Does truth help sentiment diffuse? Evidence from UK investment mandates

James Bowden and William Forbes

August 10, 2020

There is a surge of interest in whether sentiments/moods contained within the words of those trading assets predict those assets future value. A cottage industry of studies examine the ability of market sentiment, embedded in discussions on Twitter and elsewhere (see Boudoukh et al. (2019), Bowden, Burton, and Power (2018), Nekrasov, Hong, and Wu (2020)). Here we ask if such sentiment, inferred from the expressed view of traders and other financial professionals, is likely to be a helpful signal of value? Or will a focus on such measures just add to so much existing ‘noise’ in financial markets (De Long, Summers, and Waldmann 1990). Baker and Wurgler (2007) (p 129) define investment sentiment as a:

“belief about future cash flows and investment risk that are not justified by the facts at hand.”

Here the focus is on the “noisy”, perhaps giddy, element of sentiment that lies “beyond the numbers” as Kay and King (2020) express it. But we equally all know of cases where “gut feeling” is either all we have, or has served us well, (Gigerenzer 2007): in choosing friends, finding a good bar, or a good concert to go to. Often a sentiment/gut-feel over something, becomes confirmed as something much deeper, once all the data is in. For example, I cannot read all Russian literature, or even all good Russian literature. So I may choose to read those works written by the most famous - or recognisable - authors such as Tolstoy or Dostoevsky. Choosing these authors is an example of applying the “recognition heuristic” of “going with what you know” (Goldstein and Gigerenzer 2002). Such heuristics works very well it seems, not least in the context of choosing stock market portfolio ((Borges et al. 1999), (Huberman 2001),(Ortmann et al. 2008)).

When we say “it’s just a gut-feeling”, we should therefore not underestimate how powerful such feelings/sentiments can be. But clearly we cannot rely on our intuitions in an uncritical way; we must instead endeavour to “trust then verify”. We might question when we should trust our instinct in investing, and when we should not? We ask this of our investment sentiment measure, which we derive from investor mandates, in this paper.

We do this by asking whether true/predictive market sentiment has a greater influence than false sentiments, as measured using later market movements? This is inspired by the recent paper by Vosoughi, Roy, and Aral (2018) who report that “fake news” on Twitter spreads at least as quickly as real news. This make sense because we may seek out evidence which confirms our existing beliefs, or perhaps we simply enjoy a “good story” if it sounds feasible, even if it is not.

In other words, regardless of the correlation between currently expressed market sentiment and current stock market movements, is sentiment more influential in predicting future stock market movements when it is correct? For us, sentiment will therefore be a “true” guide to investment when positive sentiments are followed by equity market increases and negative sentiments are followed by equity market decreases.

1 Data and Methods.

Our data comes from a London based consultancy that seeks to obtain investors’ mandates or market perceptions, agreed on the phone, in order to match these mandates with available funds seeking investor capital. Given the breadth of the universe of funds available, this is quite a task. Our dataset contains written summaries of 2,721 calls taken during the years 2014-2019.\(^1\) Figure 1a shows the distribution of data across

\(^1\)A smaller sample of records go back to the consultancy’s formation in 2011. These call notes were ommitted from our analysis due to the low number of records per year.
our sample years. Each year before 2019 in our sample generates at least 400 calls. A small amount of data comes from 2019, although the data will be updated/expanded as our research hopefully progresses.

Here we focus on contacts located in Europe. Figure 1b gives the distribution of client calls for the ten most popular locations. After London, it is surprising to find that Lichfield, a Cathedral City in the Midlands, is the second most popular source of calls. Edinburgh, the UK’s second financial capital, also features heavily. However, a majority of the firm’s clients are London-based, and thus when we measure sentiment we measure it amongst investors largely in London. Figure 2 maps client addresses by postcode, and demonstrates that only a small number of calls originating from continental Europe are included in the dataset, with UK clients constituting the overwhelming majority of call records.

Figure 1: Distribution of Call Transcripts

(a) By Year

(b) By Client Location

(c) By Job Title

(d) By Call Manager

Representatives of client firms that engage in calls with the consultancy represent a wide range of financial roles, with the distribution of those roles accounting for more than 50 calls shown in Figure 1c. By some margin, the most common job title listed in the call notes is 'Investment Analyst', followed by 'Fund of Funds Manager' and 'Head of Public Equities', respectively. This offers some insight into who calls the consultancy, and we might also be interested in who receives these calls, and subsequently deals with the client’s request. The raw data spreadsheet does name the person who took the call, but this field is largely empty. So here we use another field, which indicates who modified or oversaw each call.

