Widespread Worry and the Stock Market

Eric Gilbert and Karrie Karahalios

Department of Computer Science
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

Abstract

Our emotional state influences our choices. Research on how it happens usually comes from the lab. We know relatively little about how real world emotions affect real world settings, like financial markets. Here, we demonstrate that estimating emotions from weblogs provides novel information about future stock market prices. That is, it provides information not already apparent from market data. Specifically, we estimate anxiety, worry and fear from a dataset of over 20 million posts made on the site LiveJournal. Using a Granger-causal framework, we find that increases in expressions of anxiety, evidenced by computationally-identified linguistic features, predict downward pressure on the S&P 500 index. We also present a confirmation of this result via Monte Carlo simulation. The findings show how the mood of millions in a large online community, even one that primarily discusses daily life, can anticipate changes in a seemingly unrelated system. Beyond this, the results suggest new ways to gauge public opinion and predict its impact.

Introduction

> Fear is an automatic response in all of us to threats to our deepest of all inbred propensities, our will to live. It is also the basis of many of our economic responses, the risk aversion that limits our willingness to invest and to trade, especially far from home, and that, in the extreme, induces us to disengage from markets, precipitating a severe falloff of economic activity. (Greenspan 2007, p. 17)

The stock market usually reflects business fundamentals, such as corporate earnings. However, we also see many events that seem rooted in human emotion more than anything else, from "irrational exuberance" during booms to panicked sell-offs during busts. It's not uncommon to see people even extend these emotions to the whole market: a recent Wall Street Journal headline read "Recession Worry Seizes the Day and Dow." A deep thread of laboratory research documents the link between a person's emotional state and the choices they make (Dolan 2002; Zajonc 1980), particularly their investment decisions (Lerner, Small and Loewenstein 2004; Lerner and Kelter 2001; Loewenstein, et al. 2001; Shiv, et al. 2005). For our purposes, one major finding stands out: fear makes people risk-averse. Still, this thread of research comes from the lab. How do real world emotions affect real world markets, like the stock market?

In this paper, we take a step toward answering this question. From a dataset of over 20 million LiveJournal posts, we construct a metric of anxiety, worry and fear called the Anxiety Index. The Anxiety Index is built on the judgements of two linguistic classifiers trained on a LiveJournal mood corpus from 2004. The major finding of this paper is that the Anxiety Index has information about future stock market prices not already apparent from market data. We demonstrate this result using an econometric technique called Granger causality. In particular, we show that the Anxiety Index has novel information about the S&P 500 index over 174 trading days in 2008, roughly 70% of the trading year. We estimate that a one standard deviation rise in the Anxiety Index corresponds to S&P 500 returns 0.4% lower than otherwise expected.

This finding is not as farfetched as it may first appear. In a 2007 paper, using Granger-causal methods, Paul Tetlock demonstrated that pessimism expressed in a high-profile Wall Street Journal column had novel information about Dow returns from 1984 to 1987. Nice, sunny weather even explains some stock market movements (Hirshleifer and Shumway 2003). Google search queries have predictive information about diseases and consumer spending (Choi and Varian 2009). Blog posts and blog sentiment have been shown to predict product sales (Gruhl, et al. 2005; Mishne and Glance 2006) and to correspond to certain high profile events (Balog, Mishne and de Rijke 2006). Previous authors have drawn comparisons between internet message boards and individual stock prices, with mixed results (De Choudhury, et al. 2008; Tumarkin and Whitelaw 2001). To the best of our knowledge, however, this is the first work documenting a clear (Granger-causal) link between web-based social data and a broad stock market indicator like the S&P. In many ways, the present work resembles an updated version of the Consumer Confidence Index: a broad, forward-looking barometer of worry. Our results show how the mood of millions in a large online community, even one that primarily discusses daily life, can anticipate changes in a seemingly unrelated system. Along

Copyright © 2009, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

with other recent work (Choi and Varian 2009; Lazer, et al. 2009) it suggests new economic forecasting techniques.

We begin by describing our LiveJournal blog dataset and our S&P 500 dataset, laying out how we constructed the Anxiety Index from the decisions of two classifiers. Next, we present the Granger-causal statistical analysis on which we base our findings. We also present the results of a Monte Carlo simulation confirming this paper's major result. The paper concludes by discussing the work's limitations (including our skepticism towards trading on it), its economic significance and where future work might lead.

Data

We now present the two datasets which form the core of this paper: a blog dataset from 2008 and an S&P 500 dataset. The blog dataset consists of the stream of all LiveJournal posts during three periods of 2008: January 25 to June 13, August 1 to September 30, and November 3 to December 18. We collected the first and last periods (Jan. 25 – Jun. 13 & Nov. 3 – Dec. 18) ourselves by listening to the LiveJournal Atom stream[2]. To compensate for the gap between these two periods, we augmented it with every LiveJournal post from the ICWSM 2010 Data Challenge dataset, which uses the same sampling method. We would have preferred a complete record of 2008, but this dataset turns out to suffice for the approach we take in this paper. In total, it comprises 20,110,390 full-text posts from LiveJournal. We made no attempt to filter the posts or the bloggers in any way. (It seemed difficult or perhaps impossible to meaningfully do so a priori; more on this topic later.)

