409	404	548	464	954	278	335	457	761
23		6.74299E-07	4.32		358.130455		270	215	625	311	402	673	654	662	588	722	394	707	403	717	463	654	584	445	826	406	248	539	621
24		3.01778E-07	4.28		354.813069		503	300	580	400	277	618	618	573	592	644	395	653	529	692	471	490	489	496	770	315	188	584	677
25		1.96431E-05	3.63		300.92394		426	777	717	488	439	641	635	558	636	765	528	832	684	820	534	521	541	502	757	376	268	570	648
26		1.07968E-09	3.51		290.98036		462	870	904	362	554	818	678	677	775	1028	514	892	911	955	818	607	641	609	852	635	303	572	728
27		1.8362E-05	3.39		281.036567		315	727	799	655	522	723	775	539	619	811	505	599	691	782	635	516	324	589	710	437	384	500	604
28																													
29				West-Virginia			0.04	-0.06	-0.10	0.10	-0.10	0.02	-0.03	-0.04	-0.03	0.00	-0.06	0.09	-0.11	0.09	-0.06	-0.02	0.00	0.13	0.02	0.54	0.49	-0.02	0.12
30		Squared Dev.	Emigration	intensity	50.2984581	Terms:	new job	unemployment	divorce	divorced	retire	rape	lawyer	prison	drugs	surgery	heart attack	health	funeral	death	accident	custody	obesity	debt	murder	college application	divorce lawyer	retired	suicide
31		1.06411E-06	2.81		141.337636		268	240	489	205	282	604	641	596	623	703	338	594	461	588	503	434	556	544	850	166	0	368	764
32		8.33843E-06	2.64		132.790817		160	239	663	351	334	517	633	679	586	571	402	601	494	518	488	551	500	493	680	23	272	479	580
33		1.27816E-05	3.03		152.400753		341	284	611	377	156	556	535	631	526	538	332	575	535	563	522	429	364	528	544	96	153	403	537
34		2.00958E-05	2.16		108.640187		236	606	793	302	120	519	618	595	633	701	427	702	683	639	530	374	393	694	557	109	129	495	496
35		5.63944E-06	2.23		112.167936		578	830	964	211	271	695	768	850	827	892	585	818	907	738	806	492	443	754	753	271	106	629	591
36		5.66756E-07	2.65		133.291667		425	575	840	580	333	728	713	586	680	754	540	559	913	700	683	460	355	628	588	187	166	499	593

Appendix (xii) Optimizing weights for individual states (formulas)

	A	B	C	D	E	F	G
1	Total Squares						
2	=SUMMER(B4:B36)			North Dakota			0.564255398266205
3		Squared Dev.	Emigration intensity		12.5101150024592	Terms:	new job
