CAPM WITH SENTIMENT

THE EFFICIENT MARKET HYPOTHESIS SPICED UP WITH SENTIMENT

Claudio BOIDO[a]
Professor of Banking and Finance
University of Siena and Rome (LUISS)

Antonio FASANO[b]
Professor of Banking and Finance
University of Salerno and Rome (LUISS)

Abstract

We analyse the relationship between large cap returns and sentiment indexes, using a Capital Asset Pricing Model (CAPM) framework. We try to provide a better explanation of asset prices and their deviations from standard theories by means of sentiment indicators, assuming the latter being measures of the very inclination to speculate. Therefore, when sentiment is high, investor demand for speculative investment is high; conversely when it is low, investor demand for speculative investments is low. Unlike other studies, based on proxies, we use the European Sentiment Indicator and its constituents, based on direct surveys, to assess business and consumer confidence.

---

[a]Claudio Boido is Professor of "Financial Markets and Institutions" at the University of Siena and at the University of Rome LUISS.

[b]Antonio Fasano earned his Ph.D. at the University of Rome "La Sapienza". He is Professor of "Equity Markets and Alternative Investments" at the University of Rome LUISS. docenti.luiss.it/fasano.

Part I

1 Introduction[1]

Before the fifties, Statman 2005 stated that the investor framework was "normal", then later with the introduction of the modern portfolio theory the investor framework became rational. In recent years it has become clear that it is possible to identify two different schools of thought, concerning the response of traders in terms of stock prices to breaking news. The first group advocates rational expectations and market efficiency (on the basis of financial market theory people choose how to invest), while the second school of thought claims that in some situations people behave irrationally (where investors buy and sell instinctively). Rational investors use instruments for an expected return and standard deviation, while normal investors are affected by cognitive biases and emotions, which may influence asset prices at least theoretically. Thaler 1999 has examined five points (volume, volatility, dividends, equity premium puzzle, predictability) where the behaviour seems to be at odds with the theories. Rational investors state that they know when it is better to buy or to sell, while fund managers often turn over their portfolios during the year. In a rational world, prices change every time news breaks, while stock and bond prices change more frequently, because they are more volatile and so the price changes without any reference to its intrinsic value. If a dividend policy is irrelevant, why is there an increase in stock prices when dividends are announced and why do most large companies pay cash dividends? It is clear that one would expect a risk premium (the difference between return equities and government bonds), because equities are riskier. A strong dividend policy justifies the reason why the stock price increases when companies pay dividend. This topic helps to explain why investors select equities when they expect a higher risk/return relationship. If the markets are strongly efficient, all existing information is included in the stock price. While, it is possible to make a profit if the investor knows confidential information before the financial market does.

In this context "normal" investor reactions could be different to what many "rational" analysts expected. Consumer confidence is an indicator which stimulates the thinking investors to react to possible "states of the world" (recession, expectation, and euphoria). In 2008 the last "real estate bubble" burst and provided evidence to support those academics and practitioners, who stated that it is not only economic factors that affect the market, but also, in particular, sentiment and irrational considerations. It's intuitive to understand that this statement relies on the fact, that in a rational and efficient market (in Fama's meaning) without "herding effects" (when the investors act together without planned direction), bubbles should not occur. Market anomalies easily support the quality of the thesis of behavioural economics and it is obvious to assert, that economics and finance paradigms should be revised to insert irrational investor behaviour in investment courses at Universities and at Business Schools.

At the end of the last century Tversky, Kahneman, et al. 1981 asserted that the perceived risk may be affected by the framing of the non-economic and non-statistical facts at hand. Long et al. 1990, Barberis et al. 1998, Baker and Wurgler 2007 established the theoretical grounds for the market in general, and that returns in particular are affected by sentiment. Emotions and mood may affect investors' perceptions and thereby their

---

[1]This working paper is the final result of a research study carried by two authors. We can attribute Part I to Claudio Boido and Part II to Antonio Fasano.

investment decisions which in some ways are often difficult to unravel. It is important to establish that the sentiment effect can explain some of the anomalies observed in the market. In this way it does not mean that all decisions are irrational, but in some situations, when people are anxious and nervous their decisions may be affected by irrational factors and cannot be explained by traditional economic models. It is accepted that it is correct to analyse investors' sentiment, because it helps to shed light on biases in investor stock market forecast and to give us the opportunity to make a profit by examining those biases.

Recently Akerlof and Shiller 2010 stated: "*To understand how economies work and how to manage them and prosper, we must pay attention to the thought patterns that animate people's ideas and feelings, their animal spirit. We will never really understand important economic events unless we confront the fact that their causes are largely mental in nature*". It is clear that the animal spirit plays a fundamental role in trying to explain most market anomalies and this would reduce the importance of the modern portfolio theory. This could define investor sentiment as the inclination to speculate, so when sentiment is high, investor demand for speculative investment is high, conversely when it is low, investor demand for speculative investments is low. It is correct to assert that some stocks are more sensitive to speculative demand and those which are more difficult to value, tend to be the riskiest to arbitrage. Baker and Wurgler 2007 made two assumptions: the first is that investors are subject to sentiment, the second assumption is that betting against sentimental investors is costly and risky. Their study showed that the sentiment effect has a stronger influence on small-firm securities. It has been shown that speculators concentrate their portfolio choices in small cap shares, because they are more volatile, non-dividend paying and often these firms are unprofitable. Cochrane 2002, Ofek and Richardson 2003 have looked at some of the reasons behind the bubble: the rise and fall of tech stocks during the internet bubble. Asness et al. 2000 considered two factors: a) the spread in valuation multiples between a value portfolio and a growth portfolio (the value spread) b) the spread in expected earnings growth between a growth portfolio and a value portfolio (the earnings portfolio) They find that the greater the value spread and the smaller the earnings growth spread the better their forecast for value. Chan et al. 2000 affirmed that the surging stock prices cannot be attributed to rational expected earnings. They provide evidence of the existence of extrapolative biases in the pricing of value and glamour stocks Shiller 2000 drew parallels between what happened in the late 1920s to the 2000 internet bubble. He offers a good description of the mood behind the market. The author analyses psychological factors and stress circumstances modifying investors' perception of the market. For example he finds an higher investor sentiment before the bubble burst. This situation seems consistent with the hypothesis that sentiment increases the relative price of stocks which are most subjective to value and the hardest to arbitrage. Fisher and Statman 2000 analysed the moods of three groups of investors: large, medium, and small. They associate Wall Street strategists as a large; writers of investment newsletters as medium and individual investors as small. They assert that there is a strong relationship between the sentiment of individual investors and the writers of the newsletters, while the other relationship is weak. The correlation of 0.47 is statistically significant, whereas the changes in the sentiment of strategists are not related to changes in the sentiment of individual investors and newsletters. It is interesting to note that a combination of the sentiment between the three groups and S&P500 index is high.

In this research we study the relationship between returns on the Italian Stock Index

regarding large and small caps quoted on Italian Stock Exchange and consumer indexes. Consumer confidence increases when economic conditions are good, so consumers and companies are stimulated into borrowing money. On the other hand economic uncertainty causes pessimism, which reduces the demand for goods and money.

As for sentiment indicators, we base the analysis on consumer and business surveys conducted in Italy by the National Bureau of Statistic, under EU guidelines. These surveys demonstrate (among other things) the economic expectations of households and firms, furthermore they show their willingness to spend in durable goods or to expand production and their perception about current and future employment levels (see section (3.2) for details).

2 Recent Literature

The rational investor makes a decision according to the axioms of the expected utility theory and he makes unbiased forecasts about the future. Efficient Market Hypothesis (EMH) is based on investors who are rational and their valuation of securities is made rationally. Normal investors are not rational and their trading is random and they do not have an effect on prices. Even if investors were irrational it would be the rational arbitrageurs who would eliminate any distortion of the market. Shiller 1980 has shown that stock market prices are more volatile than what efficient market theory affirms. In fact a stock market crash would not be justifiable if the market price took into consideration all the available information. Many authors, Bernard 1992, Jegadeesh and Titman 1993, Chan et al. 1996, Fama and French 1988, Poterba and Summers 1988, Kothari and Shanken 1997, Lo and MacKinlay 1990, have written about under reaction and overreaction, that is how security prices underreact or overreact to news which reaches the stock market. Bernard reviewed some evidence on market efficiency in relation to accounting earnings. The average initial response to earnings announcements is an underreaction, while extreme stock price movements represent over reaction to earnings. The concentration of the abnormal return in relation to subsequent earnings announcements. Jegadeesh and Titman 1993 have affirmed that the strategies to buy stock winners and to sell losers performed better in the first year, but this result was not confirmed in the following year. Chan et al. 1996 examined the predictability of future returns from past earning news. They showed that the market reacts only gradually to new information. Fama and French 1988 examined autocorrelation of stock returns and for the 1926-1985 sample period large negative autocorrelations for return horizons beyond a year are consistent with the hypothesis that mean reverting prices are important in the variation of returns. Poterba and Summers 1988 analysed transitory components account for a large fraction of the variance in common stock return. They concluded that the transitory components in stock price account for more than half of the variance in monthly returns. Kothari and Shanken 1997 studied the evidence that stock returns are predictable. They evaluated the ability of book to market ratio to track time series variations in expected market index returns and compared its forecasting ability to that of dividend yield. They noted that over the period 1926-91 the book to market ratio is stronger over the full period while the dividend yields relation was more solid in the sub period 1941-91.