Studying emotions by dissecting blog posts has its disadvantages. For one, we likely do not get a representative sample. Phone surveys like the Consumer Confidence Index achieve a roughly representative sample by random digit dialing. We can make no such claims. For instance, bloggers are likely younger than non-bloggers (Lenhart and Fox 2006). However, there are clear advantages too. This technique eliminates experimenter effects and produces a nearly continuous source of data. It also seems possible that bloggers not only speak for their own emotions, but also for people close to them (Fowler and Christakis 2008).

It may seem strange that we chose to study only LiveJournal. Why not include other blogging sites as well, such as Blogger, WordPress, etc.? Or Twitter? However, a single site study has certain advantages. For instance, it sidesteps the different norms and demographics that develop in different sites, a problem that could confound our analysis. Furthermore, LiveJournal in particular has three distinct advantages. As one of the web's earliest blogging platforms, it has a large, firmly established community, but is no longer a web darling. LiveJournal is also known as a journaling site, a place where people record their personal thoughts and daily lives (Herring, et al. 2004; Kumar, et al. 2004; Lenhart and Fox 2006). At present, it seems hard to make similar claims regarding Twitter, for instance. Perhaps most importantly, LiveJournal has a history of coupling posts with moods, something we use as we construct the Anxiety Index.

The Anxiety Index

From our blog dataset we derive the Anxiety Index, a measure of aggregate anxiety and worry across all of LiveJournal. Following in the footsteps of Mishne and de Rijke (2006) or Facebook's Gross National Happiness[3], we want to compute a LiveJournal-wide index of mood. We might do this by measuring posts which LiveJournal's users tag with certain moods. For instance, a user can tag a post with a mood like *happy* or *silly*. Unfortunately, this only constitutes a small fraction of LiveJournal's posts; most do not come with moods attached. We would prefer to use every post we see, and therefore generate more robust estimates.

Borrowing a corpus of 624,905 mood-annotated LiveJournal posts from 2004 (Balog, Mishne and de Rijke 2006), we extracted the 12,923 posts users tagged as *anxious*, *worried*, *nervous* or *fearful* (roughly 2% of the corpus). We then trained two classifiers to distinguish between these *anxious* posts and a random sample of *not anxious* posts (a proportional mix of the other 128 possible moods, including *happy*, *angry*, *confused*, *relaxed*, etc.). The first classifier (*C1*), a boosted decision tree (Freund and Schapire 1995), uses the most informative 100 word stems as features (ranked by information gain). For example, *C1* identified the following words (their stems, more precisely) as *anxious* indicators: "nervous," "scared," "interview" and "hospital." Important words indicating *not anxious* included "yay," "awesome" and "love." The second classifier (*C2*), built on a bagged Complement Naive Bayes algorithm (Rennie, et al. 2003), compensates for the limited vocabulary of the first. It uses 46,438 words from the 2004 corpus as its feature set. There is no definitive line between *anxious* and *not anxious* (e.g., "upset" might indicate *anxious* or *sad*). So, the classifiers encountered significant noise during training. Under 10-fold cross-validation, both classifiers correctly identify an *anxious* post only 28% and 32% of the time, respectively. However, they each have low false positive rates, labeling a *not anxious* post as *anxious* only 3% and 6% of the time, respectively. Clearly, these are conservative classifiers with noise. However, we care about anxiety in the aggregate, as it varies in time. It seems reasonable to us that the noise will end up uniformly distributed in time.

Admittedly, questions remain about this method's construct validity. Do *C1* and *C2* truly identify anxious, worried and fearful blog posts? Other researchers have opted

