Investor Sentiment as a Predictor of Market Returns

Kim Kaivanto, Peng Zhang*

Department of Economics, Lancaster University

Abstract

Whereas the predictability of market returns and the deviation of market returns from fundamentals have been investigated individually in a large number of studies, few if any have explicitly modelled both jointly.

The present work aims to fill the gap between the aforementioned two research streams by explicitly introducing investor sentiment as a predictor of market returns. In so doing, we explore (i) the potential for adding to the market return predictability literature, as well as (ii) the possibility for extending the literature on the influence of investor sentiment from mainly individual and portfolio returns to the aggregate-level returns.

By using major investor sentiment indicators from the literature as predictors of market returns, we implement comparison among different indicators. Our results show that the indicators are not all equally informative. Some indicators better predict returns than the others. Evidence is also in line with the statement in the literature that some indicators affect returns in a lagged way.

We also consider more complex dynamics between investor sentiment indicators and market returns by conducting Granger causality tests. We find Granger causality at neither, either, or both directions for different indicators. In general, the dynamics between indicators and market returns are not uniform.

Keywords: investor sentiment; market return predictability; long-horizon regression; bootstrap

*Correspondence to: Peng Zhang, Lancaster University Management School, Lancaster, LA1 4YX, UK E-mail: p.zhang@lancaster.ac.uk

1 Introduction

This article is concerned with two topics in asset pricing theory which independently have received a large amount of attention from financial economists over the past 30 years. One is the predictability of market returns; the other is the deviation of asset prices from fundamental values. Whereas the various streams of traditional asset pricing theory – mean-variance portfolio-based pricing, CAPM, APT, ICAPM, diffusion-based pricing, and so on – have been shown to be consistent with each other and moreover can in fact be derived from a unified Stochastic Discount Factor framework, empirical tests of the models often suggest that the programme to develop asset pricing theory is far from complete. Since the 1980s two questions have persistently vexed the profession: (i) "Are market returns (which in theory should reflect only systematic risk) in fact predictable?", and (ii) "Are asset prices determined by fundamentals (up to the point where the marginal benefits of acting on information do not exceed the marginal costs of doing so)?".

On the one hand, contrary to the traditional view that market return should be random, predictive power has been found within a variety of factors – e.g. past returns, dividend-price ratio, dividend-earning ratio etc. – to explain the returns, particularly over long-horizons. Despite several criticisms have been raised regarding sample bias, data snooping and long-horizon bias, the predictability is still commonly cited in empirical finance. Some evidence remains even after adjustments being made to eliminate biases.

On the other hand, historical data have raised a noticeable amount of stylised "puzzles", among which are excess volatility, mean-reversion in returns, extraordinary equity premium, and arbitrage opportunities in the market etc. All these findings suggest that asset prices are often apart from fundamental values. More recent theory often considers this deviation and attributes the phenomenon to the effects of "investor sentiment". Theoretical work on the role of investor sentiment in affecting asset pricing has gained significant progress since 1990's (see e.g. DeLong et al. 1990, Scheinkman and Xiong 2003). Some empirical support has been found, with a few puzzles (partially) answered.

However, despite the large number of studies devoted to investigations of both topics, it seems that explicit consideration of connecting them together has not been widely recognised. As argued by DeLong et al. (1990), investor sentiment should be "marketwide rather than idiosyncratic" (p.707). If investor sentiment affects asset prices in such a systematic way, it might be possible to find predictability in market return with investor sentiment. We aim to fill in the gap between the two topics by an attempt to explain market returns using the investor sentiment as a predictor explicitly. In so doing, we explore (i) the potential for adding to the market return predictability literature, as well as (ii) the possibility for extending the literature on the influence of investor sentiment from mainly individual and portfolio returns to the aggregate-level returns. This is the first preoccupation and contribution of this article.

The second preoccupation and contribution of this article is to conduct a comprehensive investigation on the (possibly different) effects of major investor sentiment indicators in existing literature on asset prices, in a unified framework within the same sample period. Due to the fundamental difficulty of economics in matching data with concepts, no perfect indicator of investor sentiment is available. Most of existing empirical studies on the role of investor sentiment in asset pricing focus on only one or a few (typically no more than three) particular indicators and individual or portfolio returns, and implement the analysis within different models across different sample periods. As a result one cannot easily compare findings from different studies and draw a general conclusion. Moreover, theories have not predicted the time length over which investor sentiment may affect asset prices. Only very limited indicators have been studied regarding this dimension of the topic[1]. Our approach makes it possible to run comparison among different indicators over different horizons, and our results show that the indicators do not have equal predictive power. Some indicators better predict returns than the others. Evidence is also in line with the statement in the literature[2] that some indicators affect returns in a lagged way.

Last but not least, we consider more complex dynamics between investor sentiment indicators and market returns by implementing Granger causality tests. Again our results show that the indicators are not all equally informative. We find Granger causality at neither, either, or both directions for different indicators. In general, the dynamics between indicators and market returns are not uniform. This is our third contribution.

Our study shares some investor sentiment indicators with Baker and Wurgler (2006, 2007). Particularly, we adopt six indicators and four indices from Jeffrey Wurgler's data library. As pointed out by Baker and Wurgler (2006, 2007), the six indicators are also well studied in and collected from the literature[3]. Our sample period (1978:01 - 2007:12) is a subsample of the period used in Baker and Wurgler (2007). However their analysis only uses the six indicators to generate indices and they test the predictability of only the sentiment indices in several portfolio returns in a conditional asset pricing model, while we examine each indicator and each index indicator separately and focus on the predictability of market return. We also include and analyse survey data as an additional sentiment indicator.

Neal and Wheatley (1998) and Brown and Cliff (2005) also conduct long-horizon analysis with sentiment indicators. However both studies use only particular indicators and also look at their impact on portfolio returns within conditional asset pricing models rather than market returns. We follow the single factor prediction model in Fama and French (1988b) to test the predictability in market returns, and also consider adding in the first-order lagged return as an additional regressor inspired by Campbell and Shiller (1988), Jegadeesh(1991), and Hodrick (1991) etc.

The rest of this article is structured as follows. Part 2 reviews the main strands of literature on market return predictability and the deviation of asset prices from fundamental values. It also summarises the major indicators of investor sentiment that will be used in the empirical study of this article.. Part 3

[1] The exceptions include Brown and Cliff (2005) using survey data and Neal and Wheatly (1998) using closed-end fund discount data.

[2] See, e.g. Baker and Wurgler (2006, 2007).

[3] E.g. Ibbotson et al. (1994), Neal and Wheatley (1998), Lakonishok et al. (1994), Baker and Wurgler (2000) etc.

introduces data. Part 4 describes methodology. Part 5 studies the empirical results from single factor models. Part 6 studies the empirical results from double factor models. Part 7 examines more complex dynamics between market returns and investor sentiment indicators. Part 8 concludes.

2 Literature review

2.1 Market return predictability

Contrast to the traditional view that market prices follow a random walk, financial economists have started to search for evidence of market return predictability since the 1980s. Assuming a price process consisting of a random walk and a slowly decaying part, Summers (1986) points out that traditional tests on random walk have very low power. Summers' assumption has been verified through tests on the variance-ratio of market returns. Lo and MacKinlay (1988) find positive autocorrelation at short horizon, while Poterba and Summers (1988) find that the autocorrelation turns negative over longer horizons.

Fama and French (1988a) point out that the time decaying component in price can be resulted from both an irrational market where prices can deviate from fundamental values temporarily and a rational market where prices are generated by time-varying expected returns. They also argue that the presence of this component implies predictability through time. They further suggest using long-horizon tests to better capture this slowly decaying component (though at the cost of losing a degree of statistical precision) and find U-shaped first-order autocorrelations over horizons. Jegadeesh (1991) regresses one-period market return over its lagged values and also gets similar evidence on predictability.