Daniel et al. 1998 and Hong and Stein 1999 have found that models based on investor behaviour create over or under reaction due to the presence of noise traders who are overconfident or have biased self-attribution in evaluating their performance. The problem is also caused by the presence of different classes of investors, who give different importance

4

to the news, watching the impact of the fundamentals on price or on price trends. Hong-Stein studied the behaviour of two different groups of investors "news watchers" and "momentum trader": there is a tendency for asset prices to underreact to news in the short run and there is another tendency for prices to overreact to news in the long run. The news watchers forecast is conditioned by confidential news about future fundamentals without observing past or current prices. Momentum traders follow the past price changes without observing future fundamentals.

Sentiment is an abstract concept and therefore today the market can be optimistic, pessimistic or neutral, but this does not mean that today's sentiment will be the same for future months. It believes that it is quite simple to identify sentiment when people use a survey to ask them to think about the factors that could affect the investment environment over the next six months or how they rate the performance of stock markets over the next six months.

Sometimes it is important to know if investors make a distinction between the future of the economy and stock markets. Positive sentiment is higher where consumer confidence and investor sentiment is positive. The survey could be described as a good proxy to help you if the sentiment is high or low. Obviously it must observe if the different surveys, published by database providers, are useful in improving dynamic asset allocation decisions, because they help us to know the sentiment of the stock market. The situation assumes greater importance, during periods of high volatility, in order to avoid losses or to reduce portfolio risk. In recent years active managers looking to improving dynamic strategies use active alpha management; therefore also in this context behavioural finance shows that sentiment is useful in order to increase outperformance with alphas.

Kyle 1985 and Black 1986 have put the study of market or investor sentiment in their theories of the noise trader models. They suggest that some investors negotiate on noisy signals without any link to fundamental data and so market prices can deviate from the intrinsic value, thereby not respecting Fama's efficient market hypothesis. Behavioural theory assumes that interplay between noise traders and arbitrageurs fixes the market price and this is the opposite of the efficient market theory that affirms the market price of an asset differs minimally from the present value of expected cash flows and the arbitrageurs absorb demand shocks and shifts in investor sentiment. Shleifer 2003 affirmed that the assumption of investor rationally is contradicted by psychological and institutional evidence. Efficient market hypothesis states that securities prices must equal fundamental values. He illustrates a new theoretical approach extending the research to the behavioural finance. He cites two principal foundations of behavioural finance: limited arbitrage and investor sentiment. The latter involves investors' choice which is driven by heuristic "representativeness" and "conservatism". Representativeness is the tendency for people to view events as representative of some specific class and ignore the law of probability in the process. Besides investors are conservative and do not update their behaviour models in line with possible market changes. These drivers take in overreaction and under reaction of investors on stock markets. According to traders it is well known that the best time to buy is when individual investors are pessimistic and to sell when they are optimistic. It is correct to ask if uniformed investors, called noise traders, have influence on the prices of financial assets. Long et al. 1989 have shown how traders acting on non fundamental information could affect prices in a systematic way, the changes and the volatility in stock prices are explained not only by fundamentals, but also by irrational "noise trading". They have attempted to verify or deny the existence of noise traders in the U.S. market for closed end investment which invests in securi-

ties. The difference, between the net asset value (so called NAV) of funds and its market price, is called the discount. As this value has varied since the former authors wrote the article, the behaviour of discounts may therefore be governed by the sentiment of the noise traders and so closed end funds have become a good sector of the market where researchers try to find noise traders. In this way a trader acting on non-fundamental information could affect prices and so we can identify two different types of investors. The first ones are driven by fundamentals to select on the basis of price and the second ones trade on noise signals. Consequently noise traders create an additional source of risk which adds to price volatility. The relationship between discounts and sentiment has shown that: when noise traders are bullish discounts should decrease because the market price increases and the difference is reduced and vice versa. In these model groups the most fundamental prediction of the noise trader is that irrational investors select on the basis of a noisy signal called sentiment. This causes risk and therefore price volatility. In some cases it uses a survey on a randomly selected group of members of an association of investors to measure investor sentiment directly. The question concerns contexts in which investors think the stock market will be in sixth month - bullish or bearish. On the basis of the survey's results investors understand the direction of the sentiment. Brown 1999 showed that individual investor sentiment is related to increased volatility in closed end investment. He affirmed that high levels of individual investor sentiment are associated with higher level of volatility The right form, in which sentiment influences returns or volatility, is not as clear as before. If noise traders are sensitive to sentiment changes, then sentiment variations should drive returns and volatility. We expect that the level of sentiment will influence earnings and volatility. Neal and Wheatley 1998, Wang 2001, Simon and Wiggins 2001 show that sentiment can predict returns and use the positions held by large traders in the futures markets, as a proxy for sentiment. They also discover that they are useful for predicting the future returns in the next period. Wang studied the sentiment indicator in futures market. He found that the sentiments of large speculators are a price continuation indicator, while large hedger sentiment is a contrary indicator. Investors are advised to go short when hedgers are turning bullish and to go long when they are turning bearish. So investors make profit if they buy when large speculators are bullish and large hedgers are bearish and sell when large speculators are bearish and large hedgers are bullish.

In the past articles Brown and Cliff 2004, Qiu and Welch 2004, Shiller 2000found some solutions for identifying and measuring investor sentiment: direct, indirect, meta.

The following meanings can be highlighted :

- Direct: based on surveys which directly evaluate market participants sentiment.

- Indirect: based on financial data and require theoretical backgrounds. They include buy-sell in balance, put-call ratio, Barron's confidence index.

Meta measurements: they are hybrid versions on direct and indirect measures. An example is a composite index consisting of common components coming from several sentiment proxies (closed end fund, number of IPO's, turnover ratio, dividend premium.

Direct sentiment measures are based on polling investors and so the expectations of the market participants can be measured directly. In this case, if you analyse the answers to the questionnaire, many possible sources of errors may be identified. The polling of investors suffers from deficiencies common to most survey based studies, which have problems related to inaccurate responses regarding business conditions, employment conditions and family income for the following six months.

To measure indirect sentiment financial variables are needed, which require a theory linked to sentiment. The weak point is the need to build this theory and its respective interpretation.

Meta measurements include analysts' opinion, columns, and online forums; therefore it is based on a mixture of opinions.

In this kind of research the greatest deficiency is the short sale restriction for retail investors in the equity market. In fact when the investor decides to express a negative sentiment he could sell stocks short or he does not buy stocks, but he has a short sell restriction, so he has a difficulty in expressing a negative sentiment. Investors can discover another problem of disentanglement, which is the act of segregating institutional trade from retail trade. Baker and Stein 2004 includes a trading volume as a measurement of liquidity and therefore as a measurement of investor sentiment. In fact they assert that investors reduce liquidity when they are pessimistic. Sometimes a leading independent provider of investment research, such as Investors Intelligence, obtains information from independent market newsletters and assesses each author's opinion of the market. Each newsletter is classified as bullish, bearish, or neutral regarding their expectations for future markets. Das and Chen 2007 investigate a methodology for analysing messages on stock message boards driven by providers like Yahoo! Generally it's common to use statistical and natural language processing techniques to obtain the emotive content of a bearish, bullish or neutral signal. A valid measurement of sentiment is based on opening long call and opening long put options purchased by clients. The ISEE Sentiment Index is calculated by the International Securities Exchange (ISE) and in particular they calculate the ratio as follows: (opening long calls/opening put)/100. A value above 100 implies bullish sentiment, while a value under 100 implies bearish sentiment. Baker and Nofsinger 2002 and Akhtari 2011 attribute that weather factors induce bad moods and alter investors' risk attitude. In fact Kamstra et al. 2003 have shown that stock prices tend to decline when Seasonal Affective Disorder (SAD) prevails. Returns are significantly lower when the length of the day is shorter. In the following research Kamstra et al. 2008 find that during the fall season, when daily equity returns are low, daily Government Bond returns are high, as they should be if indeed risk averse investors shun risky stock and favour safer alternatives. This phenomenon is quite well known, because it is assumed that it is related to changes in the amount of daylight and so the length of the day and temperature are the only significant factors related to SAD. Boido and Fasano 2005showed that individual investors are affected by psychological factors and that returns are low on Monday because they are inclined to sell stocks, while on Fridays investors prefer to buy stocks.(daylights saving effect). SAD is a cyclic illness with recurrent episodes of (fall/winter) depression alternating with periods (spring/summer) with a normal mood. Chan and Lakonishok 2004 confirmed the importance of sentiment investors and they affirm that value stocks outperform growth stocks, but the relationship has deteriorated because of the bubbles. During the technology bubble (2000) over optimism caused stock valuations in the technology industries to deviate from their intrinsic values. Recently Baker and Wurgler 2006 have examined the impact of investor sentiment. It is possible to associate the concept of beta with sentiment. It is imagined, after the great crisis, investors continue to be pessimistic and expect the market as a whole to decrease by 15%. They would herd toward the market index by buying and selling individual assets until their prices decreased by 15%. Those assets relative to other assets in the market might be sold rendering their beta a downward bias. This kind of herding behaviour would induce the betas on the assets to converge toward the market beta.