---

[2] http://atom.services.livejournal.com/atom-stream.xml

[3] http://apps.facebook.com/usa_gnh

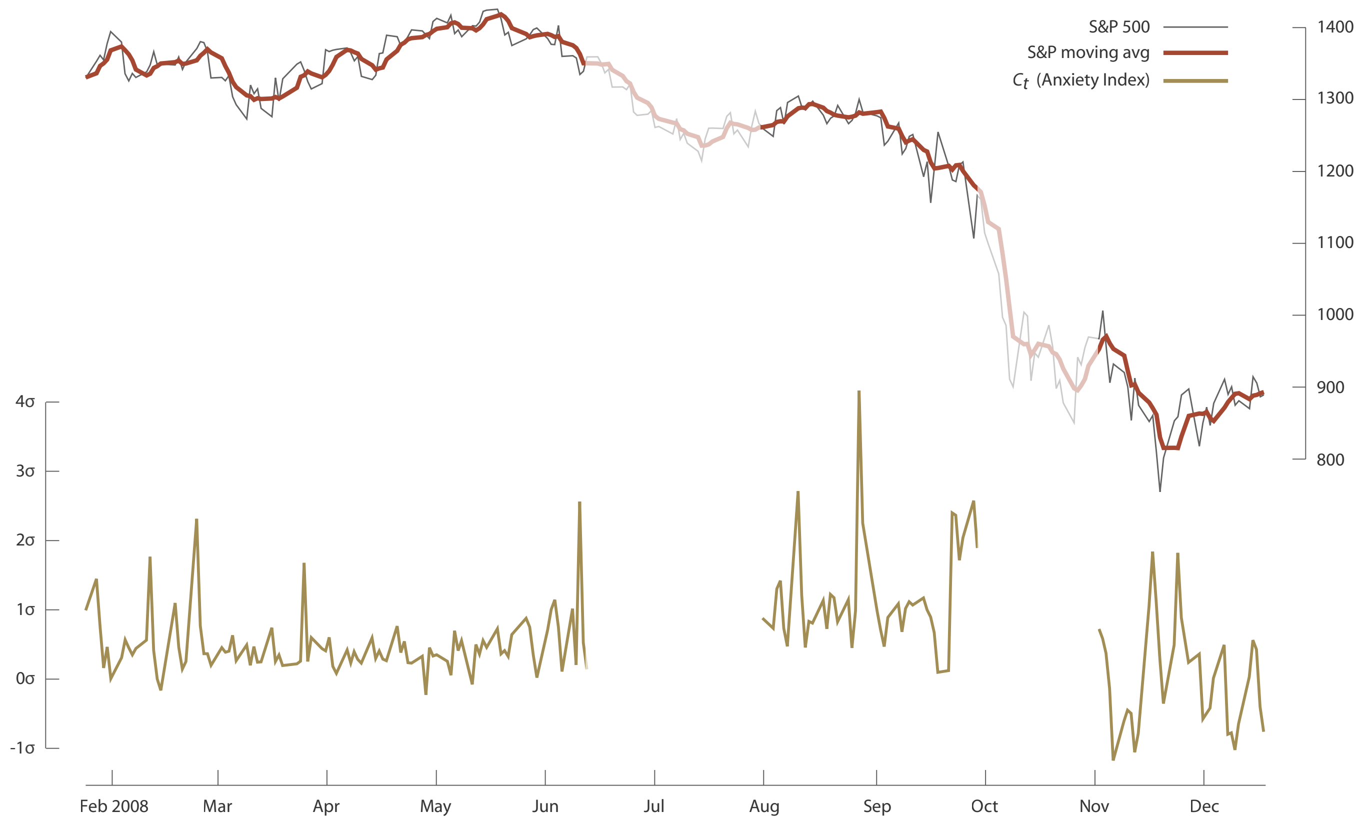

Figure 1. The computational analysis of blog text for signs of anxiety, worry and fear, plotted against the S&P 500. At bottom, C_t is the proportion of blog posts classified as *anxious* on a given trading day (in each classifier's original standardized units). The Anxiety Index is the first difference of this line. At top, the S&P 500 index over the same period, including a five-day smoothed version to help readers see trends. We faded out the S&P 500 over the two periods during which we do not have Anxiety Index data.

for calibrated and vetted dictionary-based methods (Dodds and Danforth 2009; Hancock, Landrigan and Silver 2007; Tetlock 2007). We perhaps could have made firmer validity claims had we chosen one of these tools, such as LIWC or the General Inquirer. One reason we chose classification was the specific emotion we wanted to target; anxiety and worry are not typically well-represented in affective dictionaries. For now, we rest our validity claims on the judgements of the original 2004 bloggers (who chose the mood labels) and the historical performance of the algorithms we employ. We also point out that *C1* and *C2* enjoy domain-specific training sets, meaning that they trained on LiveJournal posts (albeit from a different time) and classify LiveJournal posts.

Because we ultimately want to compare the Anxiety Index to the stock market, for which we have daily closing prices, there is a frequency problem: we need to align daily market prices to the potentially high frequency data reported by our classifiers. Let $C1_t$ (and $C2_t$) be the standardized proportion of posts classified by *C1* (and *C2*) as *anxious* between the close of trading day t-1 and the close of trading day t. This straightforward mapping has one wrinkle: trading day $t-1$ and trading day t are sometimes separated by many actual days, such as across weekends and intervening holidays. In these cases, we let $C1_t$ (and $C2_t$) correspond to the highest proportion recorded during the intervening days, where each day is treated as if it were a trading day. Other methods, such as averaging across the intervening days, seemed to unduly punish big events occurring on a weekend, for instance. We also experimented with Agresti and Coull's (1998) method for adjusting proportions, but it made no difference at this scale.