4		=(E4-C4*$E$3)^2	14.37		=$G$2*G4+$H$2*H4+$I$2*I4+$J$2*J4+$K$2*K4+$L$2*L4+$M$2*M4+$N$2*N4+$O$2*O4+$P$2*P4+$Q$2*Q4+$R$2*R4+$S$2*S4+$T$2*T4+$U$2*U4+$V$2*V4+$W$2*W4+$X$2*X4+$Y$2*Y4+$Z$2*Z4+$AA$2*AA4+$AB$2*AB4+$AC$2*AC4		292
5		=(E5-C5*$E$3)^2	14.48		=$G$2*G5+$H$2*H5+$I$2*I5+$J$2*J5+$K$2*K5+$L$2*L5+$M$2*M5+$N$2*N5+$O$2*O5+$P$2*P5+$Q$2*Q5+$R$2*R5+$S$2*S5+$T$2*T5+$U$2*U5+$V$2*V5+$W$2*W5+$X$2*X5+$Y$2*Y5+$Z$2*Z5+$AA$2*AA5+$AB$2*AB5+$AC$2*AC5		261
6		=(E6-C6*$E$3)^2	11.54		=$G$2*G6+$H$2*H6+$I$2*I6+$J$2*J6+$K$2*K6+$L$2*L6+$M$2*M6+$N$2*N6+$O$2*O6+$P$2*P6+$Q$2*Q6+$R$2*R6+$S$2*S6+$T$2*T6+$U$2*U6+$V$2*V6+$W$2*W6+$X$2*X6+$Y$2*Y6+$Z$2*Z6+$AA$2*AA6+$AB$2*AB6+$AC$2*AC6		394
7		=(E7-C7*$E$3)^2	6.59		=$G$2*G7+$H$2*H7+$I$2*I7+$J$2*J7+$K$2*K7+$L$2*L7+$M$2*M7+$N$2*N7+$O$2*O7+$P$2*P7+$Q$2*Q7+$R$2*R7+$S$2*S7+$T$2*T7+$U$2*U7+$V$2*V7+$W$2*W7+$X$2*X7+$Y$2*Y7+$Z$2*Z7+$AA$2*AA7+$AB$2*AB7+$AC$2*AC7		347
8		=(E8-C8*$E$3)^2	7.22		=$G$2*G8+$H$2*H8+$I$2*I8+$J$2*J8+$K$2*K8+$L$2*L8+$M$2*M8+$N$2*N8+$O$2*O8+$P$2*P8+$Q$2*Q8+$R$2*R8+$S$2*S8+$T$2*T8+$U$2*U8+$V$2*V8+$W$2*W8+$X$2*X8+$Y$2*Y8+$Z$2*Z8+$AA$2*AA8+$AB$2*AB8+$AC$2*AC8		429
9		=(E9-C9*$E$3)^2	7.63		=$G$2*G9+$H$2*H9+$I$2*I9+$J$2*J9+$K$2*K9+$L$2*L9+$M$2*M9+$N$2*N9+$O$2*O9+$P$2*P9+$Q$2*Q9+$R$2*R9+$S$2*S9+$T$2*T9+$U$2*U9+$V$2*V9+$W$2*W9+$X$2*X9+$Y$2*Y9+$Z$2*Z9+$AA$2*AA9+$AB$2*AB9+$AC$2*AC9		360
10							
11				Arizona			0.141018368419867
12		Squared Dev.	Emigration intensity		48.2609850800567	Terms:	new job
13		=(E13-C13*$E$12)^2	13		=$G$11*G13+$H$11*H13+$I$11*I13+$J$11*J13+$K$11*K13+$L$11*L13+$M$11*M13+$N$11*N13+$O$11*O13+$P$11*P13+$Q$11*Q13+$R$11*R13+$S$11*S13+$T$11*T13+$U$11*U13+$V$11*V13+$W$11*W13+$X$11*X13+$Y$11*Y13+$Z$11*Z13+$AA$11*AA13+$AB$11*AB13+$AC$11*AC13		486
14		=(E14-C14*$E$12)^2	11.2		=$G$11*G14+$H$11*H14+$I$11*I14+$J$11*J14+$K$11*K14+$L$11*L14+$M$11*M14+$N$11*N14+$O$11*O14+$P$11*P14+$Q$11*Q14+$R$11*R14+$S$11*S14+$T$11*T14+$U$11*U14+$V$11*V14+$W$11*W14+$X$11*X14+$Y$11*Y14+$Z$11*Z14+$AA$11*AA14+$AB$11*AB14+$AC$11*AC14		477