Basu et al. 2006 have shown that the addition of sentiment variables to business cycle indicators improves the performance of the active portfolio and these results are statistically significant. The strategies based on sentiment exploit an overreaction, which leads to active alpha strategies. Kim and Ha 2010 show, using a version of Fama-French-Carhart four factor model (i.e. market, size, book market value, momentum), that sentiment affects the stock price of companies with a small cap, a low price and low book market value. Baker and Wurgler 2006 have proposed a sentiment index based on the main principal component of the following proxies: closed end fund discount, share turnover, number of IPO, the average first day's return of IPO, the equity share in new issues and dividend premium. The stocks, that are hardest to arbitrage, tend to be the most difficult to value. Extreme growth stocks seem to be prone to bubbles consistent with the observation they are more appealing to speculators and optimists and to arbitrage. If there is a large number of investors who perceive the value of the stock to be higher than the current stock price, the sellers have low sentiment for the firms while the buyers have high sentiment. It can be seen that there are millions of investors who have different sentiment levels. The sentiment level is dynamic and it is changing. Investors may change their sentiment over time-depending on macroeconomic conditions, specific expert firms and analyst views. Consumer confidence increases (decrease) when investors grow bullish (bearish) but these investors do not always increase their stock return when consumer confidence is low.

Lawrence et al. 2007 have shown how a stock price changes by modifying the components of the existing dividend discount model with the investor sentiment as a component. In this context stock price is governed not only by firm fundamentals, but also by investor sentiment. The authors specify that investor sentiment affects both the expected growth rate and the expected discount rate. It is possible to evaluate the future performance of the company by individual beliefs. The traditional form of CAPM incorporates investor sentiment with a modified beta, B^s, as a function of beta and investor sentiment. If sentiment is high the firm will be perceived as less risky and the value of modified beta will be lower, thus reducing expected return (vice versa). Given the modified beta B^s, CAPM expected return the sentiment can be expressed:

$$\mathbb{E}r^s = r_F + \beta^s \mathbb{E}(r_M - r_F)$$

Investors who think about the relevant future performance of a company will wait for a higher growth rate than those who believe that the firm is a certain failure. The growth rate for the firm is modified from g (the growth rate in the simple Gordon and Shapiro Model) to g_s, where g_s is a function of growth rate g and investor sentiment.

For a company where the investor expected return is greater than the growth rate and constant growth, the modified Gordon equation for the stock price is:

$$P_0 = \frac{d_1}{r_s - g_s}$$

For an investors with strong sentiment for a company's future performance, the expected discount rate r_s will be low, while the expected growth rate g_s will be higher, thus making the value of the stock higher (as perceived by the investor). Similarly, the considered value of the stock will be small for investor with low sentiment for a firm's future outlook. When the market price is higher than what the low-sentiment investor expects it should be, he will sell the stock. When it is lower than what the high-sentiment investor realizes it should be, he will buy it.

Blanchard and Watson 1983, Tirole 1982 Hong and Stein 1999 have affirmed that stock markets are influenced by price bubbles when it is difficult to make a forecast on the fundamental value of stock even if investors are considered to be rational.

Part II

3 Empirical Analysis: a Sample Application

Among the main hindrances to the efficient market hypothesis is the presence of psychological biases on behalf of market participants. This makes theoretical (expected) prices, derived by CAPM models, inconsistent with actual observed prices and the market factor inadequate to explain the excess returns.

Insofar as this diversion can be ascribed to the market "sentiment", we can attempt to augment the standard model by means of this psychological factor. It is common to capture investor sentiment indirectly through a proxy and a number of alternatives used by the literature were presented in section (2). In this study, instead, we used the survey data feeding the European Sentiment Index.

To measure how and if the sentiment factor affects stock returns, we will build a multivariate factor model, in a CAPM-like fashion, where the ordinary market factor is coupled with a sentiment factor for better explanation of stock returns. We will test the model fitting with and without the behavioural factor and, as long as sample data is consistent with the theory exposed here, we expect to obtain a better fit in the latter case.

One problem we face in analysing return dynamic is given by the relatively large number of sample stock returns to be confronted with market and/or sentiment factor: in fact for some stocks, tests can confirm the theory behind the model; for others tests could be not meaningful or even contrast sharply with theoretical assumptions. We therefore need a robust method to aggregate single-stock behaviour into a market-wide dynamic. We discuss in deep the approach to address these questions, and the underlying assumptions, in section (4).

Before diving into the model technicalities, we will discuss the data set (section (3.1)), particularly the main data source of this study, the *Joint Harmonised EU Programme of Business and Consumer Surveys*, with the related sentiment indicators (section (3.2)).

3.1 Data Set

The data set, obtained via Bloomberg service and the EU public access data, is constituted of time series concerning: the Italian FTSE-MIB index, FTSE-MIB constituents, Italian Sentiment Indicators and the Italian risk free rate proxied by means of Treasury-Bill gross rate.[2]

We use monthly log-returns spanning over the time window 1999-2011, including the sentiment index too, analytically treated as a price; for free rate, instead, we use Bloomberg calculated returns for free rate[3] as is.

---

[2]Italian T-Bills are named "Buoni Ordinari del Tesoro", BOT for short. We used the 3-month BOT gross rate time series calculated and published by Bloomberg platform on the basis of the related bond issues.

[3]See previous footnote.

The Italian FTSE-MIB is constituted of the largest capitalisation stocks in Italy. Securities not available in the index as of January 1999 throughout to September 2011 were excluded.

The remaining 21 (out of 40) securities are shown in Table 1 along with their respective Bloomberg ticker.

	Company	Bloom. Ticker
1	A2A SPA	A2A
2	ATLANTIA SPA	ATL
3	AUTOGRILL SPA	AGL
4	BANCA POP EMILIA	BPE
5	BANCA POP MILANO	PMI
6	BUZZI UNICEM SPA	BZU
7	ENI SPA	ENI
8	FIAT SPA	F
9	FINMECCANICA SPA	FNC
10	FONDIARIA-SAI	FSA
11	GENERALI ASSIC	G
12	IMPREGILO	IPG
13	INTESA SANPAOLO	ISP
14	MEDIASET SPA	MS
15	MEDIOBANCA	MB
16	MEDIOLANUM SPA	MED
17	PIRELLI & C.	PC
18	SAIPEM SPA	SPM
19	STMICROELECTRONI	STM
20	TELECOM ITALIA S	TIT
21	UNICREDIT SPA	UCG

Table 1: The list of securities scrutinised included in the Italian FTSE-MIB.

3.2 Economic Sentiment Indicators in Europe and Italy

The European Economic Sentiment Indicators are produced by the European Commission, through the Directorate-General for Economic and Financial Affairs (DG ECFIN), in the context of the Joint Harmonised EU Programme of Business and Consumer Surveys. These are qualitative economic surveys, intended for short-term economic analysis, which can offer signals for predicting turning points in the economic cycle.

The surveys are conducted on a monthly basis by means of questionnaires concerning five key areas:[4] industry, construction, consumers, retail trade and services. Nearly all the questions are of a qualitative nature. These questions have a similar answer scheme, according to a three-option ordinal scale, that can be interpreted as a positive answer ("increase", "more than sufficient", "too large", etc.), a neutral answer ("remain unchanged", "sufficient", "adequate", etc.) and a negative answer ("decrease", "not sufficient", "too small", etc.); but in some cases, respondents have the choice between more questions.[5] This scheme is intended to capture managers' assessment of the trends in their sectors or households perceptions concerning the factors influencing their spending decisions.

Questionnaire data are mapped in five specific indicators, summarising judgements and attitudes of producers and consumers. Each indicator is based on a selection (not all)

---

[4]Some questions are part of the questionnaire only once a quarter.

[5]For instance, in the consumer survey respondents have a five-option ordinal scale.

of the questions comprised in the questionnaire. For example the Industrial confidence indicator is built up upon the following three questions:

- Do you consider your current overall order books to be...? ("more than sufficient", "sufficient", "not sufficient")

- Do you consider your current stock of finished products to be...? ("too large", "adequate", "too small")

- How do you expect your production to develop over the next 3 months? It will... ("increase", "remain unchanged", "decrease")

As for the methodology, the indicators are built up upon the arithmetic average of the balances in percentage points of the answers to some (not all) of the questions deployed in the related area. Often balances are seasonally adjusted.[6]

A composite index, the Economic Sentiment Indicator (ESI), is derived by the five sectoral indicators applying to each a different weight.

Table 3.2 shows both sectoral indicators and their weight used in computing ESI.

Indicator	Weight
Industrial confidence indicator	40%
Services confidence indicator	30%
Consumer confidence indicator	20%
Retail trade confidence indicator	5%
Construction confidence indicator	5%

Table 2: Sectoral indicators concurring to the definition of the Economic Sentiment Indicator (ESI)

The surveys are carried out at national level by public and private partner institutes, selected by the Commission through a call for proposals every 3-4 years. The questions are harmonised following EU guidelines, however collaborating institutes at national level may include additional questions and the sectoral breakdown may be more detailed.

The EU Directorate General of Economic and Financial Affairs is in charge to produce the aggregate surveys of the aggregate results received from the Member States, based on weighted averages of the country-aggregate replies.