Figure 1d presents the distribution of those overseeing more than 50 calls within the consultancy. It is noted that one person oversaw about a 966 of the listed calls, and a second over 600; this suggests a large
amount of concentration in the management of how calls were recorded. This suggests the possibility that management of the style, content and recording of client calls within the dataset is largely consistent.

1.1 Identifying Trends in Investor Mandates

We next run various tests on the call notes to identify some preliminary themes and trends in the data. Namely, we use measures of centrality and connectedness to identify (i) geographical locations, assets and investment strategies (i.e. “entities”) that are most frequently mentioned in investor call notes, and (ii) which entities that are often referenced in conjunction with each other; that is, they are more likely to be discussed in the same call note. The four largest clusters are represented as network graphs, which we present in Figures 3a to 3d. Large nodes represented in this network reflect those entities with the highest number of mentions, while those entities closest to the centre of the network demonstrate the highest level of centrality (or influence) amongst all other entities in the network.

To illustrate the usefulness of this form of visualisation, imagine that a client of the consultancy declares an interest in funds focussed on long-only strategies and (at least) one BRIC economy.\(^2\) It can be seen from Figure 3a that, historically, clients expressing similar interests tended to mention other emerging countries in their “search for yield” - such as Brazil and Vietnam - while an even larger number of clients in such calls mention Japan. Further interesting trends appear in the form of experimental asset classes (such as ‘Bitcoin’ and ‘GreenTech’) and investment strategies (such as ‘Quant’).

In Figure 3c, it is clear (and expected) that those calls mentioning Britain also make reference to “Brexit”, as either an opportunity or a threat.\(^3\)

Comparing the two maps of conversation flows in Figures 3b and 3c, it may suggest that those investors with a UK base, or investment focus, are also far more likely to be looking to the US - in the form of east cost

\(^2\)Brazil, Russia, India and China

\(^3\)An extension of this study could be to identify the nature of sentiment towards these specific entities over time, to infer whether investors do see significant political, economic and societal events as an opportunity, a cause for concern, or neither.
Figure 3: Network Graph Showing Entity Clusters Identified in Client Call Notes

(a) Cluster One: Asian and Emerging Economies
(b) Cluster Two: United States
(c) Cluster Three: United Kingdom
(d) Cluster Four: London
technology firms based in San Francisco, or through discussion of influential investors such as Warren Buffet - than the other way round. Conversely those who discuss the United States tend not to be so geographically varied in their mention of other countries. Instead the discussions tend to focus on regions of the United States, such as Chicago and Miami, or tax havens like the British Virgin Islands or the Bahamas.

It should be noted that associations between asset–classes and countries strengthen and weaken over time. While this section focussed on historical trends aggregated over a five–year period (2014-2019), similar analyses could be conducted on an annual (or even monthly) basis, to identify changes in investment trends and strategies over time. Further, this analysis could be combined with the sentiment measures employed in the remainder of this section to monitor changing attitudes towards these entity clusters.

1.2 Measuring sentiment

The power of text-based sentiment analysis methods in predicting future abnormal returns remains a live issue. Articles that identify high associations between text sentiment and “mood states”, such as Bollen and Zeng (2011), are countered by others declaring no evidence to suggest any such relationship (Das and Chen 2007). However, a majority of the literature generally supports the notion that textual sentiment has contemporaneous or short-term effects on abnormal returns (Kearney and Liu 2014), although such sentiment measures seldom anticipate financial movements exceeding five per cent (Nardo and Naltsidis 2016).

This consensus is also supported by the increasing influence of sentiment–driven algorithms in current financial markets: after an initially limited uptake of machine learning techniques (such as sentiment analysis), more recently a Barclay Hedge (2018) survey found that two-thirds of hedge funds now leverage AI or machine learning to generate trading ideas.

Interest in the textual analysis of investors (clients) views and beliefs is often inspired by the belief that “sentiment” embedded in such discussions is predictive of asset price movements. In other words, such feelings are “beyond the numbers” (Kay and King 2020). However, textual measures are the not the only - or even necessarily the best - measures of market sentiment. Baker and Wurgler (2006) for example present an alternative measure of investor sentiment, and we might ask to what degree text-based measures of sentiment add value to these?

1.3 Pattern in sentiment of over time

For sentiment – as derived from the records of investor call notes – to be a significant asset pricing factor it must possess predictive (and ideally causal) power-logic. Further, it must add value to other reasonable measures of market sentiment already discussed in research literature (such as Baker and Wurgler (2006)). These are central issues we address in this paper.