Besides past returns, predictability has also been found with other predictors, particularly price-related ratios. Fama and French (1988b, 1989) show that returns can be predicted by dividend-price ratio and earning-price ratio. Campbell and Shiller (1988) follow a VAR approach and also found predictive power in dividend-price ratio. Hodrick (1991) compares the performance of alternative models and again confirms significant predictive power in dividend-price ratio.

Given the importance of long-horizon analysis in this literature, it is worth noting that Campbell (2001) shows that long-horizon analysis has higher power than short-horizon analysis against persistent alternatives like the model in Summers (1986). Campbell also verifies the point in Summers (1986) that short-horizon analysis cannot detect predictability if prices contain a persistent component.

2.2 Deviation of asset prices from fundamental values

It has been three decades since pioneering financial economists first observed the deviations of asset prices from fundamental values through various asset pricing puzzles. Mehra and Prescott (1985) find the well-known equity premium puzzle using U.S. data from 1889 through 1978. Shiller (1981a, b) finds that historical volatility in stock market is too high to be explained by new information about future returns. The evidence found in Fama and French (1988a) on mean-reversion of prices especially for 3-5-year returns is also in line with the hypothesis that prices deviate from fundamental values. Poterba and Summers (1988) have also drawn similar conclusions, and in the meantime verified the findings in Shiller (1981a, b) within a different approach. Another asset pricing anomaly is the closed-end fund discount, where the Law of One Price is violated for extended periods of time, resisting explanations by agency cost, liquidity effects, and tax frictions[4]. All these findings call for further empirical and theoretical work to improve our understanding of price determinations.

To address these empirical findings, there have been some general theoretical developments on asset pricing. At the early stages, Black (1986) considers "noise" as an even more causal factor to price changes than fundamental news. Shiller et al. (1984) study a social-psychological role of "fashions" in financial markets, and furthermore introduce unsophiticated investors. The finding in Summers (1986) suggests difficulty in detecting pricing errors, and as a result implies that exploiting the errors is risky and hence arbitrages will be limited. More systematic progress has been achieved since 1990s. DeLong et al. (1990) have established a formal system of how deviation of prices from fundamental values can be generated. As long as there exist noise traders and finite investment horizons, rational arbitrageurs will face a new systematic risk resulting from the uncertainty about noise traders' beliefs, while market prices may diverge significantly from fundamental values. Campbell and Kyle (1993) also build a model that explicitly studies the interaction between noise traders and rational investors and offer an equilibrium foundation for the findings in Shiller (1981a, b). Scheinkman and Xiong (2003) focus on the effect of overconfidence in a market with short-sale constraints, and show how a bubble is likely to be generated. There is also a stream of studies on pricing errors due to investor over/underreactions to new information[5].

To summarise, we follow recent convention and use the terminology "investor sentiment" to represent systematic biases in beliefs (biases held by a large class of investors which do not net out). While there is no standard and universally accepted formal definition for investor sentiment, it can refer to any factors that will lead to incorrect estimation of the fundamental values of assets, at either individual asset or aggregate market level. Therefore it includes trading on pseudo information, incorrect expectations of future returns, as well as flawed techniques for calculating fundamentals.

2.3 Investor sentiment indicators

As stated in Baker and Wurgler (2007), although investor sentiment is not straightforward to measure, there is no particular theoretical reason why we cannot find imperfect but useful indicators. The literature has studied both direct and indirect indicators. Direct indicators consist of survey instruments that gauge

[4]See e.g. Malkiel (1977), Herzfeld (1980), and Brickley and Schallheim (1985) for early studies.

[5]See e.g. Tversky and Kahneman (1984) for a psychological explanation, DeBondt and Thaler (1985) for overreactions, Bernard and Thomas (1990) for underreactions, and Barberis et al. (1998) for a general investigation

investor attitudes directly, and indirect indicators include various variables related to investor sentiment. In this section I will briefly introduce the major indicators of investor sentiment that will be used in the empirical study of this article.

Survey measures – The most straightforward indicator of sentiment is from survey data. Shiller (1999) suggests that the Yale School of Management Stock Market Confidence Indices can reflect the attitudes of institutional investors. Qiu and Welch (2006) show that data from the UBS/Gallup surveys can explain returns, particularly small stock returns and returns of stocks held disproportionately by retail investors. Similar findings have also been found in Lemmon and Portniaguina (2006) with data from both the Index of Consumer Confidence and the University of Michigan Consumer Confidence Index. Brown and Cliff (2005) find significant explanatory power in the Investors Intelligence survey data over long horizons.

Closed-end fund discount – Zweig (1973) uses the discount to verify that prices are likely to deviate from fundamental values when "noise" is present. DeLong et al. (1990) attribute the discount to the fact that closed-end funds are mainly held by individual investors and that the noise brought by these investors will lead to an extra risk premium. Lee et al. (1991), Neal and Wheatley (1998) and Swaminathan (1996) find evidence that the discount is a measure of investor sentiment and can help explain asset returns.

Market liquidity – Empirical studies have long found coexistence of higher liquidity and lower future returns[6]. Baker and Stein (2004) argue that liquidity provides an indicator of the presence or absence of irrational investors who face short-sale constraints and are active only in optimism. Using market turnover and equity issuance as proxies of liquidity, they find predictability in returns. Scheinkman and Xiong's work (2003) also points out the link between sentiment and market liquidity.

IPO related data – There have been several studies linking IPO with investor sentiment. Specifically, both IPO volume and IPO first-day return can be viewed as indicators of investor sentiment. For instance, Lee et al. (1991) find evidence that more IPOs happen when investor sentiment is high. Ljungqvist et al. (2006) show that sentiment can lead to IPO underpricing and hence cause high returns after IPO date.

New equity issuance – Given that IPO is just one measure of equity financing, a more general indicator of investor sentiment can be measured by the fraction of equity issuance to total asset issuance. Baker and Wurgler (2000) find a negative relationship between equity issuance and market returns, and attribute this relationship to issuers shifting between equity and debt to get lower cost of financing.

Dividend premium – Baker and Wurgler (2004a, b) find that the dividend premium, defined as the difference between the average market-to-book values of dividend-paying and dividend-nonpaying stocks, is highly correlated with investor demand for dividends. It has been argued that since dividend-paying equities have characteristics like coupon bonds, they represent "safety" compared to dividend-nonpaying equities. As a result Baker and Wurgler (2006, 2007) argue that when investors perceive high risk level and look for safty, dividend premium will be higher.

[6] See, for instance, Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), and Brennan et al. (1998).

3 Data

We use real NYSE index returns to represent the market return. Both equal-weighted and value-weighted monthly index returns on NYSE are obtained from the Center for Research in Security Prices (CRSP) and then adjusted with the Consumer Price Index inflation rate. With regard to the choice of investor sentiment indicators as predictors, our decision is mainly constrained from data availability. Most of the indicators summarised in Part 2 are only available for quite limited lengths of period in the past and data frequencies are not uniform. In balancing between the number of indicators investigated and the sample size for data, we choose eleven monthly indicators from January 1978 through December 2007. The indicators chosen are introduced in the following subsections.

3.1 Direct sentiment measures

For the survey data as direct measures of investor sentiment, we choose the Index of Consumer Sentiment from the University of Michigan Consumer Confidence Index (UMCCI). Since the main constraint regarding sample size in our analysis comes from the availability of survey data, we give it the highest priority when choosing among different surveys. Although the UMCCI survey has not been designed to reflect directly the investor sentiment in asset markets but rather consumer confidence in general, it can provide the earliest survey data and therefore expand our sample size as much as possible. The survey started from as early as 1952 and monthly data became available from January 1978.