From the relatively conservative classifiers *C1* and *C2*, we define a slightly more liberal metric $C_t = \max(C1_t, C2_t)$. (Combining the two classifiers via a higher level ensemble algorithm would have been another approach to producing C_t.) Figure 1 plots C_t against the S&P 500. For reasons of stationarity we make clear in the next section, we define the Anxiety Index to be the first difference of logged C_t, $A_t = \log(C_{t+1}) - \log(C_t)$. (Logging stabilizes variance and improves normality; also, we were careful not to difference across breaks in our dataset.) The Anxiety Index has values for 174 trading days in 2008.

Market Data

We use the S&P 500 index as a broad stock market indicator, and obtained its daily closing prices from Yahoo! Finance. As is commonplace (Marschinski and Kantz 2002), we examine the S&P 500 via its *log-returns*, $R_t =$

M_t	Coeff.	Std. Err.	t	p
-1 day	-0.858	0.110	-7.83	<0.001†
-2 days	-0.655	0.095	-6.88	<0.001†
-3 days	-0.171	0.098	-1.73	0.085
VOL_t				
-1 day	0.149	0.076	1.96	0.051
-2 days	0.152	0.087	1.74	0.084
-3 days	0.176	0.117	1.51	0.132
VLM_t				
-1 day	0.054	0.080	0.68	0.497
-2 days	0.132	0.074	1.78	0.077
-3 days	0.067	0.058	1.16	0.247
Summary	**DW**	**Adj. R^2**	**$F_{9,161}$**	**p**
	2.054	0.509	20.58	<0.001†

Table 1. M1's coefficients and summary statistics. We report robust standard errors (for M2 also) to account for remaining heteroscedasticity. *DW* refers to the Durbin-Watson statistic, a measure of residual autocorrelation.

$\log(SP_{t+1}) - \log(SP_t)$, where SP_t is the closing price of the S&P 500 on trading day t. Based on a view of a limited dataset from 2007, we hypothesized that the Anxiety Index may inform the broad direction of the market, but not necessarily where it will go tomorrow (i.e., log-returns). To this end, and to meet the requirements of stationarity, in this paper our primary S&P time series is the first difference of log-returns, $M_t = R_{t+1} - R_t$, a kind of stock market acceleration metric. (Again, we were careful not to difference across the breaks in our dataset.)

However, the stock market is not just its prices. Following the lead of earlier researchers (Tetlock 2007), we include the first difference of logged trading volume, VLM_t, and the first difference of volatility, VOL_t. Squared log-returns serve as a proxy for volatility; that is, $VOL_t = R_{t+1} \cdot R_{t+1} - R_t \cdot R_t$. (In the two places where market data extends across the breaks in our data, the faded regions in Fig. 1, we calculate it from that data.)

2008 was not a typical stock market year. It started near its ten-year high and ended near its ten-year low. We had no idea what were getting into when we started collecting data. The non-representative nature of 2008 could endanger the generalizability of the findings we present next. We cover this more fully at the end of the paper, in the Limitations section.

M_t	Coeff.	Std. Err.	t	p
-1 day	-0.872	0.105	-8.36	<0.001†
-2 days	-0.667	0.085	-7.81	<0.001†
-3 days	-0.182	0.087	-2.08	0.039†
VOL_t				
-1 day	0.182	0.063	2.90	0.004†
-2 days	0.184	0.083	2.22	0.028†
-3 days	0.177	0.109	1.63	0.105
VLM_t				
-1 day	0.056	0.078	0.712	0.477
-2 days	0.133	0.071	1.89	0.061
-3 days	0.071	0.059	1.20	0.232
A_t				
-1 day	-0.064	0.052	-1.24	0.216
-2 days	-0.128	0.060	-2.14	0.034†
-3 days	-0.136	0.084	-1.62	0.107
Summary	**DW**	**Adj. R^2**	**$F_{12,158}$**	**p**
	2.032	0.527	16.77	<0.001†
Granger causality		**$F_{3,158}$**	**p**	**MC p**
		3.01	0.032†	0.045†

Table 2. M2's coefficients and summary statistics. We also summarize Granger causality results here. As expected, high anxiety negatively affects the market. *MC p* refers to the p-value obtained via Monte Carlo simulation.

Granger-causal Analysis

We now formally test the relationship between the anxiety, fear and worry expressed by millions on the Web and the stock market. We apply the econometric technique of Granger causality (Granger 1969). Informally, it asks whether the Anxiety Index provides useful information for projecting future market prices not already contained in the market itself. More formally, we compare the variance explained by two linear models: one that explains M_t wholly endogenously, and another that builds on the first but includes the Anxiety Index. Although the technique has the word "causal" in it, we clearly aren't testing true causation. We can only say whether one time series has information about another. The first model (M1) uses only lagged values of market data to predict M_t, including lagged values of M_t itself. The second model (M2) adopts M1's predictors but also includes lagged values of the Anxiety Index.

