Investor sentiment and asset returns: the case of Indian stock market

Sachin Mathur* and Anupam Rastogi

NMIMS, Mumbai,
V.L. Mehta Road, Vile Parle, West,
Mumbai, Pin Code 400056, India
Email: sachinmath@yahoo.com
Email: anupam.rastogi@nmims.edu
*Corresponding author

Abstract: According to behavioural finance theory, investor sentiment can lead to extreme mispricing of stocks. In this paper we describe the construction of an investor sentiment index for India, an emerging market, to examine the association between sentiment and stock returns. We test whether the sentiment index predicts long-term returns of the stock market, as well as of stock portfolios formed on the basis of size and value characteristics, over the sample period of 2004 to 2016. The sentiment index fails to predict broad market returns, but is inversely associated with the subsequent year’s returns of small low-priced stocks, consistent with constrained arbitrage argument of behavioural finance theory and information uncertainty associated with such stocks.

Keywords: investor sentiment; sentiment index; behavioural finance; stock returns; information uncertainty; emerging market; India.


Biographical notes: Sachin Mathur is an Associate Professor in Finance Area at the Narsee Monjee Institute of Management Studies, Mumbai. His areas of research include asset pricing, investment analysis and behavioural finance. He is a CFA Charter Holder and earned his PhD from the Narsee Monjee Institute of Management Studies, Mumbai.

Anupam Rastogi is working as a Senior Professor in Finance area at the Narsee Monjee Institute of Management Studies, Mumbai. His areas of research include strategic corporate finance, project finance, infrastructure development and finance, and emerging market economies. He has published extensively on infrastructure financing, economic modelling, capital markets and exchange rate mechanisms. He has been a consultant to the Asian Development Bank. He holds an MSc in Economics from the London School of Economics and PhD from the University of Liverpool.
1 Introduction

Examining the role of investor sentiment in forming asset bubbles is of interest to investors as well as the regulator. This is underlined by the episodes of crash of 1987 in the US stock market and the boom and bust cycle in internet related stocks during 1999–2002 across markets. These episodes provide anecdotal evidence that investor sentiment may lead to extreme mispricing. Since investor sentiment is assumed to be cyclical, it follows that periods of high sentiment would be associated with low returns in the subsequent periods, when sentiment swings downward and vice versa.

Predictability of long term returns by itself is not anomalous. In classical finance, the long horizon predictability of return has been explained away as rational variations over time (Fama, 1991). Behavioural finance proponents on the other hand argue that persistent mispricing is caused by behavioural biases and limits to arbitrage (Barberis and Thaler, 2003). As explained by Fama (1991), it is difficult to distinguish between irrational bubbles and rational time-varying returns caused by temporary changes in discount rates. For example, price multiples have been known to predict long-term returns, but since their numerator is price and not value, they may represent not only rational valuation but even mispricing. This presents a problem for researchers of isolating sentiment by finding predictors that are based on investor sentiment, but do not represent known sources of systematic risk.

Behavioural models of investor sentiment and arbitrage constraints provide insights into how sentiment becomes apparent as anomalies and hence how to observe the effect of sentiment over time periods or in a cross-section of stocks. For instance, Shleifer and Vishny (1997) argue that limits to arbitrage exist, since arbitrage requires capital (such as borrowed funds for short sales) and is not riskless. Even when informed arbitrage traders are able to identify mispricing, they are unable to correct mispricing if it is extreme, driven by uninformed noise trading. Since noise trading would be sentiment-induced and not information driven, the behavioural model of Shleifer and Vishny (1997) implies that sentiment can result in extreme mispricing in the presence of arbitrage constraints. A key insight is that the effect of sentiment will be higher where there are greater arbitrage constraints, for example, due to high volatility.

Further, sentiment-based explanations of anomalies such as discount on closed-end funds (Lee et al., 1996) and pricing of initial public offerings (IPOs) in hot issues market (Ljungqvist and Wilhelm, 2003) have laid the ground for identifying measures of sentiment. Since sentiment by itself is latent, its effect as reflected by occurrence of such anomalies can provide effective proxy measures. For example, we can measure sentiment indirectly by measuring IPO price premium or discount on closed-end funds.

Motivated by the behavioural theories and sentiment based explanations of anomalies, there have been methodological developments in identifying measures of sentiment and construction of sentiment index. One alternative is to use a single reliable indicator after suitable modifications and refinements to remove the systematic risk effects and noise. For example, Ben-Rephael et al. (2012) measure sentiment index using mutual fund flows. Da et al. (2015) build a sentiment index based on internet searches of
economic terms that reflected fear after detailed filtering and testing the association of these terms with market returns. Alrabadi (2015) uses order imbalance as a proxy for investor sentiment. Lemmon and Portniaguina (2006) use consumer confidence as a measure of investor optimism.