The Michigan Index of Consumer Sentiment ($MICS$) is calculated as a linear combination of a constant and five scores from survey questions. The five survey questions include people's perceptions of changes in their financial situation in the last 12 months, together with the expected changes in their financial situation, in the general financial condition of the country, in the unemployment and depression condition, and in major household consumptions in the next year. Each month the survey has been sent to different households in 48 US States and the District of Columbia. The households for each monthly sample are drawn according to a rotating panel consisting of 40% households from the sample interviewed six months earlier and 60% new households. Each household is interviewed no more than twice. For the period from January 1978 through December 2007, the sample size of monthly interviews varies from 492 to 1459, but has stabilised at around 500 since 1988. More detailed information about the survey can be found at http://www.sca.isr.umich.edu/.

3.2 Indirect sentiment measures

We also use the following six indirect indicators of investor sentiment from Jeffrey Wurgler's online data library, together with four sentiment indices constructed in Baker and Wurgler (2007). The six indirect indicators include closed-end fund discount ($CEFD$), NYSE turnover ($TURN$), IPO volume ($NIPO$),

IPO first day return ($RIPO$), net equity issuance fraction in total issuance ($NEIF$), and dividend premium ($PDND$). The index indicators include two level sentiment indices ($SENT$ and $SENT^{\perp}$) and two difference sentiment indices ($DSENT$ and $DSENT^{\perp}$). The indicators are calculated as followings.

The closed-end fund discount ($CEFD$) is calculated as the average percentage difference between the net asset values of closed-end fund stock shares in the open market and the prices of the closed-end funds.

NYSE turnover ($TURN$) is obtained by calculating the natural logarithm of the ratio of reported share volume over average shares listed on NYSE.

IPO volume number ($NIPO$) is the number of IPOs in the month; IPO first day return ($RIPO$) is the average first day percentage return of all IPOs in the month.

Net equity issuance fraction ($NEIF$) is the proportion of new equity issuance out of total issuance of equity and debt.

Dividend premium ($PDND$) is calculated as the log difference of the value-weighted average market-to-book ratios of dividend payers and non-payers.

Level sentiment index ($SENT$) is based on the first principal components of the six (standardised[7]) indirect indicators.

Orthogonalised level sentiment index ($SENT^{\perp}$) is based on the first principal components of the six (standardised) indirect indicators, with the six indicators being orthogonalised with respect to eight macroeconomic variables.

Sentiment difference index ($DSENT$) is based on the first principal components of changes (first-order differences) in the six (standardised) indirect indicators.

Orthogonalised difference sentiment index ($DSENT^{\perp}$) is based on the first principal components of changes in the six (standardised) indirect indicators, with the six indicators being orthogonalised with respect to eight macroeconomic variables.

As we aim to compare the performance of different indicators, we match the sample period for all the indicators from January 1978 to December 2007. This leaves us 360 observations for all the indicators, except several missing data for IPO first-day returns where no IPO is present in the month (in which case we use 0 instead).

3.3 Orthogonalised sentiment indicators

It has been argued that sentiment indicators contain information reflecting not only investor sentiment but also macroeconomic fundamentals. This idea suggests that in order to exclude the component caused by fundamentals from the data, all the indicators should be orthogonalised with respect to fundamental variables before they can be used in further analysis. Earlier studies employing this idea include Baker and Wurgler (2006, 2007), Brown and Cliff (2005), and Neal and Wheatley (1998) among others. In

[7] Standardisations in calculating $SENT$ $SENT^{\perp}$ $DSENT$ and $DSENT^{\perp}$ mean that each monthly observation value of a sentiment indicator is subtracted by its sample mean and then divided by its sample standard deviation.

this approach, each indicator is regressed on a group of fundamental variables and the residuals unexplained by fundamentals are recorded as the proposed orthogonalised sentiment indicators. To choose the fundamental variables, we adopt two sets of variables – the first set includes growth in industrial production, real growth in durable, nondurable, service and total consumptions, growth in employment, cpi, and an NBER recession indicator dummy from Baker and Wurgler (2006, 2007) following a consumption-determined asset pricing idea; the second set includes 1-month real US Treasury bill return, the difference between 3-month and 1-month real US treasury bill returns, the difference between 10-year and 3-month real US treasury bill returns, and the default spread between yields on Moody's Baa and Aaa corporate bonds from Brown and Cliff (2005) and Neal and Wheatly (1998) following the conditional asset pricing idea. The data for the first variable set come from Jeffrey Wurgler's online data library, and the data for the second variable set come from Federal Reserve Bank of US. With regard to the question that whether these two ideas are both necessary, F tests show that the two sets of variables are both jointly significant in explaining all eleven sentiment indicators. Information criteria including Akaike Information Criterion (AIC), Schwarz/Bayesian Criterion (BIC) and Hannan-Quinn Criterion (HQC) also suggest both sets to be included instead of adopting either single set.

We use two parallel approaches to generate the orthogonalised sentiment indicators. In the first approach all twelve fundamental variables are used and therefore each sentiment indicator is orthogonalised with the same fundamentals excluded. In the second approach each sentiment indicator is orthogonalised with only those of the twelve fundamental variables that are significant in explaining this indicator. The latter approach allows sentiment indicators to be orthogonalised with different fundamentals excluded. In later analysis we use "orthogonalised indicator (all)" for the data generated from the former method and "orthogonalised indicator (significant)" for the data generated from the latter method.

3.4 Data preliminary

Table 1 summarises the sample characteristics of the original sentiment indicators. Excess values are reported for kurtosis. The skewness values of all first nine sentiment indicators except $MICS$ and $CEFD$ show that investor sentiment is right skewed[8], suggesting fat tail for bullish investor sentiment in general[9]. All sentiment indicators except $MICS$ have positive excess kurtosis values, i.e. their distributions are more peaked compared to the Gaussian distribution.

Table 2 and 3 summarise the sample characteristics of the orthogonalised sentiment indicators with all twelve fundamental variables and with only those of the twelve fundamental variables that are significant

[8] $CEFD$ and $PDND$ are supposedly negatively correlated to investor sentiment whilst all other variables are positively correlated to sentiment. If the correlations were perfect, the skewness for distribution of investor sentiment would be opposite to those of $CEFD$ and $PDND$ and be the same as those of the other indicators. The predicted correlations between $DSENT$ and $DSENT^{\perp}$ and investor sentiment are not clear in the theoretical or empirical literature.

[9] One may argue that several indicators including $TURN$, $NIPO$, $RIPO$ and $NEIF$ are bounded above zero so the evidence here may not indeed imply fat tail for bullish investor sentiment. However the same conclusion of fat tail for bullish investor sentiment can be drawn from Table 2 and 3, where the orthogonalised indicators are not bounded above zero. We consider the consistent evidence through Table 1 to 3 as evidence that the implication is not a result of truncation in data.

in explaining this indicator. Again excess values are reported for kurtosis. Because the orthogonalised indicators come from the residuals of regressing original indicators on fundamental variables, the means of the othogonalised indicators are all extremely close to 0. Compared to Table 1, the signs of skewness and excess kurtosis statistics for every sentiment indicator remain unchanged after orthogonalisation except those of $CEFD$ in both Table 2 and 3 and skeness of $DSENT^{\perp}$ in Table 2. However the values of the third and fourth central moments are often quite different from those in Table 1, showing that the distribution features of sentiment indicators are only partly preserved after excluding the impact of macroeconomic fundamentals.