In Italy sentiment indicators are often referred to as "ISAE indices", since they were in charge of the Institute for Studies and Economic Analyses (ISAE), however as from January 1, 2011 ISAE surveys and updates are part of ISTAT (Italian National Bureau of Statistics) surveys.

Our study is focused on Italian surveys and indicators as shown in Table 3. The time window spans from January 1999 to September 2011.

Some descriptive analysis concerning the indices follows.

First, in Figure (1) Box-and-whisker plots show Tukey's five number summary of the confidence indicators. Cross and serial correlations are then studied. In Figure (2) correlations among sentiment indices is shown. Figure (3) shows auto-correlation with different time lags for each index. As we can see, both cross and serial correlations are not significant.

---

[6]For detailed methodology see European Commission. Directorate-General for Economic and Financial Affairs 2007.

Confidence/Sentiment Indicator
Industry (ITA)
Services (ITA)
Consumer (ITA)
Retail (ITA)
Construction (ITA)
ESI (ITA)

Table 3: The family of EU indicators used in the analysis, based on Italian surveys conducted by the Italian National Bureau of Statistics.

4 A factor model to explain returns

As depicted in the introduction, we use a multivariate factor model to explain stock returns. The factors are the market and different sentiment indicators. Investor's yield can be measured in term of excess returns as in CAPM.

Before introducing the model, we set the notation and the variables that will be used throughout this paper.

Given a matrix A with dimension $N \times M$, whose typical element is a_{ij} and enjoying the property $\varphi(a_{ij})$, it can be denoted with one of the following notations:

$$\underset{N \times M}{A}, \quad [\![a_{ij}]\!]_{ij}^{N \times M}, \quad [\![a_{ij} \mid \varphi(a_{ij})]\!]_{ij}^{N \times M}$$

Similarly, the column-vector u and the row-vector v, with n components, could be denoted with (omitting for short possible component properties):

$$\underset{n \times 1}{u} = [\![u_i]\!]_i^{n \times 1}, \quad \underset{1 \times n}{v} = [\![v_j]\!]_j^{1 \times n}$$

The model variables are as follows:

1. r_{Mt} is the log-return of the market index at time t;

2. r_{it} is the log-return of the i-th stock at time t, included in the market index;

3. r_{St} is the log-return of the sentiment index at time t;

4. r_{Ft} is the risk free rate at time t.

We denote with N the stock number and with T the number of observations, so that $t = 1 \ldots T$.

For each return, the equivalent *excess return* is defined as:

$$r_{it}^* = r_{it} - r_{Ft} \tag{1}$$
$$r_{Mt}^* = r_{Mt} - r_{Ft} \tag{2}$$
$$r_{St}^* = r_{St} - r_{Ft} \tag{3}$$

To help the vectoring of models we define also:
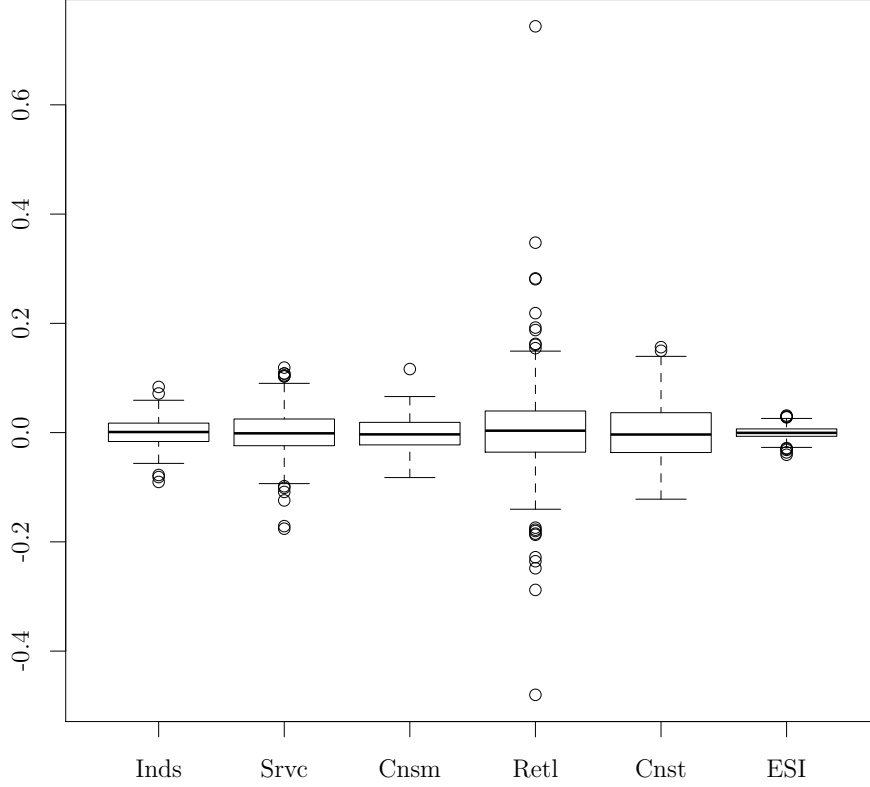
Sentiment Indices: boxplot summaries

Figure 1: Box-and-whisker plots showing Tukey's five number summary of the indices.

$$f_{1t} = r_{Mt}$$
$$f_{2t} = r_{St}$$

The variables f_{kt}^* are defined by analogy with excess returns.

We can now model r_{it} (and by analogy r_{it}^*) as a system of $N \times T$ equations:

$$r_{it} = \alpha_i + \beta_{1i} r_{Mt} + \beta_{2i} r_{St} + \epsilon_{it} \qquad (4)$$
$$= \alpha_i + \underset{1 \times K}{\beta'_{\bullet i}} \underset{K \times 1}{f_{\bullet t}} + \epsilon_{it} \qquad (5)$$

where $\beta_{\bullet i} = [\![\beta_{ki}]\!]_k^{K \times 1}$ and $f_{\bullet t} = [\![f_{kt}]\!]_k^{K \times 1}$.

The variables f_{kt} play the role of *common factors*, that is observable regressor variables which could explain the stock (excess) returns. In our analysis there are two common factors: the market factor and the sentiment factor; in spite of this, we find convenient denoting this number generically with K.

Note that, based on (4), we are assuming a constant intercept and slope in relation to time.
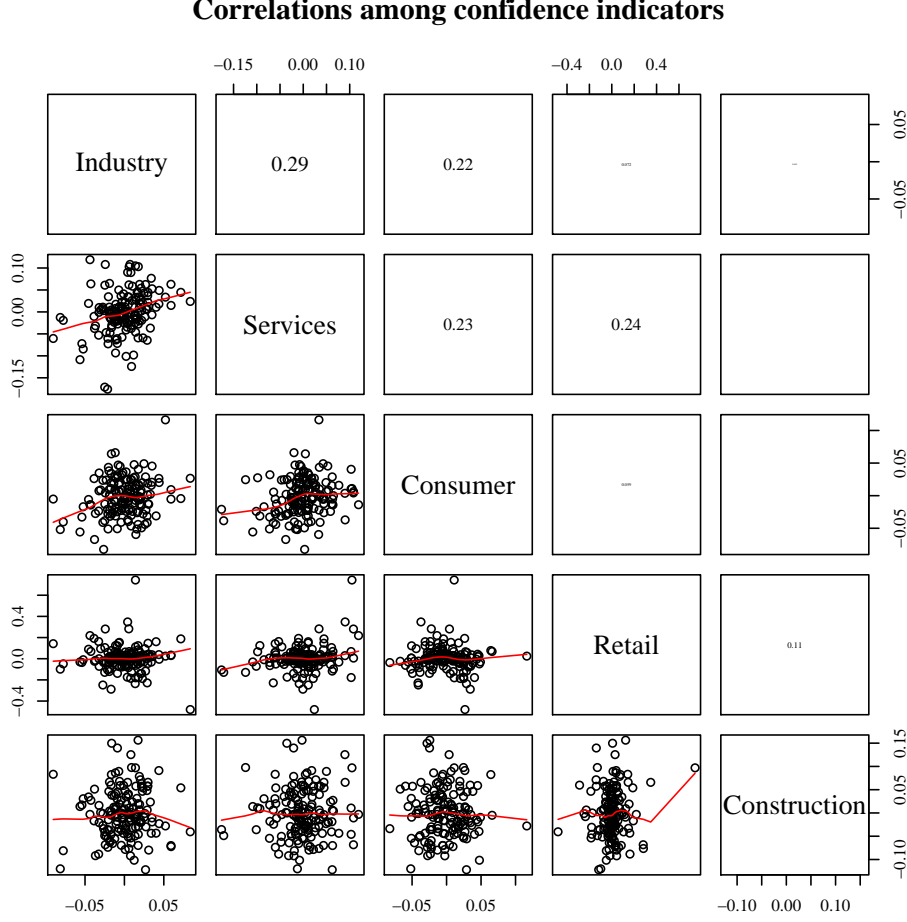
Correlations among confidence indicators

Figure 2: Cross correlation among confidence indicators. The upper triangle shows numerical values. To increase readability of the results, the more the correlation among styles the more the font dimension. The inferior triangle shows the same data in the form of dispersion plots. The fitted line (red) for each plot helps in assessing visually the degree of correlation between styles.