15		=(E15-C15*$E$12)^2	9.54		=$G$11*G15+$H$11*H15+$I$11*I15+$J$11*J15+$K$11*K15+$L$11*L15+$M$11*M15+$N$11*N15+$O$11*O15+$P$11*P15+$Q$11*Q15+$R$11*R15+$S$11*S15+$T$11*T15+$U$11*U15+$V$11*V15+$W$11*W15+$X$11*X15+$Y$11*Y15+$Z$11*Z15+$AA$11*AA15+$AB$11*AB15+$AC$11*AC15		410
16		=(E16-C16*$E$12)^2	6.53		=$G$11*G16+$H$11*H16+$I$11*I16+$J$11*J16+$K$11*K16+$L$11*L16+$M$11*M16+$N$11*N16+$O$11*O16+$P$11*P16+$Q$11*Q16+$R$11*R16+$S$11*S16+$T$11*T16+$U$11*U16+$V$11*V16+$W$11*W16+$X$11*X16+$Y$11*Y16+$Z$11*Z16+$AA$11*AA16+$AB$11*AB16+$AC$11*AC16		570
17		=(E17-C17*$E$12)^2	6.54		=$G$11*G17+$H$11*H17+$I$11*I17+$J$11*J17+$K$11*K17+$L$11*L17+$M$11*M17+$N$11*N17+$O$11*O17+$P$11*P17+$Q$11*Q17+$R$11*R17+$S$11*S17+$T$11*T17+$U$11*U17+$V$11*V17+$W$11*W17+$X$11*X17+$Y$11*Y17+$Z$11*Z17+$AA$11*AA17+$AB$11*AB17+$AC$11*AC17		669
18		=(E18-C18*$E$12)^2	6.52		=$G$11*G18+$H$11*H18+$I$11*I18+$J$11*J18+$K$11*K18+$L$11*L18+$M$11*M18+$N$11*N18+$O$11*O18+$P$11*P18+$Q$11*Q18+$R$11*R18+$S$11*S18+$T$11*T18+$U$11*U18+$V$11*V18+$W$11*W18+$X$11*X18+$Y$11*Y18+$Z$11*Z18+$AA$11*AA18+$AB$11*AB18+$AC$11*AC18		579
19							
20				Iowa			0.246319782659734
21		Squared Dev.	Emigration intensity		82.9003780908964	Terms:	new job
22		=(E22-C22*$E$21)^2	4.41		=$G$20*G22+$H$20*H22+$I$20*I22+$J$20*J22+$K$20*K22+$L$20*L22+$M$20*M22+$N$20*N22+$O$20*O22+$P$20*P22+$Q$20*Q22+$R$20*R22+$S$20*S22+$T$20*T22+$U$20*U22+$V$20*V22+$W$20*W22+$X$20*X22+$Y$20*Y22+$Z$20*Z22+$AA$20*AA22+$AB$20*AB22+$AC$20*AC22		242
23		=(E23-C23*$E$21)^2	4.32		=$G$20*G23+$H$20*H23+$I$20*I23+$J$20*J23+$K$20*K23+$L$20*L23+$M$20*M23+$N$20*N23+$O$20*O23+$P$20*P23+$Q$20*Q23+$R$20*R23+$S$20*S23+$T$20*T23+$U$20*U23+$V$20*V23+$W$20*W23+$X$20*X23+$Y$20*Y23+$Z$20*Z23+$AA$20*AA23+$AB$20*AB23+$AC$20*AC23		270
24		=(E24-C24*$E$21)^2	4.28		=$G$20*G24+$H$20*H24+$I$20*I24+$J$20*J24+$K$20*K24+$L$20*L24+$M$20*M24+$N$20*N24+$O$20*O24+$P$20*P24+$Q$20*Q24+$R$20*R24+$S$20*S24+$T$20*T24+$U$20*U24+$V$20*V24+$W$20*W24+$X$20*X24+$Y$20*Y24+$Z$20*Z24+$AA$20*AA24+$AB$20*AB24+$AC$20*AC24		503
25		=(E25-C25*$E$21)^2	3.63		=$G$20*G25+$H$20*H25+$I$20*I25+$J$20*J25+$K$20*K25+$L$20*L25+$M$20*M25+$N$20*N25+$O$20*O25+$P$20*P25+$Q$20*Q25+$R$20*R25+$S$20*S25+$T$20*T25+$U$20*U25+$V$20*V25+$W$20*W25+$X$20*X25+$Y$20*Y25+$Z$20*Z25+$AA$20*AA25+$AB$20*AB25+$AC$20*AC25		426