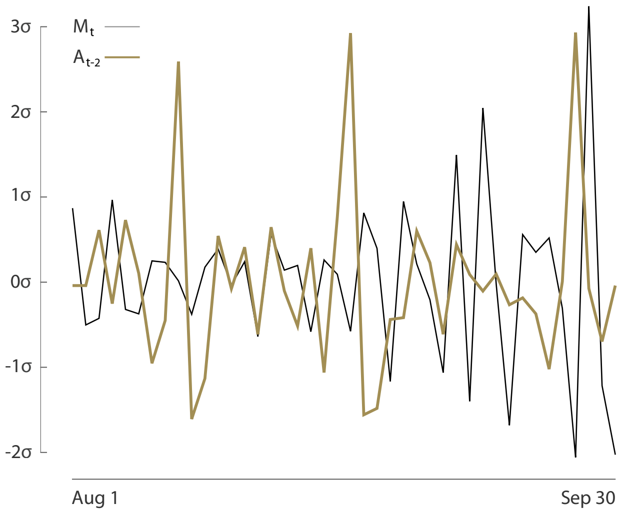

Figure 2. A view of M2, presented in Table 2, during August and September, 2008. The first difference of returns (black) is plotted alongside the Anxiety Index lagged by two days (gold). Diamond patterns illustrate the roughly negative relationship between the two.

$$\textbf{M1}: \quad M_t = \alpha + \sum_{i=1}^{3} \beta_i M_{t-i} + \sum_{i=1}^{3} \gamma_i VOL_{t-i} + \sum_{i=1}^{3} \delta_i VLM_{t-i} + \epsilon_t$$

$$\textbf{M2}: \quad M_t = \alpha + \sum_{i=1}^{3} \beta_i M_{t-i} + \sum_{i=1}^{3} \gamma_i VOL_{t-i} + \sum_{i=1}^{3} \delta_i VLM_{t-i} + \sum_{i=1}^{3} \lambda_i A_{t-i} + \epsilon_t$$

The lag parameter, 3 here, is free in Granger-causal models. We chose 3 because adding more lagged predictors does not significantly improve M1's performance, while at the same time it respects the relatively small number of samples in our dataset (174). The equations are estimated by OLS. Granger causality is particularly sensitive to non-stationary time series, whose statistical properties vary with time (Granger and Newbold 1974). However, the differencing measures explained earlier make all four times series stationary; the largest Kwiatkowski-Phillips-Schmidt-Shin (KPSS) statistic is 0.0299, $p > 0.1$. Granger causality models information flow between time series linearly. This is a limitation. Any relationship between the Anxiety Index and the S&P is almost certainly nonlinear. Recently (Schreiber 2000), techniques have emerged to model nonlinear information flow between time series. However, the statistical properties of these new techniques are not well-understood, especially significance testing (Marschinski and Kantz 2002). As we are most interested in confirming or rejecting the relationship, not necessarily modeling it optimally, we chose to apply Granger causality—a standard technique.

The major result of this paper is that M2 performs significantly better than M1, $F_{3,158} = 3.01$, $p = 0.0322$. The F-statistic here is proportional to the relative change in the residual sum of squares between models. Table 2 includes this result along with a summary of M2. Figure 2 provides a graphical view into the model. The Durbin-Watson statistic corresponding to M2's residuals indicates no serious autocorrelation problems, $DW = 2.032$, $p = 0.649$. (Deeper lags are also uncorrelated.) As expected, high anxiety negatively affects the market. The negative standardized coefficients for A_t indicate that anxiety slows a market climb and accelerates a drop. Due to remaining heteroscedasticity in the model, we report robust standard errors for the coefficient estimates. For instance, A_{t-3} has a significant ordinary coefficient, $t = -2.385$, $p = 0.0183$, but it is only marginally significant using robust standard errors. While the residuals are non-normal, Cook's distances do not indicate that outliers severely distort the regression: the two largest Cook's distances are 0.524 and 0.22.

Monte Carlo Confirmation

Despite attempts to normalize each time series and stabilize variance, the models exhibit heteroscedasticity and have non-normal residual distributions. Linear regression is resilient against these violations, in moderation. Still, since M1 and M2 violate these two assumptions, is there reason to doubt the Granger causality finding? For instance, various deviations can cause the "F-statistic" to drift from the F-distribution, distorting p-values. Or, it seems possible that differencing could create high frequency oscillations which, when lagged, make it easy to find Granger causality. So, we performed Monte Carlo simulations to confirm the F-statistic and p-value reported earlier.