Table 1: Summary statistics of level indicators

	Mean	Median	S.D.	Skewness	Kurtosis
MICS	88.0382	90.9000	12.0688	-0.5643	-0.1183
CEFD	8.6773	8.4831	5.6940	0.7749	0.4045
TURN	0.6788	0.5951	0.3483	1.8365	4.7533
NIPO	32.1922	26.0000	24.5669	0.9337	0.4199
RIPO	19.2303	14.1000	19.9166	2.4452	6.9476
NEIF	0.1607	0.1378	0.1098	1.4886	2.1435
PDND	-13.3069	-12.7761	10.3113	-0.9992	3.5558
SENT	0.2556	0.1897	2.2409	0.4220	0.4404
$\text{SENT}^{\perp}$	0.2113	0.0495	0.7338	0.5857	0.2133
DSENT	-0.0000	0.0177	1.0108	0.1049	2.9091
$\text{DSENT}^{\perp}$	0.0031	0.0223	0.9960	0.0852	0.9503

This table shows summary statistics for the data of original sentiment indicators used in the analysis. The full monthly sample contains 360 observations from Januaray 1978 through December 2007.

Table 2: Summary statistics of orthogonalised indicators (all)

	Mean	Median	S.D.	Skewness	Kurtosis
MICS	0.0000	0.3451	8.7410	-0.1600	-0.6455
CEFD	0.0000	0.5671	4.3111	-0.4517	-0.0220
TURN	0.0000	-0.0400	0.1865	1.9815	5.1790
NIPO	0.0000	-3.3587	22.3097	0.7149	0.7427
RIPO	0.0000	-4.2321	19.1223	2.0755	5.7079
NEIF	0.0000	-0.0064	0.0709	0.5940	0.6255
PDND	0.0000	1.0142	8.6264	-1.5189	5.6264
SENT	0.0000	0.0072	0.5935	0.4698	0.4706
$\text{SENT}^{\perp}$	0.0000	-0.0102	0.5760	0.6536	1.4064
DSENT	0.0000	-0.0022	0.9824	0.1171	2.5968
$\text{DSENT}^{\perp}$	0.0000	0.0295	0.8592	-0.0472	1.3667

This table shows summary statistics for the data of orthogonalised sentiment indicators with twelve fundamental variables used in the analysis. The full monthly sample contains 360 observations from Januaray 1978 through December 2007.

Figure 1 provides the histogram distributions of all sentiment indicators. Series 1 represent original indicators; series 2 represent orthogonalised indicators (all); series 3 represent orthogonalised indicators (significant). The notations $SENT\hat{}$ and $DSENT\hat{}$ stand for $SENT^{\perp}$ and $DSENT^{\perp}$ respectively in the

Table 3: Summary statistics of orthogonalised indicators (significant)

	Mean	Median	S.D.	Skewness	Kurtosis
MICS	0.0000	0.5462	8.8387	-0.1786	-0.6746
CEFD	0.0000	0.5323	4.4305	-0.4017	-0.0548
TURN	0.0000	-0.0470	0.1892	2.0256	7.3032
NIPO	0.0000	-3.5195	22.8703	0.7531	0.8933
RIPO	0.0000	-5.2771	19.8887	2.4428	7.0515
NEIF	0.0000	-0.0094	0.0719	0.6624	0.8215
PDND	0.0000	0.7839	8.7876	-1.4605	5.0888
SENT	0.0000	0.0162	0.6086	0.4240	0.2231
$SENT^{\perp}$	0.0000	0.0076	0.5848	0.5460	1.1106
DSENT	0.0000	0.0177	1.0108	0.1049	2.9091
$DSENT^{\perp}$	0.0000	0.0396	0.8860	0.1021	1.4895

This table shows summary statistics for the data of orthogonalised sentiment indicators with those fundamental variables that are significant used in the analysis. The full monthly sample contains 360 observations from Januaray 1978 through December 2007.

figure. As discussed above, the signs of skewness and kurtosis remain unchanged after orthogonalisation in all cases except those of $CEFD$, and skewness of $DSENT^{\perp}$ in series 2. We can confirm that the distributions of orthogonalised indicators are arguably similar after two orthogonalisation methods for most indicators[10]. However the differences between histogram of series 1 and those of series 2 and 3 are quite obvious, suggesting that excluding the influence of fundamental variables clearly changes the distributions of sentiment indicators.

Table 4 shows the correlation coefficients between original sentiment indicators. As found in the literature, correlations between different indicators are usually in small magnitudes, all below 0.5 except for the correlation between $PDND$ and $SENT^{\perp}$, and those between $SENT$ and $SENT^{\perp}$ as well as $DSENT$ and $DSENT^{\perp}$ which suggest that similar indices capture the common changes in the six indirect indicators in very similar ways. This finding is in line with the argument in literature that different indicators are all reflecting investor sentiment in only partial and different ways. More importantly, almost all the direct and indirect indicators are correlated with the level sentiment indices ($SENT$ and $SENT^{\perp}$) in the expected ways. $CEFD$ and $PDND$ are negatively related to the sentiment indices while all others are positively related to the sentiment indices except $RIPO$ is negatively relatively to $SENT$ and $TURN$ is to $SENT^{\perp}$. As argued by Baker and Wurgler (2007), the economic intuition of the correlations between direct and indirect indicators with the difference sentiment indices ($DSENT$ and $DSENT^{\perp}$) are less clear, as positive/negative changes in investor sentiment do not necessarily imply bullish/bearish investors. We confirm this statement by showing that the correlations cannot be easily calibrated into a straightforward economic pattern.

[10]The distributions of orthogonalised data after two orthogonalisation methods are probably more distinct for $DSENT$ and $DSENT^{\perp}$ than for the other indicators.

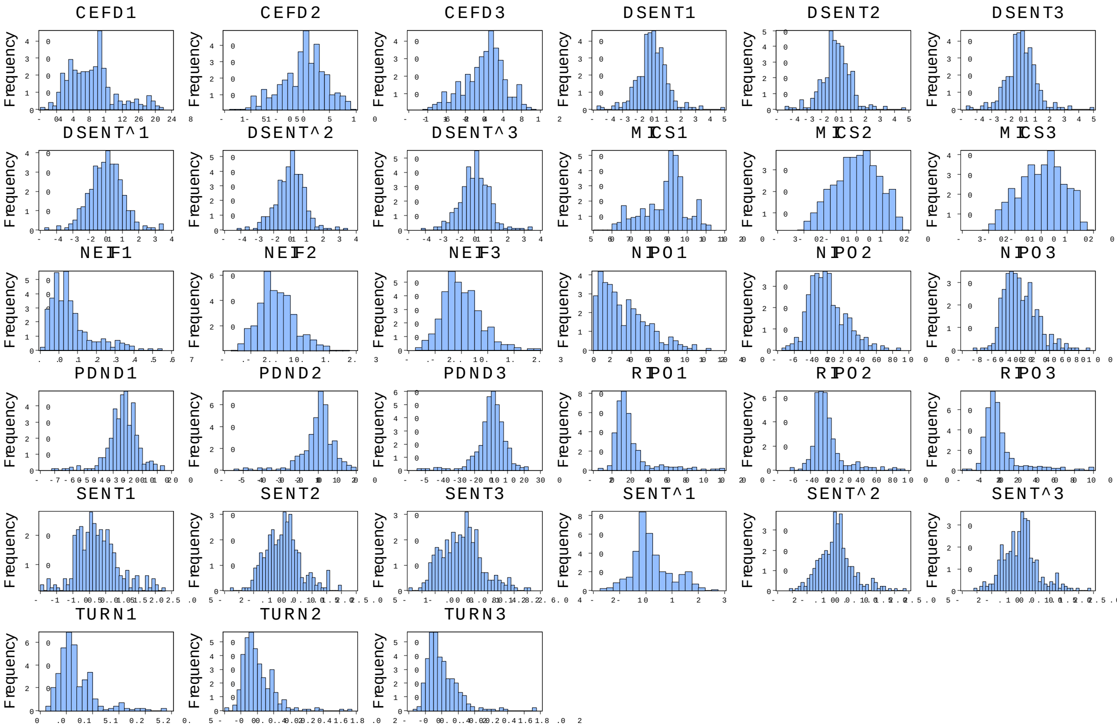


Figure 1: Histogram distributions of all sentiment indicators. Series 1 represent original indicators; series 2 represent orthogonalised indicators with twelve fundamental variables; series 3 represent orthogonalised indicators with only significant fundamental variables for each indicator. $SENT\hat{}$ and $DSENT\hat{}$ stand for $SENT^{\perp}$ and $DSENT^{\perp}$ respectively.