26		=(E26-C26*$E$21)^2	3.51		=$G$20*G26+$H$20*H26+$I$20*I26+$J$20*J26+$K$20*K26+$L$20*L26+$M$20*M26+$N$20*N26+$O$20*O26+$P$20*P26+$Q$20*Q26+$R$20*R26+$S$20*S26+$T$20*T26+$U$20*U26+$V$20*V26+$W$20*W26+$X$20*X26+$Y$20*Y26+$Z$20*Z26+$AA$20*AA26+$AB$20*AB26+$AC$20*AC26		462
27		=(E27-C27*$E$21)^2	3.39		=$G$20*G27+$H$20*H27+$I$20*I27+$J$20*J27+$K$20*K27+$L$20*L27+$M$20*M27+$N$20*N27+$O$20*O27+$P$20*P27+$Q$20*Q27+$R$20*R27+$S$20*S27+$T$20*T27+$U$20*U27+$V$20*V27+$W$20*W27+$X$20*X27+$Y$20*Y27+$Z$20*Z27+$AA$20*AA27+$AB$20*AB27+$AC$20*AC27		315
28							
29				West-Virginia			0.0388497198471782
30		Squared Dev.	Emigration intensity		50.2984580795272	Terms:	new job
31		=(E31-C31*$E$30)^2	2.81		=$G$29*G31+$H$29*H31+$I$29*I31+$J$29*J31+$K$29*K31+$L$29*L31+$M$29*M31+$N$29*N31+$O$29*O31+$P$29*P31+$Q$29*Q31+$R$29*R31+$S$29*S31+$T$29*T31+$U$29*U31+$V$29*V31+$W$29*W31+$X$29*X31+$Y$29*Y31+$Z$29*Z31+$AA$29*AA31+$AB$29*AB31+$AC$29*AC31		268
32		=(E32-C32*$E$30)^2	2.64		=$G$29*G32+$H$29*H32+$I$29*I32+$J$29*J32+$K$29*K32+$L$29*L32+$M$29*M32+$N$29*N32+$O$29*O32+$P$29*P32+$Q$29*Q32+$R$29*R32+$S$29*S32+$T$29*T32+$U$29*U32+$V$29*V32+$W$29*W32+$X$29*X32+$Y$29*Y32+$Z$29*Z32+$AA$29*AA32+$AB$29*AB32+$AC$29*AC32		160
33		=(E33-C33*$E$30)^2	3.03		=$G$29*G33+$H$29*H33+$I$29*I33+$J$29*J33+$K$29*K33+$L$29*L33+$M$29*M33+$N$29*N33+$O$29*O33+$P$29*P33+$Q$29*Q33+$R$29*R33+$S$29*S33+$T$29*T33+$U$29*U33+$V$29*V33+$W$29*W33+$X$29*X33+$Y$29*Y33+$Z$29*Z33+$AA$29*AA33+$AB$29*AB33+$AC$29*AC33		341
34		=(E34-C34*$E$30)^2	2.16		=$G$29*G34+$H$29*H34+$I$29*I34+$J$29*J34+$K$29*K34+$L$29*L34+$M$29*M34+$N$29*N34+$O$29*O34+$P$29*P34+$Q$29*Q34+$R$29*R34+$S$29*S34+$T$29*T34+$U$29*U34+$V$29*V34+$W$29*W34+$X$29*X34+$Y$29*Y34+$Z$29*Z34+$AA$29*AA34+$AB$29*AB34+$AC$29*AC34		236
35		=(E35-C35*$E$30)^2	2.23		=$G$29*G35+$H$29*H35+$I$29*I35+$J$29*J35+$K$29*K35+$L$29*L35+$M$29*M35+$N$29*N35+$O$29*O35+$P$29*P35+$Q$29*Q35+$R$29*R35+$S$29*S35+$T$29*T35+$U$29*U35+$V$29*V35+$W$29*W35+$X$29*X35+$Y$29*Y35+$Z$29*Z35+$AA$29*AA35+$AB$29*AB35+$AC$29*AC35		578