In equations (4) stock returns are in scalar form. For a given stock i, the equations (4) can be written in vector form as a system of N time series equations:

$$\underset{T \times 1}{y_{.i}} = \underset{T \times 1}{\mathbf{1}_T} \underset{1 \times 1}{\alpha_i} + \underset{T \times 1}{y_{.M}} \underset{1 \times 1}{\beta_{1i}} + \underset{T \times 1}{y_{.S}} \underset{1 \times 1}{\beta_{2i}} + \underset{T \times 1}{e_{.i}} \tag{6}$$

Here $y_{.i}$, $y_{.M}$, $y_{.S}$ are the observed time series respectively of the i-th stock, the market factor and the sentiment factor. We address the reader to appendix (A) for a more formal specification.

4.1 The Correlation Structure

Given the nature of the model, it is important to us the analysis of the correlation structure. To this end we find convenient to rewrite the model in a vectorised form. This entails: firstly stacking the equation in time series regression (6) as to obtain their matrix equivalent; secondly applying the "vec" operator.
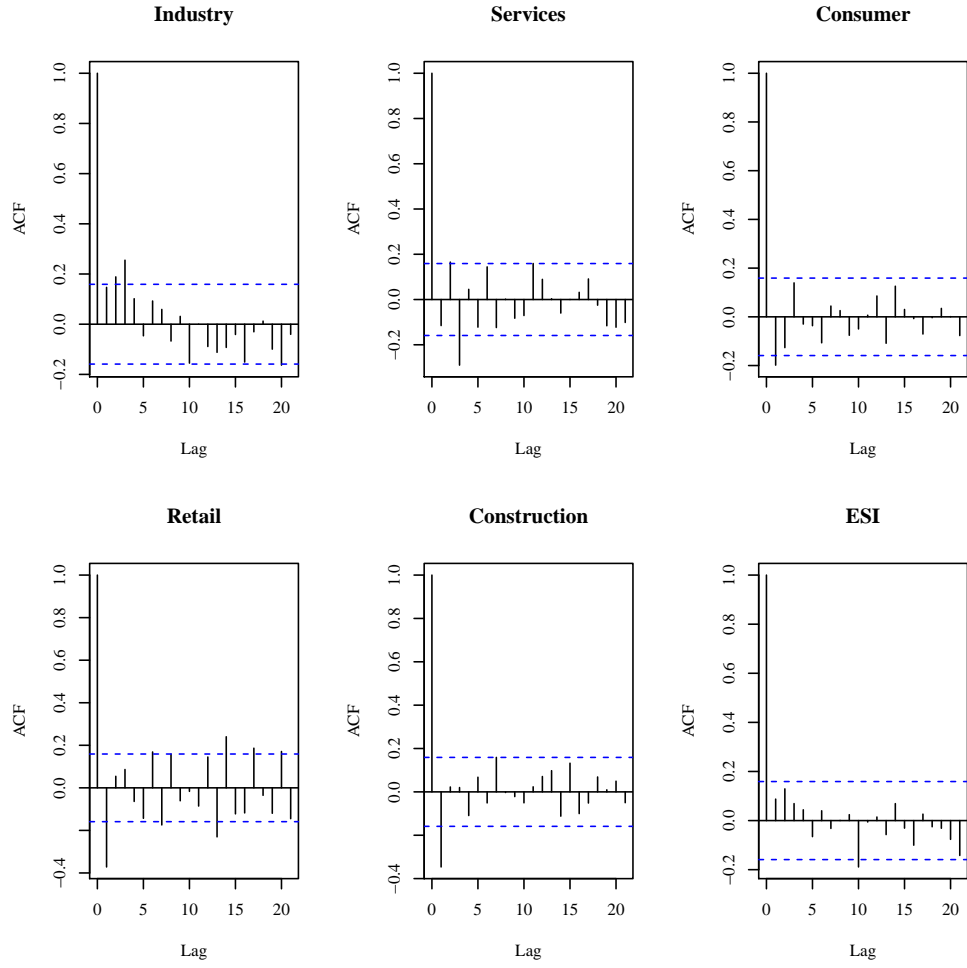
14

Figure 3: Correlograms of each confidence indicator time series. The auto-correlations allow us to asses memory effects, that is the degree of correlation of present values of the series with past ones. The correlations are reported for different time lags and they can be considered negligible when contained between the horizontal bars.

15

$$\underset{TN \times 1}{\underline{Y}} \; = \; \underset{TN \times N(K+1)}{\underline{Z}} \; \underset{N(K+1) \times 1}{\underline{\Gamma}} \; + \; \underset{TN \times 1}{\underline{E}} \tag{7}$$

While addressing the reader to appendix (A) for a formal discussion of the procedure, we observe here that $\underline{Y}$ is the vectorised return matrix, $\underline{\Gamma}$ is the vectorised coefficient matrix (including the intercept) and $\underline{Z}$ is the vectorised factor matrix (augmented with a proper 1-column vector). Clearly $\underline{E}$ is the error term for the model.

Given the vectorised-matrix model (7), the conditions concerning error term and factor correlations can be summarised as:

$$\text{Predeterminedness: } \; \mathbb{E}\left(\underline{e}_{h(ti)} \; \middle| \; \underset{N \times N(K+1)}{\underline{Z}_t} \right) = 0 \tag{8}$$

$$\text{Error correlations: } \; \mathbb{E}(\underline{e}_{h(t,i)} \underline{e}_{h(s,j)}) = \begin{cases} 0, & \text{if } t \neq s \\ \sigma_{ij}, & \text{otherwise} \end{cases} \tag{9}$$

where $e_{h(t,i)}$ is the $\underline{E}$ element equivalent to $\epsilon_{it)}$.

We will now discuss these conditions and show the construction of matrix $\underline{Z}_t$.

By the very definition of 'common factors' f_{kt}, it seems reasonable to think that they are uncorrelated with specific error terms. This assertion could possibly be false in some special case (i.e. for specific stocks with strong "systemic behaviour"), but it is supposedly true for the average stock. Formally this can be stated as:

$$\mathsf{cov}(f_{kt}, \epsilon_{it}) = 0 \tag{10}$$

If we look at (10) under another perspective, we can claim that factors have no power in explaining specific error terms. In place of the strong assumption of independence of residuals from past, present, and future factors' values, we could limit independence only to current observations: this is to say that factors, as regressors, are *predetermined* (rather than exogenous).

Eventually we replace (10) with the weaker condition:

$$\textbf{Predeterminedness: } \mathbb{E}\left(\epsilon_{it} \; \middle| \; \begin{bmatrix} r_{Mt} \\ r_{St} \end{bmatrix} \right) = \mathbb{E}\left(\epsilon_{it} \; \middle| \; [\![f_{kt}]\!]_k^{K \times 1} \right) = 0 \tag{11}$$

An important question is stock correlation. Of course disregarding stock correlation would allow us to deal with equations (6) individually. An independence (or null-correlation) assumption, while being easy for the calculations, appears too restrictive. The very fact that we take into account N equations/securities suggest that we need to take into account how they interact with each other. Since we intend to focus on portfolios, we can choose to disregard serial in favour of cross-sectional correlation. This conveys to the condition:

$$\textbf{Error Correlations: } \mathbb{E}(\epsilon_{it} \epsilon_{js}) = \begin{cases} 0, & \text{if } t \neq s \\ \sigma_{ij}, & \text{otherwise} \end{cases} \tag{12}$$

Both conditions (11) and (12) are expressed with respect to the scalar model (4); they can be restated with regard to vectorised-matrix model (7), obtaining (8) and (9). We address the reader to appendix (B) for the formal derivation. Here we observe that $\underline{Z}_t$ in (8) denotes the matrix obtained by stacking the N rows of $\underline{Z}$ containing the t-th row of Z.

4.2   Model Estimators

As showed in appendix (A), after proper transformations the $N \times T$ system of equations (4) is reduced to the linear model (7), with its related conditions for error terms.

As a general rule, since this model appears to show both heteroscedasticity and serial correlations, a simple least square (OLS) would be inefficient. Based on this a GLS approach would be necessary, which in turn would require the knowledge of the error covariance matrix:

$$\Omega = \mathsf{cov}(\underline{E}) = \mathbb{E}(\underline{E}\underline{E}') \tag{13}$$

The GLS regressor for the model (7), based on (13) is:

$$\hat{\Gamma}_{\mathsf{GLS}} = (\underline{Z}'\Omega^{-1}\underline{Z})^{-1}\underline{Z}'\Omega^{-1}\underline{Y} \tag{14}$$

On an empirical basis it could be anyway difficult to estimate (14), this would require some hypothesis concerning the error matrix, in fact Ω is $(TN)^2$, which exceeds observations available.

As an alternative approach (to GLS), we use a heteroscedasticity-corrected covariance matrix or *White adjusted*. Standard errors for regression coefficients obtained are therefore robust for non-constant variance.[7]

In order to implement the technique, we need to define a restriction matrix.[8]

Considering the vectorised coefficient matrix:

$$\underline{\Gamma} = \begin{bmatrix} \alpha_1 \\ \beta_{.1} \\ \alpha_2 \\ \beta_{.2} \\ \vdots \\ \alpha_N \\ \beta_{.N} \end{bmatrix}^{N(K+1) \times 1}$$

For each restriction to a given factor h we need to set $\beta_{hi} = 1$ for all firms. Define $\underline{\Gamma}(h, j)$ as the vector obtained setting in $\underline{\Gamma}$:

---

[7]See White 1980.