The other alternative is to use multiple proxy variables and extract the common variation as the sentiment index. Both Brown and Cliff (2004, 2005) and Baker and Wurgler (2006, 2007) use several indicators and derive the sentiment index as the first principal component of these indicators. Since specific variables may also represent aspects other than sentiment, for instance trading turnover can represent both sentiment and liquidity, taking the common variation from multiple variables mitigates such specific effects.

These sentiment indices have enabled research into the nature of relationship between investor sentiment and stock returns. By using their index, Brown and Cliff (2004) show that there is a contemporaneous relationship between investor sentiment and stock returns. Similarly, Baker and Wurgler (2006) establish that the effect of sentiment is stronger in the case of smaller and volatile segment of stocks, where arbitrage trading was more difficult. Lemmon and Portniaguina (2006) find that investor sentiments predict returns for small stocks.

In this study we focus on the relationship between stock returns and investor sentiment in the Indian stock market. The Indian market has much lower retail investor participation than developed markets. This has implications for the effect of retail investor sentiments on stock prices and hence returns. Further there is lack of trading depth or liquidity in small stocks owing to which there is limited institutional investor interest in small stocks. Hence it is probable that the relationship between sentiment indicators and stock returns would differ in the case of India from the other markets. Moreover, it would be reasonable to expect a more pronounced difference in the relationship between retail investor sentiments and stock returns between large and small stock segments, due to the liquidity concerns.

We construct an index of investor sentiment in India based on common variations between sentiment proxies. Five proxy variables are used to form this index: the number of IPOs, mutual funds flows, advances to declines ratio, concentration of trading turnover and a sentiment sub-index based on internet search history. We also build an analogous index of sentiment changes based on common variations based on changes in these variables.

We want to understand, firstly, whether investor sentiment predicts future returns and over what horizon. Secondly, we want to understand whether the sentiment effects differ by firm size and valuation. Answers to these questions would indicate the extent to which investors and policy makers can rely on an indicator on investor sentiment for predicting future returns or identifying potential stock market boom and bust cycles. Consistent with these research questions, we test three hypotheses relating investor sentiments to stock returns: first, that the sentiment affects short-term returns; second, that sentiment predicts long-term returns with an inverse association; and, third, that the effect of sentiment should be higher for smaller stocks (those with low market capitalisation) and value stocks (those with low market to book value).

These hypotheses are based on the behavioural proposition that sentiment-induced trading can move stock prices due to arbitrage constraints. While the first hypothesis is a restatement of this proposition, the second hypothesis follows as a consequence. If sentiment-induced trading moves stock prices away from their fair values, prices would
tend to mean revert over a medium to long horizon (one to three years). The third hypothesis follows from the expectation that arbitrage trading would be difficult for more volatile stocks.

We find that the sentiment index correlates well with stock prices and sentiment changes index correlates well with stock returns. However, in the short-term the direction of causality is from stock returns to sentiment and not from sentiment to returns.

We examine whether the sentiment index is inversely associated with subsequent 12 months’ market returns. The expected inverse relationship between sentiment index and future returns is modest and not statistically significant. In comparison, the price to earnings multiple predicts subsequent 12 months’ returns. Taken together, the evidence suggests that the stock price cycles during the study period can be explained more by a combination of changes in discount rates along with sentiment than by sentiment alone.

We also test the association for portfolios of stocks sorted by size and value characteristics. The subsequent 12 months returns are negatively associated with sentiment in the case of small value firms. This is because these stocks are associated with greater information uncertainty, hence with higher price volatility, and therefore with greater arbitrage constraints.

2 Data and sample

Data for sentiment indicators and stock indices is sourced from the websites of Securities and Exchange Board of India (SEBI), http://www.sebi.gov.in, the Bombay Stock Exchange (BSE), www.bseindia.com, and National Stock Exchange (NSE), http://www.nseindia.com. Data for mutual fund flows (MFF) is obtained from the website of Association of Mutual Funds in India (AMFI), www.amfiindia.com. Business cycle indicators were sourced from the website of Reserve Bank of India (RBI), http://www.rbi.org.in. The index of Google searches is sourced from http://www.google.co.in/trends/.

The length of time series of data was limited for several variables and the data frequency for most of the variables is monthly. Hence, based on the common sample, we use the monthly time series from April 2004 to March 2016 for constructing the sentiment index, which provides 144 months of observations.

Estimates for market returns and returns on portfolios are based on updated database of estimates of stock market returns ($R_m$) and Fama-French factors appended with Agarwalla et al. (2003) and are available on http://www.iimahd.ernet.in/~iffm/Indian-Fama-French-Momentum/. These estimates are based on a large sample of more than 5000 firms listed on BSE. In comparison, the broadest market indices (nifty 500 or BSE S&P 500) cover less than 10% of the listed universe, while the size sub-indices are correspondingly narrower and have limited history.