Suppose we knew the Anxiety Index's underlying distribution. The approach we take in our simulations is to draw a new Anxiety Index from this distribution, essentially scattering a new, lookalike Anxiety Index in time. (More precisely, we draw a new C_t, logging, differencing and lagging it to form A_{t-1}, A_{t-2} and A_{t-3}.) Next, we compute the F-statistic for the Granger-causal comparison between M1 and M2, using this new Anxiety Index. Repeating this process, we generate a list of F-statistics. We can count the number of times we see an F-statistic as large or larger than the one we report above, $F = 3.01$. The ratio of this number to the size of the list is an experimentally generated p-value, the probability of obtaining these results due to chance alone. As a proxy for the underlying Anxiety Index distribution, we use a Gaussian kernel density estimate. After performing one million of these simulations, we obtain an experimental $p = 0.0453$, confirming the statistical significance of this paper's main result. Figure 3 presents a visual depiction of two iterations in this process. (We find that the experimental list of F-statistics does in fact deviate from the F-distribution, Kolmogorov-Smirnov $D = 0.0337$, $p < 0.001$.) Monte Carlo simulation inflates the p-value slightly. We offer the following (most likely incomplete) explanation: violations account for some of the inflation, as does differencing.

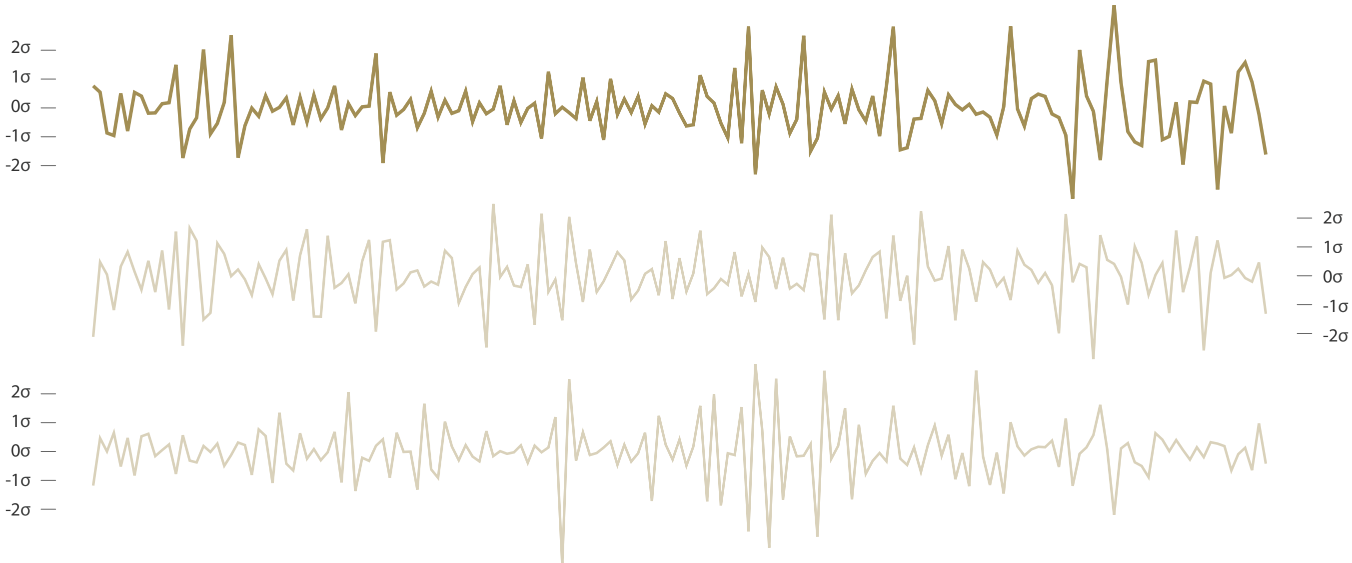

Figure 3. The real Anxiety Index (top) and two doppelgänger Anxiety Indices (bottom) generated during Monte Carlo simulation. Using a kernel density estimate, we generated new Anxiety Indices to experimentally test this paper's main result. Testing for Granger causality with one million new Anxiety Indices confirms the result, but inflates the *p*-value slightly, from 0.0322 to 0.0453.

Reverse Granger Causality

If, as these results suggest, the Anxiety Index informs future movements of the stock market, it is natural to ask the question in reverse. Does the market's recent past give us predictive information about the Anxiety Index? Anecdotally at least, the market itself is often a significant source of anxiety for many of us—certainly in 2008. Swapping the roles of M_t and A_t in M1 and M2, however, we do not find evidence that M_t Granger-causes A_t, $F_{3,158} = 1.08$, $p = 0.36$. The only significant predictor (aside from the very significant lagged A_t) is VLM_{t-2}, $t = -2.409$, $p = 0.0172$. This seems to suggest that the Anxiety Index is reacting to something other than the market, or at least that the market is only one among many inputs driving A_t. This is markedly different from the results reported by Tetlock (2007), in which past market values significantly drove media pessimism.