Table 5 and 6 show the correlation coefficients between the orthogonalised sentiment indicators, generated with the two orthogonalisation methods respectively. Similar results are obtained after two orthogonalisation methods, suggesting that the correlation structure of the indicators is robust to different fundamental variable choices in orthogonalisation. Again the correlations stay within small magnitudes, with the only exceptions of correlations of MICS with $SENT$ and $CEFD$ with $SENT$ increased reasonably to above 0.5. All the correlations of the direct and indirect indicators with level sentiment indices ($SENT$ and $SENT^{\perp}$) maintain the same signs and similar values as in Table 4 except that between $TURN$ and $SENT$. In general the correlation structure of the eleven indicators stay the same after orthogonalisation except a few cases involving $CEFD$ or $TURN$.

Table 4: Correlation of original indicators

	MICS	CEFD	TURN	NIPO	RIPO	NEIF	PDND	SENT	SENT$^{\perp}$	DSENT	DSENT$^{\perp}$
MICS	1										
CEFD	-0.3376	1									
TURN	0.2923	-0.3472	1								
NIPO	0.2941	-0.2044	-0.1656	1							
RIPO	0.1224	0.2978	-0.0365	-0.0024	1						
NEIF	-0.3530	0.3282	-0.4663	0.3353	0.1728	1					
PDND	-0.1410	-0.1385	-0.0094	-0.3765	-0.4560	-0.4326	1				
SENT	0.2342	-0.3887	0.0529	0.3057	-0.0573	0.2450	-0.4027	1			
SENT$^{\perp}$	0.0612	-0.1588	-0.0203	0.2221	0.0545	0.4136	-0.5368	0.9124	1		
DSENT	-0.1115	0.1332	0.0558	0.0396	0.4467	0.0495	-0.1492	-0.2538	-0.2040	1	
DSENT$^{\perp}$	-0.0880	0.0753	0.0574	0.0853	0.3034	0.0291	-0.1065	-0.1792	-0.1218	0.6867	1

This table shows the correlation coefficients of original indicators.

Table 5: Correlation of orthogonalised indicators (all)

	MICS	CEFD	TURN	NIPO	RIPO	NEIF	PDND	SENT	SENT$^{\perp}$	DSENT	DSENT$^{\perp}$
MICS	1										
CEFD	-0.1915	1									
TURN	-0.1533	0.1312	1								
NIPO	0.2627	-0.3488	-0.1652	1							
RIPO	0.1534	0.2732	-0.0472	-0.0411	1						
NEIF	-0.0248	0.0125	-0.1665	0.3928	0.1383	1					
PDND	-0.2406	0.0060	0.0985	-0.4120	-0.4402	-0.4179	1				
SENT	0.5342	-0.5347	-0.0866	0.4161	-0.0707	0.1609	-0.3264	1			
SENT$^{\perp}$	0.4127	-0.3563	-0.1246	0.3233	0.0284	0.2693	-0.4368	0.9228	1		
DSENT	-0.1584	0.1851	0.1087	0.0385	0.4562	0.0518	-0.2197	-0.2373	-0.1988	1	
DSENT$^{\perp}$	-0.0669	-0.0079	0.0912	0.1681	0.2830	0.0137	-0.1445	-0.1235	-0.1196	0.7335	1

This table shows the correlation coefficients of orthogonalised indicators with twelve fundamental variables.

Table 6: Correlation of orthogonalised indicators (significant)

	MICS	CEFD	TURN	NIPO	RIPO	NEIF	PDND	SENT	SENT$^{\perp}$	DSENT	DSENT$^{\perp}$
MICS	1										
CEFD	-0.1901	1									
TURN	-0.1427	0.1501	1								
NIPO	0.2638	-0.3661	-0.1728	1							
RIPO	0.1500	0.2901	-0.0169	-0.0539	1						
NEIF	-0.0241	0.0242	-0.1566	0.3684	0.1468	1					
PDND	-0.2362	0.0018	0.0823	-0.4003	-0.4427	-0.4142	1				
SENT	0.5297	-0.5520	-0.1034	0.4293	-0.0893	0.1542	-0.3106	1			
SENT$^{\perp}$	0.4138	-0.3549	-0.1279	0.3250	0.0332	0.2701	-0.4393	0.9081	1		
DSENT	-0.1503	0.2067	0.1331	0.0155	0.4469	0.0589	-0.2193	-0.2457	-0.1996	1	
DSENT$^{\perp}$	-0.0573	0.0122	0.1117	0.1339	0.2821	0.0202	-0.1338	-0.1280	-0.1303	0.7252	1

This table shows the correlation coefficients of orthogonalised indicators with only significant fundamental variables.

4 Methodology

4.1 Model specification

Although investor sentiment's putative effect on asset prices has been studied extensively, giving rise to a rich literature, theoretical models nevertheless offer no guidance as to the length of the time horizon over which sentiment becomes impounded into asset prices. Evidence has been found in both short-horizon and long-horizon analysis[11]. Therefore we choose to follow the convention in the literature of market return predictability and conduct our analysis over both short and long horizons. Existing studies of market return predictability mainly follow three model specifications. Fama and French (1988b) try to explain future multiple-period returns using current predictor value; Jegadeesh (1991) predicts single-period returns with sum of lagged predictors as the regressor; Campbell and Shiller (1988) adopt a VAR model instead and show that it also implies long-horizon analysis. Further discussions about the similarities among these specifications and the advantage of each model can be found in Hodrick (1991) and Campbell (2001).

In this article we first follow the model in Fama and French (1988b) and hence conduct single factor regressions over multiple horizons. In this approach we examine the null hypothesis that investor sentiment indicators have no predictive power in market returns. We also consider a parallel approach

[11] See e.g. Fisher and Statman (2000) and Brown and Cliff (2000) for short-term results; Neal and Wheatley (1998) and Brown and Cliff (2005) for long-term evidence.

by adding in the first-order lagged future returns as an additional predictor. This approach tests the null hypothesis that investor sentiment indicators have no incremental predictive power with lagged returns in market returns. This latter consideration takes into account the self-predictive power of returns found in Poterba and Summers (1988) and Fama and French (1988a), which for instance may come as a result of either a time decaying component in the price process or only aggregation biases. Similar incremental predictive power tests in this literature can be found in e.g. Campbell and Shiller (1988), Jegadeesh(1991), Hodrick (1991) among others. Note that this latter model can also help reduce the autocorrelations in the residuals as a result of overlapping future long-horizon returns. We follow Fama and French (1988b, 1989) in setting the horizon lengths to 1, 3, 12, 24, 36, and 48 months.

In Part 5, we regress future k-month average returns on each sentiment indicator. A constant term is included as well.

$$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \beta^{(k)} S_t + \epsilon_t^{(k)} \tag{1}$$

In Part 6, we regress future k-month average returns on first-order lagged returns and each sentiment indicator. A constant term is included as well.

$$\frac{1}{k}\sum_{i=1}^{k} r_{t+i} = c^{(k)} + \alpha^{(k)}\left(\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}\right) + \beta^{(k)} S_t + \epsilon_t^{(k)} \tag{2}$$

In both model specifications:

(i) r can refer to equal-weighted return (EWR) or value-weighted return (VWR);

(ii) S represents one of sentiment indicators and can refer to $MICS$, $CEFD$, $TURN$, $NIPO$, $RIPO$, $NEIF$, $PDND$, $SENT$, $SENT^{\perp}$, $DSENT$, or $DSENT^{\perp}$;

(iii) k represents the horizon length and can take the values 1, 3, 12. 24, 36, or 48.