3 Construction of sentiment index

3.1 Variable identification and selection

The identification and selection process for proxy variables for sentiment is as follows. First, we list all the proxy variables of sentiment as identified in research literature. From
the list we eliminate variables for which sufficient long and reliable data series are not readily available, variables which are also significantly associated with known sources of systematic risk, or those which could be affected by structural changes in the stock market.

As done by Brown and Cliff (2004), we group the select proxy variables in categories. The first category relates to the flow of funds. We consider four proxy variables in this category, the share of equity issues in total issues, the number of IPOs, first day return on IPOs and mutual fund flows. The share of equity issues indicator is constrained by a less developed market for corporate debt in India which affects the denominator, making the ratio noisy and less meaningful. The first day premium on IPO indicator is affected by several months of nil IPO activity. Only the number of IPOs and mutual fund flows are retained in the final index from this category.

The second category consists of price-based indicators. Based on the available information we consider two indicators, the number of advances to declines and the difference between implied and realised volatility. The data series on implied volatility is available only from April 2008 onwards, resulting in its elimination from our list. However the advance to decline ratio is retained.

The third category consists of volume-based indicators. We find trading turnover, estimated as the ratio of trading volumes to outstanding shares, declining unevenly owing to significant growth in the derivatives segment, and having a low correlation with other sentiment indicators after adjusting for the declining trend. We instead choose concentration of turnover, estimated as the share of top 50 shares in the total turnover as a more suitable indicator.

The fourth category covers consumer surveys and searches. We do not find a suitable historical monthly index of consumer confidence for India. Instead we use the index of internet searches, noting the high frequency with which this data is available.

3.2 Estimation of proxy variables

The selection process results in five proxy variables as described below.

Number of initial public offerings (IPON): the number of IPOs is considered as a sentiment indicator because previous research shows that managements time their IPOs based on investor sentiment. IPON is used as a sentiment indicator by Brown and Cliff (2004) and Baker and Wurgler (2006). We estimate IPON as log (1 + Number of IPOs), one being added to account for nil values prior to the log transformation. IPOs listed on the small and medium enterprises (SME) platform are excluded since they are not of comparable size. For the sentiment change index we use the monthly change in IPON (CIPON) as the indicator.

MFF: MFF is estimated as the net monthly purchases less redemptions or repurchases of equity mutual funds. Hybrid (equity and debt) funds and tax-incentive based funds are not included in this measure. For the sentiment changes index we find the changes in MFF itself very noisy and therefore use changes in increments in monthly assets under management held in equity mutual funds or CMFF as the indicator, which is found to be better correlated with other proxy variables of sentiment change.

Advances to declines ratio (AD): AD is the ratio of the number stocks whose prices have increased to the number of stocks whose prices have declined. It is also a commonly used technical indicator. We use the monthly figures for advances and declines as
In order to ensure that level index captures persistence in advances or declines, we estimate AD as the log of six month moving average of advances to six month moving average of declines on the BSE. However, for estimating the corresponding index of sentiment change we use the log of the monthly AD ratio, CAD as change indicator.

Concentration in trading turnover (CTT): we use the share of top 50 securities in the total turnover as reported by NSE, the stock exchange having the largest monthly trading volumes in India. The investor interest in smaller securities is expected to increase during periods of high sentiment, but fall during periods of low sentiment. The concentration of top 50 securities in turnover is therefore expected to vary inversely with investor sentiment. To convert the bounded percentages into an unbounded figure, the concentration ratio is logit transformed, so that CTT is measured as the log of the ratio of share of top 50 securities to the share of the remaining securities. Change in CTT (CCTT) is used for the sentiment changes index.

Index of internet searches for sentiment related terms (SEARCH): SEARCH is estimated as the average of selected investor sentiment related terms. Historical indices of searches on Google search engine for each search term are available on a weekly basis from 2004 onwards. Following Da et al. (2015), we select search terms that have strong statistical association with market returns. We use the levels of search data for each search term to estimate the SEARCH variable and the changes in monthly search data to estimate an analogous CSEARCH variable.

For this purpose, we prepare a list of 50 search terms related to investments, personal finances or the economy. From the searches data, two sets of estimates are made for each search term. The first set is based on monthly average search levels for each term. The second set is based on weekly changes in search index for each term. We winsorise, deseasonalise and standardise each series to have a mean of zero and variance of one. From the initial list of 50 search terms, we select those terms whose weekly changes have a statistically significant correlation with index returns. This process results in 11 search terms, including ‘share prices’, ‘mutual funds’, ‘stocks’, ‘money’, ‘dividend’ and ‘savings’ that have positive correlation with nifty 50 returns, and ‘nifty’, ‘economy’, ‘economic slowdown’, ‘recession’ and ‘crisis’ that have negative correlation with nifty 50 returns. SEARCH is estimated as a simple average of the signed monthly search levels, the sign for each term being based on the sign of their correlation with nifty 50 index. For the changes index, a proxy CSEARCH is estimated by the same method, averaging the signed monthly changes in search data series.