The VIX

The Chicago Board Options Exchange produces an index called the VIX, often commonly referred to as the "investor fear gauge" (Whaley 2000). Given its common name, many have suggested that we compare the Anxiety Index with the VIX. It measures the expected future volatility of the stock market and is constructed wholly from financial data. Do the Anxiety Index results still hold when controlling for the VIX?

Let VIX_t be the first difference of the logged VIX on trading day t. Adding three lagged VIX_t predictors to the models M1 and M2, we can estimate the Anxiety Index's predictive information after controlling for the VIX. Adding the VIX does reduce the predictive power of the Anxiety Index, $F_{3,155} = 2.058$, $p = 0.108$. Perhaps this is to be expected: the VIX and the Anxiety Index measure (at least marginally) related concepts. In fact, this could be a positive sign for the Anxiety Index's construct validity.

However, the VIX does not completely obliterate the Anxiety Index's predictive value: one Anxiety Index predictor has a significant OLS coefficient estimate in the VIX formulation, $p = 0.0436$, and another borders on significance, $p = 0.0714$. These results suggest that the Anxiety Index offers some of the same information as the VIX, and perhaps some relevant information not contained in the VIX. This is the most we can say on the subject given the size of our current dataset, but it seems like a rich area for future study.

Limitations

2008 is an outlier in the history of the stock market. Although our data precedes the true market crash by many months, it remains unclear whether these results will generalize to more "normal" times. For instance, we think it is entirely possible that 2008 itself acted as the filter we left out. In other words, financial and economic events often made top headlines in the mainstream press. It seems reasonable to conclude that the blogosphere also disproportionately reflected these events. Perhaps in normal times, a pop culture event like Michael Jackson's death utterly swamps the Anxiety Index. For example, the Anxiety Index has a blip coinciding with Valentine's Day. Maybe in normal times this is a towering event—but not in 2008.

When we talk about this work, the typical first reaction is, "So how much money have you made?" In truth, it remains unclear whether any clever trading strategy can derive from the Anxiety Index. While at first glance it appears that such a strategy exists, we point to the VIX as a complication. Perhaps, given Tetlock's (2007) index, the VIX and every other related indicator, the Anxiety Index is rendered ineffectual. On the other hand, proponents of the efficient market hypothesis might argue that those other indicators are already priced in. Or maybe the Anxiety Index only works in extraordinary times. Even the publication of this very paper may nullify its impact. For now, we can only express a desire to see more work on these questions.

Conclusions and Future Work

This paper statistically shows that a broad index of mood from an online community has novel predictive information about the stock market. The anomalous gyrations of the stock market have been explained by network effects (Ivkovic and Weisbenner 2007; Scherer and Cho 2003) (i.e., listening to your neighbor's stock tips), media pessimism (Tetlock 2007), the weather (Hirshleifer and Shumway 2003), and even the outcomes of soccer games (Edmans, Garcia and Norli 2007). Here, we identify a Granger-causal relationship between the stock market and an algorithmic estimate of the mood of millions of people. Specifically, a one standard deviation increase in the Anxiety Index corresponds to 0.4% lower returns (actual returns, not log-returns). This is a much bigger decrease than the 0.07% reported by Tetlock (2007) and the 0.022% reported by Hirshleifer and Shumway (2003). We think this big jump is due to a combination of the broadness of the Anxiety Index and the extraordinary market swings we saw during 2008. The present work makes three main contributions beyond previous research. First, we tap the emotional state of a vast group of people without needing a proxy, like the mainstream media. Second, we demonstrate that this method provides timely information about the stock market. Third, this all happens without human intervention, relying solely on algorithmic techniques.

After the demonstration of new predictive information, the next most interesting part of this work is that we made no attempt to filter bloggers or what they write about. This is remarkably different than previous attempts in this domain (Tetlock 2007; Rumarkin and Whitelaw 2001). At first glance, it seems strange that LiveJournal should provide any information at all. Investing experts sell investor sentiment reports. What on earth could LiveJournal know about the stock market? Certainly, this is one of the most fascinating pieces of the present work. We would argue that one of the reasons the Anxiety Index has *novel* information is because it derives from somewhere other than financial circles. It seems altogether likely that the market has already priced-in information from financial circles—but perhaps not from elsewhere, like the social web.

Many have suggested that we constrain our data to those related to financial topics. We steered away from this approach because we found it difficult to find the dividing line: Does a post about a flood count? If that flood is in a small Iowa town, maybe not. If it's Katrina, then perhaps so. What about political posts? Some data in the Anxiety Index may constitute unrelated blips; witness the spike around Valentine's Day in Figure 1. On the other hand, these data could reflect a deep concern that will manifest in the market. It seemed impossible to know a priori. However, filtering may represent one of the most profitable lines of future work. Consider reinforcement learning. Sources that add distinctively predictive information to the Anxiety Index now could receive more weight in the future. Of course, this would have to be done with care and give credit to random chance. If anxiety and worry have new predictive information about the market, then what about other emotions? This also looks like a very interesting and productive line of future work. For instance, does an increase in anxiety coupled with a decrease in happiness have more predictive information than the Anxiety Index alone? These are provocative questions.