(iv) the coefficient $\beta^{(k)}$ represents how sensitive the future return is to investor sentiment, giving the horizon length k. If $\beta^{(k)}$ is statistically significant then evidence of predictive power in the investor sentiment indicator is present.

The horizon lengths we use here contain both monthly and long-horizon frequencies. As well studied in the literature on long-horizon regression, the overlapping dependent variables will introduce strong autocorrelations into the residuals and therefore lead to biased and in most cases inconsistent estimates for least square coefficients (see e.g. Valkanov 2003). Furthermore, the distributions of the estimated coefficients are often not normal, together with the calculated standard errors being incorrect. As a result standard hypothesis tests on significance of coefficients will not provide reliable results. We use bootstrap methods to correct the bias for hypothesis tests and introduce different methods used in the next subsection.

4.2 Bootstrap

Different methods have been proposed to obtain adjusted results for hypothesis tests. For instance, Hansen and Hodrick (1980) and Newey and West (1987) propose to use adjusted standard errors in calculating the t statistic. Valkanov (2003) proposes $t/\sqrt{T}$ to be used instead of the standard t statistic[12], as the latter does not converge to any well defined distribution whilst the former does. Our preferred approach, however, is to use bootstrap simulations to generate empirical distribution for the t statistic under the null hypothesis, and test the hypothesis based on the empirical distribution. This approach has several advantages in implementation. Firstly, it is not based on as strict asymptotic assumptions as the alternatives and therefore will not perform significantly less well in finite (and particularly small) samples or when the degree of overlapping is relatively "large"[13]. Secondly, it can deal with not only autocorrelation but also possible heteroskedasticity (with the right bootstrap method) in the residuals, while method like Hansen and Horick standard errors does not correct for heteroskedasticity. Thirdly, bootstrapping is relatively flexible as different approaches have been developed for the simulation, each suitable under a particular circumstance. Last but not least, the bootstrap can even overcome the initial small sample problem by careful choice of the most suitable data generating process to increase the sample size. For further discussion on bootstrap method in general, see MacKinnon (2006).

4.2.1 Data generating process

We use the moving-block bootstrap approach in our bootstrap simulations to deal with both possible autocorrelation and possible heteroskedasticity in the residuals. In order to take account of possible autocorrelation and heteroskedasticity even at short horizon lengths[14], the bootstrap is implemented at all horizon lengths. Given a horizon length in regression Equation 1 or 2 and hence a sample size, for each (averaged) future return and sentiment indicator pair, overlapping moving blocks of 10 residuals are generated until the last nine residuals are left[15]. Then residuals are drawn one by one with replacement, and for each residual drawn the moving block following this residual is chosen into the generated residual series until the sample size is reached. For unlikely but possibly the same residuals in the residual series from the regression (Equation 1 or 2), it is recognised as the first of these same residuals in the series and the following moving block is chosen. In case that the residual drawn comes from the last nine residuals, another moving block is chosen randomly. When the number of residuals needed is smaller than 10, the moving block is truncated and chosen into the generated residual series.

For example, with the horizon length set to be 3 months our data sample size is 358, therefore 349 overlapping moving blocks of residuals are generated, with each block containing 10 sequential residuals

[12] t is the standard t-statistics, and T is the sample size.

[13] See e.g. Mishkin (1992) and Goetzmann and Jorion (1993) for evidence on limited performance of the adjusted standard error approach.

[14] which may come as a result of, e.g. small sample biases as discussed in Stambaugh (1999).

[15] Moving blocks of fixed length tend to work better. See Lahiri (1999) for example. We relax this setting and match block lengths to horizon lengths later in the robustness tests.

from the regression (Equation 1 or 2). Then 36 residuals are drawn, and each residual is found in the original residual series. The residual values showing up more than once in the original residual series will be viewed as the one first showing up. The 36 moving blocks following these 36 residuals are then included in the generated residual sample. In case that a residual drawn is recognised to be one of the last 9 residuals from the original residual series, a random moving block from the 349 blocks is chosen instead. The last of the 36 moving blocks is truncated, as only the first 8 residuals in this block are needed.

Then a pseudo series of the dependent variable (average future returns) is generated. The series in Part 5 is generated according to the following equation:

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \overline{\epsilon_t^{(k)}} \tag{3}$$

where $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$ is the generated dependent variable; $\widehat{c^{(k)}}$ is the estimate of $c^{(k)}$ from the regression Equation 1; $\overline{\epsilon_t^{(k)}}$ is the bootstrapped series of the residuals.

The series in Part 6 is generated recursively according to the following equation:

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \widehat{\alpha^{(k)}}(\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}}) + \overline{\epsilon_t^{(k)}} \tag{4}$$

where $\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}}$ is the generated dependent variable; $\widehat{c^{(k)}}$ and $\widehat{\alpha^{(k)}}$ are the estimates of $c^{(k)}$ and $\alpha^{(k)}$ from the regression Equation 2; $\overline{\epsilon_t^{(k)}}$ is the bootstrapped series of the residuals. Following suggestions in Mackinnon (2006), the pre-sample value of $\frac{1}{k}\sum_{i=1}^{k} r_{t+i}$ is used to start the recursive process.

4.2.2 Hypothesis test

We test the null hypothesis $\widehat{\beta^{(k)}} = 0$ from regression Equation 1 in Part 5 and the null hypothesis $\widehat{\beta^{(k)}} = 0$ from regression Equation 2 in Part 6. To obtain the empirical distribution of the t statistic under the null, we regress the generated pseudo dependent variable from last subsection on the estimates of constant and the regressor(s) in

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + \beta^{(k)} S_t + \varepsilon_t^{(k)} \tag{5}$$

and

$$\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t+i}} = \widehat{c^{(k)}} + +\widehat{\alpha^{(k)}}(\overline{\frac{1}{k}\sum_{i=1}^{k} r_{t-1+i}}) + \beta^{(k)} S_t + \varepsilon_t^{(k)} \tag{6}$$

respectively in Part 5 and 6.

As the dependent variable is generated by Equation 3 or 4, our null hypothesis $\beta^{(k)} = 0$ is true in this bootstrap regression. As a result the estimated $\widehat{\beta^{(k)}}$ from the regression Equation 5 or 6 should be statistically insignificant.

We repeat the bootstrap for 4999 times. The number of bootstrap sample size is chosen according to the fact that $\alpha(1+B)$ should be an integer to make the simulation closer to be exact, where B is the bootstrap sample size (MacKinnon, 2006).

For each time of the bootstrap, we record the t statistic of $\widehat{\beta^{(k)}}$ from the regression Equation 5 or 6. The empirical distribution of the t statistic under the null hypothesis is then obtained by combining the 4999 values together. We then calculate the p-value of $\widehat{\beta^{(k)}}$ from the regression Equation 1 or 2 according to the empirical distribution and make inference based on the p-value. As suggested in MacKinnon (2006, p. 21), for hypothesis tests based on signed statistics, we may or may not wish to assume symmetry when calculating p-values. In present study we do not assume symmetry and therefore calculate the p-value under the null as in a single-tail test. This choice is validated by the fact that the empirical distribution generated from data is often heavily skewed in our sample.

5 Single factor analysis

Tables 7 to 9 present the sentiment indicator coefficients from regression Equation 1. Table 7 is based on original investor sentiment indicators, whereas Table 8 is based on orthogonalised sentiment indicators with twelve fundamental variables and Table 9 on orthogonalised sentiment indicators with only significant fundamental variables for each indicator. In all three tables coefficient estimates $\widehat{\beta^{(k)}}$ are reported, with the adjusted p-values from bootstrap distributions in the parentheses. Each p-values below 5% is denoted by a star (*) following the value in the parenthesis.