Common variation in the five proxy variables as described above is used to construct the sentiment index. Prior to deriving the index, each proxy variable is detrended and then made orthogonal to business cycle indicators by obtaining regression residuals with year-on-year growth in index of industrial production, the yield on 91-day treasury bill and term structure of interest rates. These business cycle indicators are chosen from the list of macro-economic indicators, which according to Chen et al. (1986) are related to expected returns, and for which reliable monthly data is available in India.

Table 1 presents the correlation matrix between the transformed proxy variables. Most of the correlations are statistically significant but moderate, indicating that none of the proxies is redundant.
Table 1  Correlation between proxy variables

<table>
<thead>
<tr>
<th></th>
<th>IPON</th>
<th>MFF</th>
<th>AD</th>
<th>CTT</th>
<th>SEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPON</td>
<td></td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFF</td>
<td>0.13</td>
<td></td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
<td>0.46</td>
<td>0.25</td>
</tr>
<tr>
<td>CTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.57</td>
</tr>
<tr>
<td>SEARCH</td>
<td>0.39</td>
<td>0.36</td>
<td>0.50</td>
<td></td>
<td>-0.50</td>
</tr>
<tr>
<td>SENT</td>
<td>0.52</td>
<td>0.32</td>
<td>0.74</td>
<td>-0.86</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Note: Shaded cells represent statistically significant correlations.

3.3 Estimation of SENT and SENTCH indices

The sentiment index (SENT) is derived as the first principal component of five proxy variables, using standardised values of these variables. The equation for SENT is as follows:

\[ \text{SENT} = 0.39\text{IPON} + 0.32\text{MFF} + 0.45\text{AD} - 0.53\text{CTT} + 0.51\text{SEARCH} \]  

The first principal component contributes to 50% of the cumulative variance and all the variables in the equation have the expected signs. It also satisfies the heuristic of being the only component with eigen value greater than one.

The analogous index of sentiment change (SENTCH) is similarly derived from changes in the five proxy variables. The equation for SENTCH is as follows:

\[ \text{SENTCH} = 0.31\text{CIPON} + 0.35\text{CMFF} + 0.52\text{CAD} - 0.52\text{CCTT} + 0.49\text{CSEARCH} \]  

The first principal component in this case contributes to 47% of the cumulative variance and signs of all the variables are theoretically consistent. Though the second component also has an eigen value of 1.04 and added an incremental 21% to the cumulative variance, we are unable to ascribe any economic interpretation to it due to factor loadings that were either low or having incorrect sign. Therefore we retain only the first component even in the case of SENTCH.

The time series characteristics of SENT and SENTCH are represented graphically in Figure 1. The SENT time series is found to have statistically significant positive serial correlations up to four months lag, while SENTCH has a significant negative serial correlation only at the second lag (ignoring higher order lags beyond six months). Tests using alternative ARMA models show that SENT is best represented (using Bayesian information criterion) by a first-order autoregressive process, while SENTCH is best represented as an ARMA (0, 0) process. Finally both SENT and SENTCH are found to be stationary using augmented Dickey-Fuller test (the null hypothesis of presence of a unit root is rejected at p < 0.01). The tests of unit root are rejected even when structural breakpoints are taken into consideration based on Perron (1997).
Investor sentiment and asset returns

Figure 1  Time series of SENT and SENTCH indices

4  Short-term relationship of sentiment indices with market returns

Figure 2 shows the trend for the SENT and the five proxy variables in six panels. In each panel the local high and low points of nifty 50 index have also been marked. The high and low points are one standard deviation higher and lower respectively than the mean detrended values of the index. Detrending of the index is done by subtracting past moving average value of the past 12 months. Since CTT is proposed to have a negative relationship with sentiment, the trend line for CTT is based on negative of CTT values. From the panels, it can be seen that SENT has captured most of the low and high phases of the stock index in India during the sample period.

SENT has a contemporaneous Pearson’s correlation coefficient of 0.38 with detrended index, but negligible correlation with nifty returns (rho = –0.02). This is not surprising since the stock index returns are expected to be correlated to sentiment changes rather than to the sentiment levels.

Figure 3 shows the scatterplot of SENTCH with returns on nifty 50 indexes. There is fair degree of correlation between SENTCH and nifty 50 index returns (Pearson’s rho = 0.44).