We begin by plotting monthly call note sentiment across the months of our sample time period, from January 2014 to September 2019. Figure 4a plots the sentiment time-series, and shows very large deviations in sentiment, especially in the months from late 2018 continuing into mid-2019. We see an initial spike in sentiment upwards in late 2017/early 2018, which is dissipated in September 2018, and reappears in early 2019. Given the sharp monthly volatility in average sentiment, derived from our text-based sentiment measure, we also plot a three month moving-average of sentiment scores in Figure 4b. As can be seen, this transformation makes very little difference to the time-series pattern of our text-based sentiment measure. We therefore use the 3-month rolling-average mean of sentiment expressed in our analysis, and report differences when they arise.

But not all months within our sample have equally reliable estimates of average clients. This is primarily because, in some months, clients demonstrate a tendency to express similar sentiments, whereas in other months clients express contrasting sentiments. To investigate this possibility we also examine the time-series variation in the standard deviation of sentiment within each month. This plot is presented in Figure 4c.

Here again we see how unusual the period of mid-2018 and late 2019 are. In both these periods, as well as in August 2017, we see a spike upwards in the dispersion of expressed sentiment within the data. We cannot

---

4 a hedge fund based solely on the premise of Bollen and Zeng (2011) produced a positive return of 1.86 percent in its first month, before being closed down due to a lack of investor interest.

5 We feel that this time horizon works well, in that it occurs before the Coronavirus pandemic wreaked havoc across the world, and thus excludes any effects of this event.

6 based on the textual data in the ‘description’ field of our raw data
derive a measure of the standard deviation for January to April 2019, or in July 2019, reflecting the paucity of data we have for that year currently. Finally, in Figure 4d we construct the coefficient of variation of our sentiment, text based, metric, or the ratio of its standard deviation to its mean, \( \frac{\sigma}{\mu} \). This confirms mid-2017, late 2018 and late 2019 as the most volatile periods for expressed investor sentiment in our data set.

1.4 Do current asset pricing reflect expressed market sentiment?

An efficient market reflects publicly available, value relevant, information (Fama 1970). So any empirical test of value relevance is a test of a particular asset pricing model and the value relevance of the information advanced for investor use. A wide range of literature already confirms the value relevance of investor sentiment for asset pricing ((Fisher and Statman 2000) and (Baker and Wurgler 2007) review the early literature). Here we confine ourselves initially to univariate tests of the relation between current returns and the sentiment currently expressed in our textual data, drawn from the description of the call.

We begin by calculating the Spearman rank correlation between the average sentiment expressed by clients in call notes and various measures of equity market performance in the UK and US. Results, reported in Figure 5a, suggest significant correlations with the index level of the UK FTSE 100 and the US S&P 500, but much weaker correlations with stock market returns. Indeed the correlation with current returns on the S&P 500 is negative, if statistically insignificant. Only the correlation with the level of the US S&P
500 is statistically significant, and even then only when a 90% confidence interval is used. So while some conditionally does appear to be present here it would be unwise to exaggerate it.

To uncover the extent of the relationship between our textual sentiment measure and stock market movements we repeat the analysis of Figure 5a in Figure 5b using a three–month moving–average of recorded textual sentiments expressed rather the monthly measure of Figure 5a. We do this in the belief that lasting emotional dispositions, persisting over some (here, three) months, have a more consistent impact on stock markets; in the same way as impulse purchases are less important drivers of aggregate consumption/GDP than standing orders (to utilities, mortgage payments, newspapers/ sports clubs subscriptions, etc.).

So we now examine the relationship between various market indices and a three-month moving–average of our text-based sentiment measure. We find that the Spearman rank correlations are improved and the perverse correlation between our sentiment measure and the return on the S&P 500 disappears. In this reformulation of how sentiment moves markets the implied Spearman rank correlations are higher, with the level of the FTSE 100 and S&P 500 index as well being statistically significant at the 95% confidence interval (which academics seem to hold in, perhaps strange, esteem as a benchmark for results worthy of discussion/publication).

However, the correlation with equity market returns returns – both in the US and the UK – remains small and statistically insignificant. This is unfortunate as most asset pricing models we have, such as the CAPM and Fama-French three-factor model, are all stated in terms of (expected) shareholder returns and not asset price levels ((Sharpe 1964), (Fama and French 1993)).