Table 7: Coefficients of original sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000321	-0.000684	-0.000334	-0.000746	-0.000214	-0.000520	-0.000155	-0.000322	-0.000185	-0.000257	-0.000228	-0.000214
	(0.047*)	(0.002*)	(0.027*)	(0.000*)	(0.055)	(0.000*)	(0.059)	(0.000*)	(0.013*)	(0.000*)	(0.000*)	(0.000*)
CEFD	0.000102	0.000163	0.000121	0.000092	0.000199	0.000273	0.000128	0.000245	-0.000011	0.000170	0.000015	0.000263
	(0.402)	(0.381)	(0.372)	(0.421)	(0.238)	(0.218)	(0.287)	(0.121)	(0.468)	(0.146)	(0.473)	(0.020*)
TURN	-0.007972	-0.009969	-0.007157	-0.008858	-0.009255	-0.006883	-0.013667	-0.008347	-0.013691	-0.006644	-0.013285	-0.004340
	(0.100)	(0.108)	(0.106)	(0.119)	(0.041*)	(0.143)	(0.004*)	(0.058)	(0.002*)	(0.061)	(0.001*)	(0.126)
NIPO	-0.000078	-0.000300	-0.000105	-0.000297	-0.000042	-0.000239	0.000014	-0.000147	0.0000233	-0.000142	0.000018	-0.000125
	(0.207)	(0.006*)	(0.092)	(0.003*)	(0.241)	(0.000*)	(0.382)	(0.000*)	(0.293)	(0.000*)	(0.310)	(0.000*)
RIPO	-0.000014	0.000102	-0.000009	0.000066	-0.000114	-0.000046	-0.000171	-0.000065	-0.000149	-0.000038	-0.000121	-0.000007
	(0.451)	(0.246)	(0.467)	(0.296)	(0.058)	(0.290)	(0.000*)	(0.085)	(0.000*)	(0.153)	(0.000*)	(0.407)
NEIF	-0.016613	-0.017561	-0.010525	-0.010381	-0.002942	-0.010815	0.007607	0.000695	0.008825	-0.003956	0.012378	-0.002609
	(0.216)	(0.262)	(0.293)	(0.336)	(0.417)	(0.267)	(0.245)	(0.471)	(0.165)	(0.322)	(0.049*)	(0.352)
PDND	0.000267	0.000426	0.000293	0.000547	0.000330	0.000479	0.000245	0.000314	0.000190	0.000233	0.000100	0.000122
	(0.107)	(0.074)	(0.072)	(0.024*)	(0.018*)	(0.005*)	(0.015*)	(0.002*)	(0.024*)	(0.003*)	(0.124)	(0.039*)
SENT	-0.005059	-0.008363	-0.004461	-0.007123	-0.002917	-0.004279	-0.001987	-0.003268	-0.000797	-0.002240	-0.000908	-0.002401
	(0.054)	(0.020*)	(0.061)	(0.030*)	(0.111)	(0.066)	(0.122)	(0.024*)	(0.302)	(0.045*)	(0.240)	(0.009*)
$SENT^{\perp}$	-0.005001	-0.006773	-0.004580	-0.005569	-0.003668	-0.003779	-0.002715	-0.002916	-0.001240	-0.001557	-0.000839	-0.001319
	(0.046)	(0.037*)	(0.055)	(0.068)	(0.044*)	(0.077)	(0.052)	(0.036*)	(0.188)	(0.107)	(0.241)	(0.091)
DSENT	0.002364	0.006642	0.000014	0.001449	-0.000288	0.000252	-0.000100	0.000143	-0.000292	-0.000119	-0.000021	0.000077
	(0.156)	(0.005*)	(0.490)	(0.235)	(0.360)	(0.407)	(0.421)	(0.402)	(0.263)	(0.380)	(0.497)	(0.402)
$DSENT^{\perp}$	-0.002748	-0.003013	-0.000823	-0.000961	-0.000355	-0.000203	-0.000237	-0.000108	-0.000288	-0.000216	-0.000124	-0.000017
	(0.112)	(0.135)	(0.243)	(0.282)	(0.273)	(0.389)	(0.289)	(0.396)	(0.209)	(0.240)	(0.347)	(0.473)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where residual is resampled in moving block bootstrap with block length fixed to 10.

Table 8: Coefficients of orthogonalised sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000408	-0.000874	-0.000392	-0.000898	-0.000189	-0.000481	-0.000234	-0.000400	-0.000262	-0.000318	-0.000336	-0.000308
	(0.055)	(0.004*)	(0.041*)	(0.002*)	(0.143)	(0.010*)	(0.032*)	(0.000*)	(0.008*)	(0.000*)	(0.000*)	(0.000*)
CEFD	0.000521	0.001186	0.000155	0.000727	0.000145	0.000629	-0.000063	0.000276	-0.000197	0.000201	-0.000146	0.000326
	(0.157)	(0.035*)	(0.355)	(0.111)	(0.343)	(0.068)	(0.426)	(0.133)	(0.205)	(0.137)	(0.231)	(0.019*)
TURN	-0.005890	-0.011929	0.000325	-0.006439	-0.001220	-0.000791	-0.003227	-0.004255	-0.002326	-0.004344	-0.001457	-0.003652
	(0.305)	(0.215)	(0.486)	(0.321)	(0.420)	(0.469)	(0.326)	(0.304)	(0.355)	(0.246)	(0.385)	(0.239)
NIPO	-0.000080	-0.000299	-0.000117	-0.000273	-0.000041	-0.000184	-0.000012	-0.000133	0.000012	-0.000125	0.000013	-0.000111
	(0.214)	(0.010*)	(0.089)	(0.009*)	(0.256)	(0.008*)	(0.398)	(0.001*)	(0.380)	(0.000*)	(0.364)	(0.000*)
RIPO	0.000058	0.000236	0.000032	0.000179	-0.000103	0.000019	-0.000171	-0.000054	-0.000162	-0.000028	-0.000129	-0.000001
	(0.314)	(0.055)	(0.374)	(0.080)	(0.070)	(0.393)	(0.001*)	(0.136)	(0.000*)	(0.231)	(0.000*)	(0.470)
NEIF	-0.029730	-0.019921	-0.035571	-0.015521	-0.026337	-0.025895	-0.009744	-0.005539	-0.006321	-0.011541	0.003263	-0.002141
	(0.168)	(0.314)	(0.084)	(0.335)	(0.079)	(0.123)	(0.230)	(0.340)	(0.284)	(0.128)	(0.388)	(0.402)
PDND	0.000159	0.000092	0.000266	0.000282	0.000411	0.000403	0.000374	0.000379	0.000260	0.000267	0.000127	0.000133
	(0.273)	(0.393)	(0.126)	(0.177)	(0.008*)	(0.019*)	(0.001*)	(0.000*)	(0.006*)	(0.001*)	(0.075)	(0.032*)
SENT	-0.008252	-0.012235	-0.006925	-0.010521	-0.004777	-0.007087	-0.003006	-0.004949	-0.001640	-0.003649	-0.001665	-0.003294
	(0.012*)	(0.004*)	(0.018*)	(0.008*)	(0.033*)	(0.012*)	(0.077)	(0.004*)	(0.177)	(0.006*)	(0.139)	(0.002*)
$SENT^{\perp}$	-0.009121	-0.010505	-0.007851	-0.008409	-0.006863	-0.006878	-0.005210	-0.005348	-0.002842	-0.002893	-0.002172	-0.002014
	(0.006*)	(0.018*)	(0.011*)	(0.035*)	(0.005*)	(0.014*)	(0.005*)	(0.003*)	(0.055)	(0.026*)	(0.080)	(0.058)
DSENT	0.003156	0.006913	0.000449	0.001502	0.000010	0.000394	0.000025	0.000224	-0.000239	-0.000086	0.000005	0.000098
	(0.091)	(0.009*)	(0.382)	(0.228)	(0.498)	(0.348)	(0.472)	(0.342)	(0.307)	(0.410)	(0.475)	(0.379)
$DSENT^{\perp}$	-0.000785	-0.000852	-0.000442	-0.000820	0.000310	0.000085	0.000110	-0.000323	0.000015	-0.000525	0.000119	-0.000409
	(0.384)	(0.592)	(0.374)	(0.345)	(0.333)	(0.471)	(0.421)	(0.271)	(0.480)	(0.091)	(0.353)	(0.087)

This table shows the coefficients of orthogonalised sentiment indicators in regressions at six horizon lengths. Each orthogonalised indicator is calculated from the corresponding original indicator orthogonalised with 12 fundamental control variables. Each new orthogonalised indicator is then used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where the residual is resampled in moving block bootstrap with block length fixed to 10.