In order to probe causality, we test a VAR (vector auto-regression) model with stock market return ($R_m$) and SENT as covariates. Since the correlation of $R_m$ is stronger with SENTCH, we also test a second VAR model with $R_m$ and SENTCH as covariates. The two VAR models are tested with variables in levels, as SENT, SENTCH and $R_m$ are all found to be stationary. For deciding the number of lags we use the Akaike information criterion (AIC). The AIC suggests a lag of one period in the first case and two periods in the second case.

Tables 2 and 3 show the results of estimation of the two VAR models. From the results it is apparent that sentiment level does not predict near-term market returns. In the case of sentiment change, market return shows a weak relationship with SENTCH at one period lag. However, one-period lagged market return explains both SENT and SENTCH. Further, one-period lagged SENT predicts itself with a positive sign, suggesting persistence. On the other hand, one-period lagged SENTCH also predicts itself but with a negative sign, which suggests reversion towards mean value.
Figure 2  Trend of SENT index and proxy variables

Table 2  Results of VAR model: SENT and market returns

<table>
<thead>
<tr>
<th></th>
<th>SENT</th>
<th>( R_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{NT,t-1} )</td>
<td>0.677</td>
<td>−0.007</td>
</tr>
<tr>
<td>t-statistics</td>
<td>13.02**</td>
<td>−1.13</td>
</tr>
<tr>
<td>( R_{m,t-1} )</td>
<td>4.655</td>
<td>0.107</td>
</tr>
<tr>
<td>t-statistics</td>
<td>6.91**</td>
<td>1.27</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.046</td>
<td>0.010</td>
</tr>
<tr>
<td>t-statistics</td>
<td>−0.88</td>
<td>1.56</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.617</td>
<td>0.005</td>
</tr>
<tr>
<td>F-statistics</td>
<td>115.20**</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Notes: **p < 0.01; *p < 0.05; ~p < 0.10. The columns represent the dependent variables and the rows represent the independent variables of the VAR estimation equations.
The results are further confirmed using the Granger causality test. The test fails to reject the hypothesis that SENT does not the Granger cause $R_m$ (p-value of 0.26) and rejects the hypothesis that SENTCH Granger causes $R_m$, but only at a p-value of 0.09. On the other hand $R_m$ is found to the Granger cause both SENT and SENTCH (p-value < 0.01), again showing that the direction of causality is from market returns to sentiment and not the other way round.

Figure 3 Scatterplot of SENTCH with nifty 50 index (see online version for colours)

5 Long-term return predictability

5.1 Two-way analysis

We analyse the association of SENT with subsequent 12 month returns for overall market, as well as for stocks sorted by size and value. For this purpose we categorise the months when SENT is greater than one standard deviation above its mean (zero) as high sentiment and months when SENT is lesser than one standard deviation below zero as low sentiment. Table 3 presents the two-way analysis of firms sorted by size and value as well as by sentiment.

From the analysis three patterns are evident. First, the average returns are high subsequent to low sentiment periods and vice versa across portfolios and horizons. Second, the differences in returns based on sentiment are higher for small firms than for large firms. Third, the differences increase when moving from growth to value firms.

The greater impact of sentiment in the case of small stocks and value stocks can be explained with reference to the model of Shleifer and Vishny (1997). The model predicts that it would be difficult for rational traders to arbitrage away noise trade based mispricing in segments where there are greater risks to arbitrage trading, for instance highly volatile stocks. Thus we expect the impact of sentiment to be greater in the case of
stocks where volatility is higher. Figure 4 shows that volatility in Indian stocks indeed tends to increase inversely with size and directly with market to book value.

Table 3  Results of VAR model: SENC H and market returns

<table>
<thead>
<tr>
<th></th>
<th>SENTCH</th>
<th>Rm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENTCHt–1</td>
<td>–0.286</td>
<td>–0.017</td>
</tr>
<tr>
<td>t-statistics</td>
<td>–2.93**</td>
<td>–1.93~</td>
</tr>
<tr>
<td>SENTCHt–2</td>
<td>–0.151</td>
<td>–0.013</td>
</tr>
<tr>
<td>t-statistics</td>
<td>–1.79~</td>
<td>–1.65</td>
</tr>
<tr>
<td>Rm,t–1</td>
<td>8.093</td>
<td>0.193</td>
</tr>
<tr>
<td>t-statistics</td>
<td>7.66**</td>
<td>2.00*</td>
</tr>
<tr>
<td>Rm,t–2</td>
<td>–0.504</td>
<td>0.146</td>
</tr>
<tr>
<td>t-statistics</td>
<td>–0.39</td>
<td>1.25</td>
</tr>
<tr>
<td>Constant</td>
<td>–0.087</td>
<td>0.891</td>
</tr>
<tr>
<td>t-statistics</td>
<td>–1.21</td>
<td>1.35</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.324</td>
<td>0.021</td>
</tr>
<tr>
<td>F-statistics</td>
<td>17.88**</td>
<td>1.74</td>
</tr>
</tbody>
</table>

Notes: **p < 0.01; *p < 0.05; ~p < 0.10. The columns represent the dependent variables and the rows represent the independent variables of the VAR estimation equations.