[8]In a restriction matrix (or hypothesis matrix) we have a row for each restriction applied to a coefficient. Each row is constituted of ones in the positions of restricted coefficient and zeros in the positions of unrestricted coefficients. We also need the vector of restricted values, which is constituted of a scalar for each of those restrictions; the scalar value is the value which the coefficient is constraint to. See e.g. Rencher 2003.

$$\beta_{ki} = 1 \text{ if } k = h, i = j$$
$$\beta_{ki} = 0 \text{ otherwise}$$

It follows that, if we want to test the hypothesis of factor h coefficient being null, we set:

$$\begin{bmatrix} \underline{\Gamma}(h,1) \\ \cdots \\ \underline{\Gamma}(h,N) \end{bmatrix}^{N \times N(K+1)} \underline{\Gamma} = \mathbf{0}_N$$

4.3   Implementation

The empirical application was developed using the open source environment and language R. The code relevant to the scope of this paper is presented in appendix (C) with detailed comments, to the benefit of other researchers interested to play with it.

As detailed in the previous section, tests are implemented using a heteroscedasticity-corrected coefficient covariance matrix (so called "White-corrected" matrix). The code is based on the R package "Car", maintained by John Fox and described in Fox and Weisberg S. 2011.

Table 4 shows the models obtained varying the sentiment indices, with and without intercepts. The single R-Squared for each model is presented connected to the vectorised multivariate model and the average of R-Squared's of the sub-models obtained by splitting the multivariate models into a multiple regression for each security. There is a minimal difference between the two methods. Intercept models are also tested against the hypothesis of the factor coefficients being null (restriction matrix is built as detailed in section 4.2). The p-values for these models appear highly significant.

The results in Table 4, while proving the effectiveness of the model, do not involve that there is an improvement or a significant contribution on behalf the sentiment factor. Therefore in Table 5 unrestricted models are tested against restricted versions without the sentiment factor (i.e. setting the sentiment coefficient as null) in order to check the contribution of psychological factors. It can be seen from the data here that many indices show an acceptable degree of statistical significance (below 0.05), while the industry index and the general ESI provide a noteworthy contribution. Most notably, as we can see from the table, the impact of the sentiment factors on returns is stronger when the test is run without the intercept, which proves coherent with the CAPM model.

5   Final Remarks on Empirical Investigation

In this study we address some behavioural biases connected with the CAPM model. Since the market factor may be unable to explain stock excess returns, which may also be affected by market sentiment, we augmented the standard model with this behavioural factor.

Dissimilarly to the majority of the research reviewed, based on proxies, we measured the sentiment with direct data based on the Joint Harmonised EU Programme of Business and Consumer Surveys, producing the European Sentiment Index.

	Mean R^2	R^2	F-stat	F p-value
Market only	0.36	0.37		
Int + Market only	0.36	0.40	55.68	1.06E-197
Mkt + Industry	0.36	0.37		
Int + Mkt + Industry	0.36	0.41	28.63	3.27E-187
Mkt + Services	0.36	0.37		
Int + Mkt + Services	0.36	0.40	28.53	1.43E-186
Mkt + Consumer	0.36	0.37		
Int + Mkt + Consumer	0.36	0.40	28.66	1.95E-187
Mkt + Retail	0.37	0.39		
Int + Mkt + Retail	0.37	0.42	31.21	1.70E-203
Mkt + Construction	0.36	0.37		
Int + Mkt + Construction	0.36	0.40	28.52	1.56E-186
Mkt + ESI	0.36	0.37		
Int + Mkt + ESI	0.36	0.40	28.15	3.43E-184

Table 4: Different sentiment models with fitting parameters, based on equation (7). Each model is obtained varying the sentiment factor and fitted both with and without intercept. The single R-Squared for each model is presented connected to the vectorised multivariate model and the average of R-Squared's of the sub-models obtained by splitting the multivariate models into a multiple regression for each security, as by equation (6). Fitting procedures uses the White adjusted standard errors.

We compared the standard model, without sentiment factors, against our sentiment augmented model to understand if the latter was able to give a better explanation of return dynamics.

Testing different stocks can easily bear different and possibly conflicting results; therefore care has been taken in aggregating test results in the presence of heteroscedasticity, to avoid deriving meaningless conclusions with respect to the whole sample. We described in details our assumptions and the related tests in the methodological section, particularly the approach we used was based on a heteroscedasticity-corrected covariance matrix (White adjusted).

Our analysis shows an improvement of the inference results when using sentiment indicators. Several sub-indices connected with ESI were tested, but the best results were obtained with the main ESI index.

Sampled blue chips, having a strong correlation with market movements, showed extremely low p-values when tested with sentiment indicators, and F-tests show that a significant contribution in explaining returns is given by psychological factors. Adding different kind and size of stocks (small or mid-caps) could possibly bring novel news to the subject matter.

References

Akerlof, G. A., and R. J. Shiller. 2010. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism (New in Paper)*. Princeton University Press.

Akhtari, M. 2011. "Reassessment of the weather effect: Stock prices and wall street weather". *Undergraduate Economic Review* 7 (1): 19.

Asness, C. S., J. A. Friedman, R. J. Krail, and J. M. Liew. 2000. "Style timing: Value versus growth". *The Journal of Portfolio Management* 26 (3): 50–60.

	F-stat	F p-value
Mkt + Industry	2.70	4.55E-05
Int + Mkt + Industry	1.72	2.12E-02
Mkt + Services	1.20	2.38E-01
Int + Mkt + Services	0.96	5.13E-01
Mkt + Consumer	1.63	3.46E-02
Int + Mkt + Consumer	0.40	9.93E-01
Mkt + Retail	1.64	3.32E-02
Int + Mkt + Retail	1.83	1.17E-02
Mkt + Construction	0.81	7.16E-01
Int + Mkt + Construction	0.82	6.93E-01
Mkt + ESI	4.20	4.94E-10
Int + Mkt + ESI	1.55	5.23E-02

Table 5: In order to measure the contribution of the sentiment factor to the models presented in Table 4, here the unrestricted version of these models (with the sentiment factor) are tested against restricted versions (without the sentiment factor). Only the ESI index contributes to a better explanation of the returns; besides consistent with CAPM theory, lower p-value are obtained in the absence of the intercept.

Baker, H. K., and J. R. Nofsinger. 2002. "Psychological biases of investors". *Financial Services Review* 11 (2): 97–116.

Baker, M., and J. C. Stein. 2004. "Market liquidity as a sentiment indicator". *Journal of Financial Markets* 7 (3): 271–299.

Baker, M., and J. Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns". *The Journal of Finance* 61 (4): 1645–1680.

— . 2007. *Investor sentiment in the stock market.*

Barberis, N., A. Shleifer, and R. Vishny. 1998. "A model of investor sentiment". *Journal of financial economics* 49 (3): 307–343.

Basu, D., C.-H. Hung, R. Oomen, and A. Stremme. 2006. "When to Pick the Losers: Do Sentiment Indicators Improve Dynamic Asset Allocation?" In *EFA 2006 Zurich Meetings Paper*.

Bernard, V. L. 1992. "Stock price reactions to earnings announcements: A summary of recent anomalous evidence and possible explanations".

Black, F. 1986. "Noise". *The Journal of Finance* 41 (3): 529–543.

Blanchard, O. J., and M. W. Watson. 1983. *Bubbles, rational expectations and financial markets.* Paul Wachtel, editor., Crisis in Economic and Financial Structure, Lexington, MA: D.C: Heathland Company 295-316.

Boido, C., and A. Fasano. 2005. "Calendar Anomalies: Daylights Savings Effect". *The ICFAI Journal of Behavioral Finance* 2(4) (4): 7–24.

Brown, G. W., and M. T. Cliff. 2004. "Investor sentiment and the near-term stock market". *Journal of Empirical Finance* 11 (1): 1–27.

Brown, G. 1999. "Volatility, sentiment, and noise traders". *Financial Analysts Journal* 55(2):82–90.

Chan, L. K., and J. Lakonishok. 2004. "Value and growth investing: Review and update". *Financial Analysts Journal* 60(1):71–86.

Chan, L. K., N. Jegadeesh, and J. Lakonishok. 1996. "Momentum strategies". *The Journal of Finance* 51 (5): 1681–1713.

Chan, L. K., J. Karceski, and J. Lakonishok. 2000. "New paradigm or same old hype in equity investing?" *Financial Analysts Journal*:23–36.

Cochrane, J. H. 2002. *Stocks as money: convenience yield and the tech-stock bubble.* Tech. rep. NBER working paper W8987.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam. 1998. "Investor Psychology and Security Market Under- and Over-reactions". *Journal of Finance* 53(6):1839–1886.

Das, S. R., and M. Y. Chen. 2007. "Yahoo! for Amazon: Sentiment extraction from small talk on the web". *Management Science* 53 (9): 1375–1388.

European Commission. Directorate-General for Economic and Financial Affairs. 2007. *Directorate-General for Economic and Financial Affairs. The Joint Harmonised EU Programme of Business and Consumer Surveys. User Guide (updated 4 July 2007).*

Fama, E. F., and K. R. French. 1988. "Permanent and temporary components of stock prices". *The Journal of Political Economy*:246–273.