36		=(E36-C36*$E$30)^2	2.65		=$G$29*G36+$H$29*H36+$I$29*I36+$J$29*J36+$K$29*K36+$L$29*L36+$M$29*M36+$N$29*N36+$O$29*O36+$P$29*P36+$Q$29*Q36+$R$29*R36+$S$29*S36+$T$29*T36+$U$29*U36+$V$29*V36+$W$29*W36+$X$29*X36+$Y$29*Y36+$Z$29*Z36+$AA$29*AA36+$AB$29*AB36+$AC$29*AC36		425

Appendix (xiii) Volatility Indicator

Firm Search Volatility (Yahoo)

Search Term	Optimized Weighting
yahoo merger	0.030736322
yahoo merge	-0.071063999
yahoo bankrupt	0.067125404
marissa mayer	0.112605169
yahoo bankruptcy	0.051840057
yahoo buy	0.077692206
yahoo sell	-0.071523049
yahoo crisis	-0.013638537
yahoo good	-0.072958386
yahoo bad	0.029703353
yahoo new	0.548934907
yahoo takeover	0.042889406
yahoo fear	0.022467302
yahoo buyout	0.03304012
yahoo risk	-0.042491168
yahoo cash	-0.006848179
yahoo earnings	0.090803038
yahoo announcement	0.095317921
yahoo press release	0.022938962
yahoo income	-0.048111846
yahoo assets	0.038445653
yahoo bond	-0.056232649
yahoo net worth	0.05774685
yahoo ceo	0.130963947
yahoo option	-0.015008689
yahoo fails	-0.055374117
sum weights	0.0000

Appendix (xiv) Calculating predictive power in excel

Date	Implied Volatilty		Indicator		indicator	yahoo merger: (Hele	yahoo merge: (H	yahoo bankrupt	marissa mayer:
01/01/2015		45.79655869		58.86	58	46	38	17
02/01/2015	36.31	39.52067953		50.80	100	29	32	42
03/01/2015		29.86541172		38.39	26	20	34	31
04/01/2015		26.58345922		34.17	52	53	33	28
05/01/2015	39.62	33.28861128	0	42.79	45	64	58	43
06/01/2015	40.68	38.02100539	1	48.87	23	47	29	64
07/01/2015	40.42	34.22128249	1	43.99	88	51	0	36
08/01/2015	40.73	39.25419177	1	50.46	44	23	29	53
09/01/2015	39.33	33.39447743	1	42.92	22	52	0	42
10/01/2015		33.23889738		42.72	38	50	33	30
11/01/2015		25.52120344		32.80	38	20	32	36
12/01/2015	43.94	37.65515976	1	48.40	44	63	29	54
13/01/2015	48.2	35.42494707	0	45.53	65	34	70	38
14/01/2015	48.62	34.23023777	0	44.00	22	23	28	31
15/01/2015	51.5	39.6364501	1	50.95	55	59	56	31
16/01/2015	50.41	35.18091661	1	45.22	0	61	44	48