Table 4  One-year forecast returns (%) by sentiment level

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Sentiment level</th>
<th>High minus low sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High (&gt; 1 S.D.)</td>
<td>Low (&lt; -1 S.D.)</td>
</tr>
<tr>
<td>Market</td>
<td>10.9</td>
<td>35.5</td>
</tr>
<tr>
<td>Large-value</td>
<td>–37.1</td>
<td>20.7</td>
</tr>
<tr>
<td>Large-neutral</td>
<td>5.1</td>
<td>41.1</td>
</tr>
<tr>
<td>Large-growth</td>
<td>13.3</td>
<td>32.8</td>
</tr>
<tr>
<td>Small-value</td>
<td>13.2</td>
<td>64.5</td>
</tr>
<tr>
<td>Small-neutral</td>
<td>4.3</td>
<td>54.8</td>
</tr>
<tr>
<td>Small-growth</td>
<td>11.2</td>
<td>46.9</td>
</tr>
<tr>
<td>Average small minus</td>
<td>–8.06</td>
<td></td>
</tr>
<tr>
<td>Average value minus</td>
<td>–26.99</td>
<td></td>
</tr>
</tbody>
</table>

The greater price volatility in the case of small stocks is intuitive because these stocks are associated with higher information uncertainty and hence more subjective to value. However, there is a need to clarify the relationship for growth versus value stocks, defined on the basis of market to book value ratios, especially because the trends seem to be different from Baker and Wurgler (2006), who find that the sentiment effect for the US market is greater for both extreme growth and distress stocks.

We expect greater uncertainty to be associated with low market to book value stocks than with high market to book value stocks, noting that the ‘value’ label for low market to book value stocks and ‘growth’ label for high market to book value stocks may be misnomers in the Indian case. A large proportion of high price to book ratio companies in India may not necessarily be associated with growth stage of their lifecycle, for example,
Investor sentiment and asset returns

several companies that own consumer brands or have profitable but stable businesses. These companies are likely to be associated with low levels of information uncertainty or ambiguity in valuation. On the other hand, there are a large number of companies that have low price to book value, even though they are still in nascent or early growth stage, as these are associated with either distress or high information uncertainty.

Figure 4  Volatility of characteristics-based portfolios

5.2 Regression tests

In order to confirm the results of the descriptive two-way analysis, we formally test the relationship between sentiment and future returns using the following equation:

\[
RET = a + b_1 \text{SENT} + c_1.IIPB + c_2.IIPC + c_3.IIPI + c_4.IIPCD \\
+ c_5.IIPND + c_6.TB + c_7.TERM + e
\]

(3)

In the above equation, RET represents the long-term return for the market or the relevant portfolio, SENT for the sentiment levels index, IIPB, IIPC, IIPI, IIPCD and IIPND for the year-on-year growth in the index of industrial production for basic goods, capital goods, intermediate goods, consumer durables and consumer non-durables respectively, TB for yield on 91-day Treasury Bill and TERM for the term structure of interest rates (estimated as regression residual of ten year G-Sec yield with 91-day T-Bill yield). OLS regression is used to estimate the equation.\(^1\)

There are two problems associated with testing association with long-term returns using the above regression equation. One, correlations between changes in SENT and the changes in returns tend to bias the regression coefficients. Two, auto-correlations in the time series of overlapping long-term returns data, results in understated estimates of standard errors of the coefficients. Following Brown and Cliff (2005), we use bootstrapping simulations to correct the bias in coefficients and under-estimation of
standard errors. Estimating robust standard errors using Newey West method to mitigate
the problem caused by overlaps is inadequate since in this case the prediction horizons
are long relative to the length of the time series.

For bootstrapping purposes, we use a VAR model with returns and sentiment as
endogenous variables, and business cycle indicators listed in equation (3) as exogenous
variables. We use AIC to suggest the number of lags but cap the number at three lags
considering parsimony. A constraint corresponding to the null hypothesis of no effect of
SENT on returns is imposed on the coefficient matrix. The residuals derived from the
estimation are re-sampled and added to the fitted returns to create new samples of return
series. A sample size of 10,000 is used for bootstrapping trials.

From the resultant samples of returns, long-term returns are estimated at 12, 24 and
36 month horizons. Equation (3) is then estimated for each case to provide a large
distribution of coefficients of SENT. In each case the boot-strapped p-value is derived
using the b1 coefficient of original OLS regression of equation (3) and the bootstrapped
distribution of b1.