Fisher, K. L., and M. Statman. 2000. "Investor sentiment and stock returns". *Financial Analysts Journal* 56(2):16–23.

Fox, J., and H. S. Weisberg S. 2011. *An R Companion to Applied Regression.* Sage Publications (second edition).

Hong, H., and J. C. Stein. 1999. "A unified theory of underreaction, momentum trading, and overreaction in asset markets". *The Journal of Finance* 54 (6): 2143–2184.

Jegadeesh, N., and S. Titman. 1993. "Returns to buying winners and selling losers: Implications for stock market efficiency". *The Journal of Finance* 48 (1): 65–91.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2003. "Winter Blues: A SAD Stock Market Cycle." *American Economic Review,* 93(1):1257–63.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2008. *Opposing Seasonalities in Treasury versus Equity Returns.* Tech. rep. University of Toronto.

Kim, T., and A. Ha. 2010. "Investor sentiment and market anomalies". In *23rd Australasian Finance and Banking Conference.*

Kothari, S., and J. Shanken. 1997. "Book-to-market, dividend yield, and expected market returns: A time-series analysis". *Journal of Financial Economics* 44 (2): 169–203.

Kyle, A. S. 1985. "Continuous auctions and insider trading". *Econometrica: Journal of the Econometric Society*:1315–1335.

Lawrence, E. R., G. McCabe, and A. J. Prakash. 2007. "Answering financial anomalies: Sentiment-based stock pricing". *The Journal of Behavioral Finance* 8 (3): 161–171.

Lo, A. W., and A. C. MacKinlay. 1990. "When are contrarian profits due to stock market overreaction?" *Review of Financial studies* 3 (2): 175–205.

Long, J. B., A. Shleifer, L. H. Summers, and R. J. . Waldmann. 1990. "Noise trader risk in financial markets," *Journal of Political Economy* 98(4):703–738.

Long, J, A. Shleifer, L. H. Summers, and R. J. Waldmann. 1989. "The size and incidence of the losses from noise trading". *The Journal of Finance* 44 (3): 681–696.

Neal, R., and S. M. Wheatley. 1998. "Do measures of investor sentiment predict returns?" *Journal of Financial and Quantitative Analysis* 33 (4).

Ofek, E., and M. Richardson. 2003. "Dotcom mania: The rise and fall of internet stock prices". *The Journal of Finance* 58 (3): 1113–1138.

Poterba, J. M., and L. H. Summers. 1988. "Mean reversion in stock prices: Evidence and implications". *Journal of Financial Economics* 22 (1): 27–59.

Qiu, L., and I. Welch. 2004. *Investor sentiment measures.* Tech. rep. National Bureau of Economic Research.

Rencher, A. 2003. *Methods of Multivariate Analysis.* Wiley Series in Probability and Statistics. Wiley.

Shiller, R. J. 1980. *Do stock prices move too much to be justified by subsequent changes in dividends?*

— . 2000. *The Irrational Exuberance.*

Shleifer, A. 2003. *Inefficient markets: An introduction to behavioral finance.* OUP Oxford.

Simon, D. P., and R. A. Wiggins. 2001. "S&P futures returns and contrary sentiment indicators". *Journal of Futures Markets* 21 (5): 447–462.

Statman, M. 2005. "Normal investors, then and now". *Financial Analysts Journal*:31–37.

Thaler, R. H. 1999. "The end of behavioral finance". *Financial Analysts Journal*:12–17.

Tirole, J. 1982. "On the possibility of speculation under rational expectations". *Econometrica: Journal of the Econometric Society*:1163–1181.

Tversky, A., D. Kahneman, et al. 1981. "The framing of decisions and the psychology of choice". *Science* 211 (4481): 453–458.

Wang, C. 2001. "Investor sentiment and return predictability in agricultural futures markets". *Journal of Futures Markets* 21 (10): 929–952.

White, H. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". *Econometrica: Journal of the Econometric Society*:817–838.

Appendices

A   Vectorisation Procedure

We recall equation (6) for formal variable specification.

$$\underset{T\times 1}{y_{.i}} = \underset{T\times 1}{\mathbf{1}_T}\ \underset{1\times 1}{\alpha_i} + \underset{T\times 1}{y_{.M}}\ \underset{1\times 1}{\beta_{1i}} + \underset{T\times 1}{y_{.S}}\ \underset{1\times 1}{\beta_{2i}} + \underset{T\times 1}{e_{.i}} \tag{15}$$

$$= \underset{T\times 1}{\mathbf{1}_T}\ \underset{1\times 1}{\alpha_i} + \underset{T\times K}{X}\ \underset{K\times 1}{\beta_i} + \underset{T\times 1}{e_{.i}} \tag{16}$$

where:

$$\underset{T\times 1}{y_{.\Lambda}} = [\![y_{t\Lambda} \mid y_{t\Lambda} = r_{\Lambda t}]\!]_t^{T\times 1} \text{ with } \Lambda = i, M \text{ or } S \tag{17}$$

$$\underset{T\times K}{X} = [\![x_{tk} \mid x_{tk} = f_{kt}]\!]_{tk}^{T\times K} \tag{18}$$

$$\underset{T\times 1}{e_{.i}} = [\![e_{ti} \mid e_{ti} = \epsilon_{it}]\!]_t^{T\times 1} \tag{19}$$

The model (4) can be written also in matrix form as:

$$\underset{T\times N}{Y} = \underset{T\times 1}{\mathbf{1}_T}\ \underset{1\times N}{\alpha} + \underset{T\times K}{X}\ \underset{K\times N}{B} + \underset{T\times N}{E} \tag{20}$$

where:

$$\underset{1\times N}{\alpha} = [\![\alpha_i]\!]_i^{1\times N} \tag{21}$$

$$\underset{T\times N}{Y} = [\![y_{ti} \mid y_{ti} = r_{it}]\!]_{ti}^{T\times N} \tag{22}$$

$$\underset{K\times N}{B} = [\![\beta_{ki}]\!]_{ki}^{K\times N} \tag{23}$$

$$\underset{T\times N}{E} = [\![e_{ti} \mid e_{ti} = \epsilon_{it}]\!]_{ti}^{T\times N} \tag{24}$$

It could be noted that, based on (20), the typical element of the matrix Y is:

$$y_{ti} = \alpha_i + \sum_k x_{tk}\beta_{ki} + e_{ti} \tag{25}$$

$$= \alpha_i + x_{t1}\beta_{1i} + x_{t2}\beta_{2i} + e_{ti} \tag{26}$$

If in (26) we substitute values from y_{ti}, x_{tk}, e_{ti} based resp. on definitions (22), (18), (24), we obtain:

$$r_{it} = \alpha_i + f_{1t}\beta_{1i} + f_{2t}\beta_{2i} + \epsilon_{it}$$

that is (4). It follows that (20) contains all the $N \times T$ equations (4) obtained by varying i, t.

Clearly estimating the matrix B and therefore its elements β_{ki} — by means of the single matrix equation (20) — could bring results different by the individual estimates of the same parameters obtained, instead, by means of the single equations (4). The latter method could be potentially less efficient since two estimates $\hat{\beta}_{ki}, \hat{\beta}_{hj}$, despite individually optimal (that is with regard to the equations (4) whom they refer to), are not necessarily such contemporaneously.

It seems wise to operate on the (20) and, to this end, we vectorise it. So we set preliminarily:

$$\underset{T\times(K+1)}{Z} = \left[\underset{T\times1}{\mathbf{1}_T} : \underset{T\times K}{X} \right] \tag{27}$$

$$\underset{(K+1)\times N}{\Gamma} = \begin{bmatrix} \underset{1\times N}{\alpha} \\ \underset{K\times N}{B} \end{bmatrix} \tag{28}$$

Therefore (20) can be written in matrix form as:

$$\underset{T\times N}{Y} = \underset{T\times(K+1)}{Z} \underset{(K+1)\times N}{\Gamma} + \underset{T\times N}{E} \tag{29}$$

We now apply the 'vec' operator:

$$\underset{TN\times1}{\text{vec}Y} = \underbrace{(\mathbf{I}_N \otimes Z)}_{TN\times N(K+1)} \underset{N(K+1)\times1}{\text{vec}\Gamma} + \underset{TN\times1}{\text{vec}E} \tag{30}$$

that is:

$$\underset{TN\times1}{\text{vec}Y} =$$

$$\overset{TN\times N(K+1)}{\begin{bmatrix} Z & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & Z & \cdots & \mathbf{O} \\ \vdots & \vdots & & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & Z \end{bmatrix}} \overset{N(K+1)\times1}{\begin{bmatrix} \alpha_1 \\ \beta_{.1} \\ \alpha_2 \\ \beta_{.2} \\ \vdots \\ \alpha_N \\ \beta_{.N} \end{bmatrix}} + \underset{TN\times1}{\text{vec}E} \tag{31}$$

Eventually it can be worth rewrite (30) as:

$$\underset{TN\times1}{\underline{Y}} = \underset{TN\times N(K+1)}{\underline{Z}} \underset{N(K+1)\times1}{\underline{\Gamma}} + \underset{TN\times1}{\underline{E}} \tag{32}$$

where $\underline{Y} = \text{vec}Y$, $\underline{Z} = \mathbf{I}_N \otimes Z$, $\underline{\Gamma} = \text{vec}\Gamma$, $\underline{E} = \text{vec}E$.