**Figure 5: Correlations between Levels of Call Note Sentiment and Index Levels/Returns**

(a) Sentiment Index

<table>
<thead>
<tr>
<th>FTSE100 level</th>
<th>FTSE100 return</th>
<th>S&amp;P500 level</th>
<th>S&amp;P500 return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman rank correlation</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

(b) 3-Month Rolling Average Sentiment Index

Plot of Sentiment 3-month moving average in description field of Murano data January 2014 to November 2019

Sentiment Index 3-month moving average

<table>
<thead>
<tr>
<th>FTSE100 level</th>
<th>FTSE100 return</th>
<th>S&amp;P500 level</th>
<th>S&amp;P500 return</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
<td>0.15</td>
</tr>
</tbody>
</table>

An approximate randomisation of the relationship between the sentiment metric and the level of the FTSE 100

The previous section identified that the correlation between the level and the FTSE 100 Index and sentiment is large and statistically significant. But we might wonder what this means exactly? One way to envisage this is by imagining us repeatedly shuffling the pack of our data points on observed client sentiment, and constructing a “pseudo” Spearman rank statistic measuring how well our shuffled sentiment data is correlated with the level of the FTSE 100 index.

For example, suppose we took the first observation of our text-based sentiment measure, and replaced it with the 77th observation. Suppose we then replaced the 77th observation with the sixth observation, and so on, to obtain a new vector of sentiment measure data. The original Spearman rank correlation between our implied market sentiment, averaged over three months, and the level of the FTSE 100 was 0.255. How many times can we beat/match this correlation purely by chance using shuffled data? One way to know this is to recalculate the Spearman rank correlation test many times (here 10,000 times) for different
shuffled reconstructions of our sentiment index data, and see how many match/exceed 0.255. This is what
an approximate randomisation test does in essence (Noreen 1989).

Figure 6 presents the results of our test. Some 180 of the 10,000 pseudo Spearman rank replications
match, or exceed, our Spearman rank correlation value of the unshuffled/“true” data of 0.255, so we have a
$1 - \frac{180}{10,000} = 98.2\%$ confidence that the original correlation did not get produced merely by chance.

1.5 Do changes in expressed market sentiment drive current equity returns?

In an attempt to capture movements in returns we also examine the relationship between changes in expressed
client sentiment and contemporary equity returns. The results of this analysis are presented in Figure 7. Here
the correlations are low, statistically insignificant, and - in the case of US equity returns - negative, suggesting
that a decline in US markets raises the sentiments of UK investors. This deters us from modelling shareholder
returns in the manner typically done by standard asset pricing models.

Figure 6: Randomised Spearman Rank Test

Figure 7: Correlations between Changes in Call Note Sentiment and Index Levels/Returns
2 Can our sentiment index form the basis of a profitable trading rule?

At this stage, an investor may reasonably suggest: "this is all very interesting, but show me the money!". Our text-based sentiment index is fairly strongly and significantly correlated with the level of the FTSE 100 index, but can that finding be integrated into a profitable trading strategy? We attempt to answer that question in this section.

We begin with a simple trading rule which buys FTSE 100 index data when our textual measure of sentiment is "high" and sells that same measure when the sentiment metric is "low". So a question arises: what is high and what is low?

We begin by forming a portfolio using a simple rule to buy the FTSE 100 index when the three-month moving average sentiment index lies above its mean value of 0.136, and selling the FTSE 100 index when the three-month moving average of our sentiment index lies below its mean value of 0.136. This trading strategy is simple in nature, but does require continuous trading.

Figure 8a shows the pattern of trading return, while Figure 8b shows the net profit over the buy and sell portfolio. First, looking at Figure 8a we see that, from mid-2017, low sentiment months have quite good current returns on the FTSE 100 index. The current returns on the FTSE 100 index are low and frequently negative throughout our sample period. The strong current returns to low sentiment stocks drive the losses from this sentiment-based trading strategy reported in Figure 8b.

Can mean sentiment predict future FTSE 100 returns?

Perhaps the issue is that it takes time for market sentiment to be impounded into FTSE 100 returns? So using current sentiment to predict current returns might not necessarily yield meaningful results. Our performance metric is a three-month moving average of sentiment, so part of the current return embeds knowledge of the first two months of our three–month sentiment metric. But to give us confidence that this is not the cause of our trading rule’s poor performance, we repeat our tests by seeing how a three–month average of past performance predicts next month’s returns.