The findings presented here have important implications for economic forecasting and, more broadly, for predicting the impact of public opinion. As Google (Choi and Varian 2009) and others (Gruhl, et al. 2005) have shown, we can draw important and timely conclusions about the data people leave behind on the web. Introducing the Anxiety Index (and related indices) into economic models may reduce the need for costly phone surveys and offer ways to identify inflection points in real time.

Acknowledgements

This work was supported by funding from Microsoft and Google. We thank Richard Sowers, David Hirshleifer, David Ikenberry, Evangelos Benos, Martha Pollack, Marek Jochec and the entire Social Spaces group for early feedback on this work.

References

Agresti, A. and Coull, B. 1998. Approximate Is Better Than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician* 52:119-126.

Balog, K., Mishne, G., de Rijke, M. 2006. Why Are They Excited? Identifying and Explaining Spikes in Blog Mood Levels. In *11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL)*.

Choi, H. and Varian, H. 2009. Predicting the Present with Google Trends. Google Technical Report.

Dolan, R. 2002. Emotion, Cognition, and Behavior. *Science* 298: 1191–1194.

De Choudhury, M., et al. 2008. Can Blog Communication Dynamics Be Correlated With Stock Market Activity? In *Proceedings of the Nineteenth ACM conference on Hypertext and Hypermedia (HT)*, 55–60.

Dodds, P. and Danforth, C. 2009. Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. *Journal of Happiness Studies* 10.

Edmans, A., Garcia, D. and Norli, O. 2007. Sports Sentiment and Stock Returns. *The Journal of Finance* 62: 1967–1998.

Fowler, J. and Christakis, N. 2008. Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 years in the Framingham Heart Study. *British Medical Journal* 337: a2338+.

Freund, Y. and Schapire, R. 1995. A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting. In *Proceedings of the Second European Conference on Computational Learning Theory (EuroCOLT)*, 23–37.

Granger, C. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37: 424–438.

Granger, C. and Newbold, P. 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2: 111–120.

Greenspan, A. 2007. *The Age of Turbulence: Adventures in a New World*. London: Penguin Books.

D. Gruhl. et al. 2005. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (KDD)*, 78–87.

Hancock, J., Landrigan, C., and Silver, C. 2007. Expressing Emotion in Text-based Communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI), 929–932.

Herring, S., et al. 2004. Bridging the Gap: A Genre Analysis of Weblogs. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS)*.

Hirshleifer, D. and Shumway, T. 2003. Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance* 58: 1009–1032.

Ivkovic, Z. and Weisbenner, S. 2007. Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices. *Review of Financial Studies* 20: 1327–1357.

Kumar, R., et al. 2004. Structure and Evolution of Blogspace. *Communications of the ACM* 47: 35–39.

Lazer, D., et al. 2009. Computational Social Science. *Science* 323: 721–723.

Lenhart, A. and Fox, S. 2006. Bloggers: A Portrait of the Internet's New Storytellers. Technical report. Pew Internet & American Life Project.

Lerner, J. and Keltner, D. 2001. Fear, Anger, and Risk. *Journal of Personality and Social Psychology* 81: 146–159.

Lerner, J., Small, D. and Loewenstein, G. 2004. Heart Strings and Purse Strings. *Psychological Science* 15: 337–341.

Loewenstein, G., et al. 2001. Risk as Feelings. *Psychological Bulletin* 127: 267–286.

Marschinski, R. and Kantz, H. 2002. Analysing the Information Flow Between Financial Time Series. *European Physical Journal B* 30: 275–281.

Mishne, G. and de Rijke, M. 2006. Capturing Global Mood Levels using Blog Posts. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*.

Mishne, G. and Glance, N. 2006. Predicting Movie Sales from Blogger Sentiment. In *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*.

Rennie, J., et al. 2003. Tackling the Poor Assumptions of Naive Bayes Text Classifiers. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*.

Scherer, C. and Cho, H. 2003. A Social Network Contagion Theory of Risk Perception. *Risk Analysis* 23: 261-267.

Schreiber, T. 2000. Measuring Information Transfer. *Physical Review Letters* 85: 461–464.

Shiv, B., et al. 2005. Investment Behavior and the Negative Side of Emotion. *Psychological Science* 16: 435–439.

Tetlock, P. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62: 1139–1168.

Tumarkin, R., Whitelaw, R. 2001. News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal* 57: 41+.

Whaley, R. 2000. The Investor Fear Gauge. *The Journal of Portfolio Management* 26: 12–17.

Zajonc, R. 1980. Feeling and Thinking: Preferences Need No Inferences. *American Psychologist* 35: 151–175.