B Correlation Structure in Vectorised-Matrix form

We derive the correlation structure in vectorised form, i.e. with respect to (32).

As for error correlations, given (12), the condition (12) can be easily reformulated in terms of (29) as:

$$\mathbb{E}(\epsilon_{it}\epsilon_{js}) = \mathbb{E}(e_{it}e_{js}) = \begin{cases} 0, & \text{if } t \neq s \\ \sigma_{ij}, & \text{otherwise} \end{cases} \tag{33}$$

In order to refactor (12) for model (32), we set:

$$\underset{TN\times 1}{\underline{E}} = \underset{TN\times 1}{\mathsf{vec}E} = \left[\!\left[\underline{e}_h \mid e_{ti} = \underline{e}_{h(t,i)} = \underline{e}_{(i-1)T+t} \right]\!\right]_h^{TN\times 1} \tag{34}$$

Given (34), (33) becomes:

$$\mathbb{E}(e_{ti}e_{sj}) = \mathbb{E}(\underline{e}_{h(t,i)}\underline{e}_{h(s,j)}) = \begin{cases} 0, & \text{if } t \neq s \\ \sigma_{ij}, & \text{otherwise} \end{cases} \tag{35}$$

We now turn to predeterminedness. In order that we may express (11) in terms of (29), we set:

$$\underset{T\times(K+1)}{Z} = \left[\!\left[z_{tk} \mid z_{tk} = x_{tk} \vee z_{tk} = 1 \right]\!\right]_{tk}^{T\times(K+1)} \tag{36}$$

Now, (11) becomes:

$$\mathbb{E}\left(\epsilon_{it} \mid [\![f_{kt}]\!]_k^{K\times 1} \right) = \mathbb{E}\left(e_{ti} \mid [\![x_{tk}]\!]_k^{K\times 1} \right) = \mathbb{E}\left(e_{ti} \mid \begin{bmatrix} 1 \\ [\![x_{tk}]\!]_k^{K\times 1} \end{bmatrix} \right) \tag{37}$$

$$= \mathbb{E}\left(e_{ti} \mid [\![z_{tk}]\!]_k^{(K+1)\times 1} \right) = 0$$

As for the model (32), the $\underline{Z}$ regressor is a bit more cumbersome to work with. Looking at the (31), we see that the Z matrix appears in N row of $\underline{Z}$, so the t-th row of Z appears N times in $\underline{Z}$. We denote with $\underline{Z}_t$ the matrix obtained by stacking the N rows of $\underline{Z}$ containing the t-th row of Z (plus the adjacent zeros). The predeterminedness condition is therefore:

$$\mathbb{E}\left(e_{ti} \mid [\![z_{tk}]\!]_k^{K\times 1} \right) = \mathbb{E}\left(\underline{e}_{h(ti)} \mid \underset{N\times N(K+1)}{\underline{Z}_t} \right) = 0 \tag{38}$$

C Main R Code

This code could be omitted from printed paper. It can be anyway useful for the review process or possibly made available on request to interested readers.

```r
#===========
# Models
#==========

#Data description (zoo objects):
#sents  sentiment indices
#mark   market index
#secs   security prices
#free   risk free rate


# Some stat convenience funcs
#============================

library(car)
#
#Define the 'vec' operator
vec=function(x) as.matrix (c(as.matrix(x)))
#

#Multivariate Linear model
# int = T/F: intercept
mod=function(Y, X, int ){

  #Space x Time
  N= NCOL(Y)
  T= NROW(Y) #=NROW(X)

  #Vectoring
  vecY=vec(Y)

  #Add intercept
  if(int)
      X = merge ( zoo(1, index(X)), X)

  #Kronecker (I_N, X)
  ubarX=diag(N) %x% X


  #Here intercept is implicit in ubarX
  lm=lm(vecY ~ 0 + ubarX)
  attr(lm,"rsquared") <- rs(vecY,ubarX)
  lm
}

#Get rsquared for model with one security only
#intended for use with "apply" for average rsquared
rs=function(y,X){
#here the intercept has to be set as 1-col in design matrix

  #Has a constant first column?
  int=FALSE
  p=NCOL(X)
```

```r
    if(all (X[,1]==X[1,1])){
        X = merge ( zoo(1, index(X)), X)
        int=TRUE
    }


    #OLS
    hatBeta= solve( t(X) %*% X )  %*% t(X) %*% y

    y.hat=X %*% hatBeta
    mss = if(int)
                sum((y.hat - mean(y.hat))^2)
            else sum(y.hat^2)

    e=y-y.hat
    rss = sum(e^2)

    #r2 non-adjusted
    mss/(mss + rss)

}




#Create hypothesis matr. (one factor restriction for all firms)
hypMat=function(firmCount,
            afactCount, #factor count including alpha
            restFact){  #factor to constrain including alpha

    #zero placeholder matrix
    hm=matrix(
        rep(0, firmCount*  afactCount*firmCount)
        , nrow=firmCount)

    #Elements subject to constraint
    cel=cbind( 1:firmCount,
        seq(restFact, afactCount*firmCount, by=afactCount) )

    #Apply constraint
    hm[cel]=1
    hm
}

#Get model parameters using white.adjust
# int = T/F: intercept
getModPars=function(Y, X, int){

    unrestricted = mod(Y, X, int)
    s=summary(unrestricted)

    #Calculate F for intercept models
    if(int){
        #Define hypothesis/restriction matrix
        K=NCOL (X)
        N=NCOL (Y)
        hm=vector()
```

```r
    for(i in 2:(K+1) ){
       h=hypMat(N, K+1, i)
       hm =rbind(hm,h)
    }
    rhs=rep(0, K*N)

    l=linearHypothesis(unrestricted, hm, rhs,
               test="F", white.adjust=TRUE)
    f.stat= l$F[2]
    f.pv  = l$"Pr(>F)"[2]
  }else{
    f.stat= NA
    f.pv  = NA
  }

  #data.frame( s£r.squared, f.stat, f.pv)
  data.frame( mean(apply(Y,2,rs,X)), attr(unrestricted , 'rsquared'),
             f.stat, f.pv )


}


#Get mod parameters (white adj) for mod comparisons
# int = T/F: intercept
getModPars.anova=function(Y, X, int){

  N=NCOL (Y)
  K=NCOL (X)
  unrestricted = mod(Y, X, int)


  #Exclude sentiment for intercept models
  if(int){
    #Define hypothesis/restriction matrix
    hm=hypMat(N, K+1, K+1)
    rhs=rep(0, N)
    l=linearHypothesis(unrestricted, hm, rhs,
               test="F", white.adjust=TRUE)
     f.stat= l$F[2]
     f.pv  = l$"Pr(>F)"[2]
  }else{   #Exclude sentiment for NO intercept models
    hm=hypMat(N, K, K)
    rhs=rep(0, N )
    l=linearHypothesis(unrestricted, hm, rhs,
               test="F", white.adjust=TRUE)
     f.stat= l$F[2]
     f.pv  = l$"Pr(>F)"[2]
  }

  data.frame( f.stat, f.pv)


}
#

#Stack model pars
#Run model, get pars and stack'em for nice printing
parStack=function(Y, X, mod.name, pars.stack, do.anova=FALSE){
```

```r
    #int = T/F: intercept
    for(int in c(FALSE, TRUE)) {
        if(do.anova) p= getModPars.anova(Y, X, int)
        else p= getModPars(Y, X, int)
        if (int) row.names(p)= paste("Int +", mod.name )
        else    row.names(p)= mod.name
        pars.stack= rbind (pars.stack, p)
    }
    pars.stack
}


#Stack model pars (compare models)
#Run model, get pars and stack'em for nice printing
parStack.anova=function(Y, X, mod.name, pars.stack){
    parStack(Y, X, mod.name, pars.stack, do.anova=TRUE)
}
#

#Set regression variables
#========================

#Log returns (assume a zoo series)
ret=function(x){
    x1=x[-1,]
    x0=x[-NROW(x),]
    zoo(log(coredata(x1)+100)- log(coredata(x0)+100)
        , index(x)[-1])
}

Y  =ret(secs)
X.m=ret(mark)
#Scale sents component up to 100
X=sents
X[,-NROW(sents)] = X[,-NROW(sents)]+100
X.s=ret(X)

#Get Excess Returns
Y  = Y-(free)[-1]/12
X.m=X.m-(free)[-1]/12
X.s=X.s-(free)[-1]/12


#Run models and stack parmeters
#  with/out sentiment and interc.
#==============================

  pars=data.frame()

  #Market only
  pars= parStack(Y,  X.m,  "Market only", pars)

  #Market + one sentiment at time
  scount= NCOL(X.s)
  for(i in 1:scount) {
    #i=1
    sname=paste("Mkt +", names(X.s)[i])
```

```
    X= merge(X.m, X.s[,i], all=F)
    pars= parStack(Y, X, sname, pars)
  }#===end loop sentiments


#Check removing Sent_i from Mark+Sent_i and Int+Mark+Sent_i
  pars.anova = data.frame()
  for(i in 1:scount) {
    #i=1
    sname=paste("Mkt +", names(X.s)[i])
    X= merge(X.m, X.s[,i], all=F)
    pars.anova= parStack.anova(Y, X, sname, pars.anova)
  }


#===end Run models
```