Table 5 presents the coefficients for SENT and p-values from regressions with market
returns at various horizons. Newey West t-statistics and p-values are also presented for
comparison. Boot-strapped coefficients are less significant than those using Newey West
methods, since they are adjusted for bias and have higher standard error estimates. We
estimate Newey West standard errors for each bootstrap simulation trial, and then
compare the properties of t-distribution of the estimated Newey West t-statistics with the
distribution of t-statistics of the bias-adjusted coefficients (bias adjustment is done by
subtracting the mean of the bootstrapped coefficients obtained from all the trials). The
distribution of boot-strapped t-statistics approximates normal distribution well in all
cases, whereas the distribution of Newey West t-statistics is skewed with fatter lower tails
and thinner upper tails than a standard normal or student’s t distribution.

Since the values of SENT are standardised, the coefficients imply change in average
monthly returns (in percent) for one standard deviation change in SENT. At all horizons,
the coefficients of SENT, though negative, do not have any statistical significance.

Table 5  Coefficients of SENT as predictor of market returns

<table>
<thead>
<tr>
<th>Return horizon</th>
<th>Boot-strapped values</th>
<th>Newey West values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>12 months</td>
<td>–0.33</td>
<td>0.26</td>
</tr>
<tr>
<td>24 months</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>36 months</td>
<td>–0.03</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Note: Boot-strapped coefficients are adjusted for bias.

Table 6 presents the coefficients for SENT and p-values for regressions for the size-value
portfolios. Confirming the two-way analysis, the coefficients decrease with size and
increase when moving from growth to value. The inverse relationship between returns
and SENT is significant for only the small-value firms, though there is a weak association
in the case of small-neutral and large-neutral firms.
Table 6  Coefficients of SENT as predictor of 12 months portfolio returns

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>Boot-strapped values</th>
<th>Newey West values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>Large-value</td>
<td>–0.27</td>
<td>0.49</td>
</tr>
<tr>
<td>Large-neutral</td>
<td>–0.59</td>
<td>0.07</td>
</tr>
<tr>
<td>Large-growth</td>
<td>–0.22</td>
<td>0.46</td>
</tr>
<tr>
<td>Small-value</td>
<td>–1.08</td>
<td>0.03</td>
</tr>
<tr>
<td>Small-neutral</td>
<td>–0.79</td>
<td>0.07</td>
</tr>
<tr>
<td>Small-growth</td>
<td>–0.53</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Note: Boot-strapped coefficients are adjusted for bias.

We also conduct the Granger causality tests to check for statistical evidence that sentiment causes long-term returns. The tests fail to reject the null hypothesis that sentiment does not cause returns for any portfolio at any horizon. Even in the case of small-value firms, where the regression estimates show that the sentiment index predicts long-term returns, there is no evidence that sentiment actually influences stock returns in the subsequent period.

5.3 Comparative analysis with price multiples

According to research literature, valuation multiples are inversely related to future returns. Campbell and Shiller (1998) introduced the cyclically adjusted price to earnings multiple as an inverse predictor of long-term returns. In the Indian context, Sehgal and Pandey (2014) found evidence that stocks with lower valuation multiples outperformed stocks having higher valuation multiples.

Hence, we also test equation (3) using price to earnings multiple (P/E) and log of price to dividend multiple (log of P/D, the inverse of dividend yield) in place of SENT to predict subsequent 12 months’ market returns. These multiples are valuation indicators, and hence their relationship with long-term returns is assumed to be primarily due to changes in discount rates rather than due to sentiment. However, it is incorrect to assume that the changes in multiples represent only rational time varying behaviour. For illustration, we can decompose the price to earnings multiple as follows:

\[
\log (P_t / E_{t+1}) = \log (V_t / E_{t+1}) + [\log (P_t) - \log (V_t)]
\]  

In the above equation, \(P_t\) denotes the current price, \(V_t\) denotes intrinsic value and \(E_{t+1}\) denotes one year forward multiple. The first term in equation (4) is a function only of the expected growth in future cash flows and of the associated discount rates, and can be estimated using a valuation model. At an aggregate level, the first term is therefore expected to change with macro-economic expectations. The second term denotes the gap between the stock price and the fair value, and can therefore be explained by reasons other than expected returns, for instance by sentiment or random error.
Thus it becomes difficult to interpret whether the changes in price multiples are due to rational changes in economic expectations and hence discount rates, or due to cyclicality induced by investor sentiment. However, to the extent that SENT provides a measure of sentiment, which is expected to be correlated with the second term in equation (4), a comparison of predictability of future returns by the valuation multiples vis-à-vis SENT can indicate whether the return predictability is associated primarily with investor sentiment or changes in macroeconomic expectations.