Appendix (xv) Calculating predictive power in excel (formulas)

B	C	D	E	F	
				0.778002547370479	0.0
Date	Implied Volatilty (30)	Indicator		indicator	yal
42005		=F3*$F$1		=$G$1*G3+$H$1*H3+$I$1*I3+$J$1*J3+$K$1*K3+$L$1*L3+$M$1*M3+$N$1*N3	58
42006	36.31	=F4*$F$1		=$G$1*G4+$H$1*H4+$I$1*I4+$J$1*J4+$K$1*K4+$L$1*L4+$M$1*M4+$N$1*N4	100
42007		=F5*$F$1		=$G$1*G5+$H$1*H5+$I$1*I5+$J$1*J5+$K$1*K5+$L$1*L5+$M$1*M5+$N$1*N5	26
42008		=F6*$F$1		=$G$1*G6+$H$1*H6+$I$1*I6+$J$1*J6+$K$1*K6+$L$1*L6+$M$1*M6+$N$1*N6	52
42009	39.62	=F7*$F$1	=HVIS((((D7/D4)-1)*((C7/C4)-1))<0,0,1)	=$G$1*G7+$H$1*H7+$I$1*I7+$J$1*J7+$K$1*K7+$L$1*L7+$M$1*M7+$N$1*N7	45
42010	40.68	=F8*$F$1	=HVIS((((D8/D7)-1)*((C8/C7)-1))<0,0,1)	=$G$1*G8+$H$1*H8+$I$1*I8+$J$1*J8+$K$1*K8+$L$1*L8+$M$1*M8+$N$1*N8	23
42011	40.42	=F9*$F$1	=HVIS((((D9/D8)-1)*((C9/C8)-1))<0,0,1)	=$G$1*G9+$H$1*H9+$I$1*I9+$J$1*J9+$K$1*K9+$L$1*L9+$M$1*M9+$N$1*N9	88
42012	40.73	=F10*$F$1	=HVIS((((D10/D9)-1)*((C10/C9)-1))<0,0,1)	=$G$1*G10+$H$1*H10+$I$1*I10+$J$1*J10+$K$1*K10+$L$1*L10+$M$1*M10+$N	44
42013	39.33	=F11*$F$1	=HVIS((((D11/D10)-1)*((C11/C10)-1))<0,0,1)	=$G$1*G11+$H$1*H11+$I$1*I11+$J$1*J11+$K$1*K11+$L$1*L11+$M$1*M11+$N$1*	22
42014		=F12*$F$1		=$G$1*G12+$H$1*H12+$I$1*I12+$J$1*J12+$K$1*K12+$L$1*L12+$M$1*M12+$N	38
42015		=F13*$F$1		=$G$1*G13+$H$1*H13+$I$1*I13+$J$1*J13+$K$1*K13+$L$1*L13+$M$1*M13+$N	38
42016	43.94	=F14*$F$1	=HVIS((((D14/D11)-1)*((C14/C11)-1))<0,0,1)	=$G$1*G14+$H$1*H14+$I$1*I14+$J$1*J14+$K$1*K14+$L$1*L14+$M$1*M14+$N	44
42017	48.2	=F15*$F$1	=HVIS((((D15/D14)-1)*((C15/C14)-1))<0,0,1)	=$G$1*G15+$H$1*H15+$I$1*I15+$J$1*J15+$K$1*K15+$L$1*L15+$M$1*M15+$N	65
42018	48.62	=F16*$F$1	=HVIS((((D16/D15)-1)*((C16/C15)-1))<0,0,1)	=$G$1*G16+$H$1*H16+$I$1*I16+$J$1*J16+$K$1*K16+$L$1*L16+$M$1*M16+$N	22
42019	51.5	=F17*$F$1	=HVIS((((D17/D16)-1)*((C17/C16)-1))<0,0,1)	=$G$1*G17+$H$1*H17+$I$1*I17+$J$1*J17+$K$1*K17+$L$1*L17+$M$1*M17+$N	55
42020	50.41	=F18*$F$1	=HVIS((((D18/D17)-1)*((C18/C17)-1))<0,0,1)	=$G$1*G18+$H$1*H18+$I$1*I18+$J$1*J18+$K$1*K18+$L$1*L18+$M$1*M18+$N	0
42021		=F19*$F$1		=$G$1*G19+$H$1*H19+$I$1*I19+$J$1*J19+$K$1*K19+$L$1*L19+$M$1*M19+$N	0