Figure 9a shows the high and low sentiment portfolio returns if we buy/sell the FTSE 100 index in the month following abnormal positive/negative sentiment, and the cumulative profit to this strategy is given in Figure 9b. The mean cumulative return following this strategy is -0.107, which suggests that little has changed by introducing a lag to the sentiment metric. Comparing Figures 8a and 8b with Figures 9a and 9b certainly confirms this view. So, based on our preliminary study of the FTSE 100 index, a trading system
based on buying/selling when sentiment is higher/lower than it’s three-month mean value appears to have little attraction for investors.

Figure 9: FTSE100 Index Returns Generated by a Lagged Sentiment-Driven Portfolio

(a) High/Low Sentiment Portfolios

(b) Long/Short Portfolio

3 Modelling sentiment and stock markets levels

If we can’t find a model of investor returns relation to expressed sentiment in our data how can we model the relationship between sentiment and market levels? Once we have such a model what could it mean?

One answer comes from the idea that our sentiment index and the level of the FTSE 100 are co-integrated time–series: being integrated of order one/random-walk series themselves, yet forming a stable, equilibrium, relationship as a pair (Engle and Granger 1987). In other words, sentiment and the level of the market might be out of kilter for a while, but tend to come into balance in the longer/medium term. Looking at the superimposed plot of the two time-series in Figure 10 this at least looks possible, if not likely, to be the case.

Co-integrated time–series that are individually integrated of order one and form a pair integrated of order zero, will drift towards one another over time. So while the elements of the co–integrated pair are individually random-walk processes their combination forms a regression which yields residuals of order zero, or swiftly reverting to a mean value.

Therefore, the two random-walk like variables form an error-correcting pair that are stable in their shared evolution over time. More formally we estimate a regression of the form ((Wooldridge 2013), p 625):

\[
\Delta \text{FTSE}_t = \alpha_0 + \gamma_0 \text{SENT}_t + \delta (\text{FTSE}_{t-1} - \beta \text{SENT}_{t-1}) + \epsilon_t
\]

where the term \(\delta (\text{FTSE}_{t-1} - \text{SENT}_{t-1})\) denotes the “error-correction-term” deriving from a prior regression of the FTSE 100 index (FTSE) on our text based sentiment error (SENT). So we might obtain the error-correction term, \(\delta (\text{FTSE}_{t-1} - \text{SENT}_{t-1})\), in the above error correction mechanism, of equation 1, by retrieving the residual from a regression of the FTSE 100 on our sentiment measure, and lagging it by one month.

Using a standard Dickey-Fuller test (Dickey and Fuller 1979) for the stationarity, mean–reverting nature of our two time series, it turns out that while the FTSE 100 monthly index is a random walk our sentiment measure is not. So sentiment by its nature mean–reverts. This might make sense to us: after the elation of victory, comes the tedium of the “new normal”, even if this is an improvement on our past life. Undeterred, we nevertheless form the implied error–correction mechanism of equation 1 above.

The results of our estimation of an error–correction representation of the relationship between the FTSE 100 and our market sentiment measure is given in Table 1. We see that the resulting error-correction representation is both statistically insignificant, with the coefficient on the lagged residuals from the level regression being significant at the 90% confidence interval, but very close to zero.
The level of investor sentiment and the FTSE 100 in our data

Table 1: Error correction Mechanism regression for FTSE100 index and sentiment as expressed in the description field of our data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-value</th>
<th>R-squared</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td>0.275</td>
<td>0.05</td>
<td>63</td>
</tr>
<tr>
<td>ΔSentiment</td>
<td>-0.003</td>
<td>-0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals_{t-1}</td>
<td>0.00001</td>
<td>1.77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Discussion

What have we learned from this tentative examination of the predictability of textual measures of sentiment for market movements? Certainly that at a market level we are not optimistic about its predictive power or its ability to drive a reliable trading rule. But we know in trading that "ripeness is all" and, because an aggregate relationship is not found in the current study, this does not necessarily mean that a relationship between sentiment measures and market movements never exists. More data over longer time-series and more clients should reveal this.

If text-based sentiment measures can drive profitable trading, then it will be interesting to see when these strategies work and fail. Thus, we need to trace an ecology of sentiment-versus value-based (earnings, dividends, sales) trading strategies. Further we need to understand more about the time horizon, client group, and sentiment based measures that demonstrate the greatest associations when aggregated.

Are clients depressed by the poor performance of emerging markets at the same as those primarily invested in US firms, or do these sentiments occur at different – even diametrically opposed – times? If good news for emerging market investors is bad news for US investors, this would explain the poor performance of our aggregate trading rule. We defer discussion of such issues to a time when we have more data on investor sentiment, derived from our investment mandate service data.
References