With this background, we test equation (3) for the price to earnings and price to dividend multiple using bootstrapped simulations for corrected standard errors. The values of price to earnings ratio and log of price to dividend were standardised before the regression for meaningful comparison of the coefficients. The results are shown in Table 7.

**Table 7** Comparison of predictors of market returns

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Sample period</th>
<th>Boot-strapped values</th>
<th>Newey West values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coefficient</td>
<td>p-value</td>
</tr>
<tr>
<td>SENT</td>
<td>Apr 2004–Mar 2015</td>
<td>–0.33</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Jun 2009–Mar 2015</td>
<td>–0.38</td>
<td>0.09</td>
</tr>
<tr>
<td>P/E</td>
<td>Apr 2004–Mar 2015</td>
<td>–1.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Jul 2007–May 2013</td>
<td>–1.50</td>
<td>0.01</td>
</tr>
<tr>
<td>P/D</td>
<td>Apr 2004–Mar 2015</td>
<td>–0.91</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Nov 2009–Apr 2013</td>
<td>–0.85</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Note: Boot-strapped coefficients are adjusted for bias.

The tests show that unlike in the case of SENT, the coefficient of P/E is statistically significant ($p = 0.03$), though the coefficient of log P/D is not significant ($p = 0.11$). In terms of economic significance, whereas one standard deviation change causes a –4.0% change in annualised market return over the subsequent 12 months, one standard deviation change in P/E results in a –12.4% change in annual return, while one standard deviation change in log P/D results in –10.9% change in annual return. Further, the coefficient for P/E is also found to weakly cause (using the Granger test) subsequent one year’s market return ($p < 0.07$).

The above analysis is further extended to take into account structural breaks. The presence of structural breaks was confirmed using Quandt Andrews breakpoint test in the case of all the three equations (that is with SENT, P/E and log P/D as alternative variables). To identify the breakpoints, multiple breakpoints test was conducted, following Bai and Perron (2003). The tests reveal multiple breakpoints in each case, on November 2007 and June 2009 in the case of SENT, November 2005, July 2007 and June 2013 in the case of P/E, and September 2006, November 2009 and May 2013 in the case of log P/D. In each case, only one of the sub-periods defined by the breakpoints could be used for retesting the null hypothesis of no predictability of 12 month returns, since other sub-periods had inadequate data points. The second part of Table 7 shows the results considering only the relevant sub-period in each case. The hypothesis of no return predictability is rejected in the case of P/E at $p < 0.05$, but not in the case of log P/D or SENT. These results are consistent with those found for the entire sample period.
In investor sentiment and asset returns

The comparisons imply that during April 2004 to March 2016, investor sentiment played a modest role in influencing future returns in comparison to changes in discount rates and therefore economic expectations. One reason for this finding could be the relatively low share of retail investors in India in the total stock market turnover, as compared with the share of institutional investors. Hence, institutional sentiments, economic expectations and liquidity effects induced by foreign institutional flows could have played a greater role in influencing market returns than retail investor sentiment.

6 Conclusions

This paper describes the construction of an index of investor sentiment for India based on common variation among five sentiment proxy variables and an analogous index of changes in investor sentiment. The changes in sentiment index co-vary with stock returns, though the causality proceeds from returns to sentiment. Further we find evidence that the level of investor sentiment did not significantly predict the subsequent returns for the overall stock market in India over the study period of 2004 to 2016, though it does predict the returns in the case of small, low-priced stocks.

Since sentiment indicators appear to be contrarian only for small stocks and value stocks, investors can use these to a limited extent as predictors of future returns over 12 months horizon. However, sentiment indicators would not be useful predictors of stock returns in the case of large stocks and growth stocks. In comparison, price to earnings ratio has been a statistically significant contrarian indicator. These results imply that investors should primarily consider relative market valuation using price to earnings ratio for timing aggregate changes in equity allocations whereas sentiment indicators can be used as contrarian indicators while making timing decision only in the case of small and value stock segments.

Policy makers and regulators concerned about asset bubbles should also primarily focus on the relative valuation rather than on sharp increases in sentiment indicators. For this purpose refined earnings multiples similar to the cyclically adjusted price-earnings ratio estimated by Campbell and Shiller (1998) can be used.

While this research focuses on time series behaviour of stock returns based on investor sentiment, there is further scope of work in examining the role of sentiment in cross-section of expected returns in an emerging market context.

References


**Notes**

1. Collinearity between the IIP components is not expected to affect the coefficient of SENT, since SENT itself has low correlations with these components, as the variables used to construct SENT have been made orthogonal to IIP. This is also confirmed by the low variance inflation factors for the regression coefficients of SENT in all the estimations.

2. The same equation can be stated in terms of trailing price to earnings multiple by including one year expected growth rate in cash flows on the right hand side of the equation. Further, the same equation can be modified without loss of generality to represent other price multiples such as price to dividend ratio or market to book value.