Table 9: Coefficients of orthogonalised sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000443	-0.000913	-0.000367	-0.000880	-0.000193	-0.000503	-0.000223	-0.000404	-0.000253	-0.000325	-0.000326	-0.000312
	(0.042*)	(0.003*)	(0.048*)	(0.002*)	(0.135)	(0.006*)	(0.037*)	(0.000*)	(0.010*)	(0.000*)	(0.000*)	(0.000*)
CEFD	0.000245	0.000857	0.000058	0.000606	0.000086	0.000588	-0.000091	0.000283	-0.000213	0.000217	-0.000157	0.000345
	(0.323)	(0.089)	(0.452)	(0.154)	(0.404)	(0.081)	(0.381)	(0.138)	(0.174)	(0.130)	(0.226)	(0.014)
TURN	-0.008242	-0.012753	-0.000568	-0.006521	-0.002850	-0.002128	-0.004290	-0.004710	-0.003334	-0.004390	-0.002113	-0.003204
	(0.240)	(0.210)	(0.466)	(0.317)	(0.368)	(0.422)	(0.288)	(0.273)	(0.290)	(0.224)	(0.352)	(0.263)
NIPO	-0.000032	-0.000247	-0.000100	-0.000262	-0.000031	-0.000185	0.0000001	-0.000131	0.000020	-0.000128	0.000021	-0.000116
	(0.369)	(0.024*)	(0.121)	(0.013*)	(0.310)	(0.005*)	(0.517)	(0.000*)	(0.311)	(0.000*)	(0.270)	(0.000*)
RIPO	-0.000021	0.000094	-0.000016	0.000060	-0.000120	-0.000049	-0.000158	-0.000067	-0.000154	-0.000039	-0.000126	-0.000007
	(0.416)	(0.254)	(0.419)	(0.677)	(0.051)	(0.268)	(0.002*)	(0.085)	(0.000*)	(0.153)	(0.000*)	(0.408)
NEIF	-0.026510	-0.017770	-0.036362	-0.017253	-0.025857	-0.024230	-0.010171	-0.004816	-0.006136	-0.009675	0.002689	-0.001168
	(0.202)	(0.332)	(0.083)	(0.313)	(0.068)	(0.140)	(0.215)	(0.352)	(0.302)	(0.168)	(0.410)	(0.440)
PDND	0.000060	0.000005	0.000268	0.000318	0.000417	0.000464	0.000359	0.000403	0.000254	0.000290	0.000128	0.000152
	(0.595)	(0.493)	(0.121)	(0.147)	(0.005*)	(0.011*)	(0.002*)	(0.000*)	(0.006*)	(0.000*)	(0.084)	(0.018*)
SENT	-0.006969	-0.010827	-0.006334	-0.010102	-0.004417	-0.007098	-0.002646	-0.004968	-0.001363	-0.003770	-0.001496	-0.003518
	(0.021*)	(0.010*)	(0.026*)	(0.011*)	(0.041*)	(0.009*)	(0.089)	(0.002*)	(0.217)	(0.003*)	(0.170)	(0.001*)
$SENT^{\perp}$	-0.008224	-0.010309	-0.007476	-0.008756	-0.006729	-0.007713	-0.004775	-0.005740	-0.002561	-0.003336	-0.002060	-0.002477
	(0.012*)	(0.020*)	(0.014*)	(0.028*)	(0.005*)	(0.010*)	(0.010*)	(0.001*)	(0.075)	(0.010*)	(0.100)	(0.020*)
DSENT	0.002364	0.006642	0.000014	0.001449	-0.000288	0.000252	-0.000100	0.000143	-0.000292	-0.000119	-0.0000021	0.000077
	(0.148)	(0.009*)	(0.490)	(0.238)	(0.359)	(0.396)	(0.427)	(0.402)	(0.259)	(0.372)	(0.483)	(0.416)
$DSENT^{\perp}$	-0.002467	-0.001900	-0.000972	-0.000925	-0.000156	0.000201	-0.000238	-0.000134	-0.000191	-0.000238	-0.000069	-0.000183
	(0.162)	(0.258)	(0.223)	(0.311)	(0.405)	(0.410)	(0.311)	(0.391)	(0.326)	(0.245)	(0.417)	(0.259)

This table shows the coefficients of orthogonalised sentiment indicators in regressions at six horizon lengths. Each orthogonalised indicator is calculated from the corresponding original indicator orthogonalised with a subset of 12 fundamental control variables. Only those control variables that are significant in explaining each original indicator are included in each (different) subset. Each orthogonalised indicator is then used as the only regressor to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where the residual is resampled in moving block bootstrap with moving block length fixed to 10.

5.1 Signs

Most coefficient estimates have the same signs as predicted and found in the literature. When investor sentiment is high, current asset price will be driven up and therefore we can expect a lower future return in the market. When investor sentiment is low the opposite is true. Our empirical results substantiate this effect of sentiment. The direct indicator $MICS$ survey variable is positively reflecting sentiment, consistent with negative signs for the coefficients in all 36 regressions from Table 7 to 9. With regard to the indirect indicators, market turnover ($TURN$), IPO volume ($NIPO$), new equity issuance fraction ($NEIF$) are viewed to be positively related to sentiment; closed-end fund discount ($CEFD$) and dividend premium ($PDND$) are considered to be negatively related to sentiment. Our estimates of their coefficients also prove the predictions in 52 out of 60 cases in Table 7. The fraction increases slightly to 53/60 in both Table 8 and 9. Level sentiment indices ($SENT$ and $SENT^{\perp}$) also have negative signs in all the return-horizon combinations as expected. Since the economic intuition of the relationship between differenced sentiment indices ($DSENT$ and $DSENT^{\perp}$) is not clear, it is not surprising that the signs of coefficients for $DSENT$ and $DSENT^{\perp}$ do not seem to follow any consistent pattern across horizon lengths and across different measures of indicators (original and two orthogonalised measures).

Perhaps the most interesting finding regarding the signs of coefficients comes from $RIPO$. The sign of coefficient stays positive for short horizons (mainly at 1 month and 3 months although also at 12 months in Table 8), at least in regressing EWR, and turns negative for longer horizons. While behavioural asset pricing theories and anecdotal evidence generally agree that firms and investment banks are "timing" the market by launching IPO when investor sentiment is high and therefore predict that high $RIPO$ will be followed by low future returns, our empirical finding suggests that the conclusion is true only for longer horizons. In other words, in the short term there must exist more complicated dynamics between $RIPO$ and investor sentiment or other sentiment indicators. Firstly, $RIPO$ is also part of market return and may be driven up by low or negative sentiment level in previous periods. Secondly, it is widely agreed that pricing initial public offerings is extremely difficult and that even professionals can make mistakes quite often. Therefore high (low) $RIPO$ may simply be a consequence of undervalued (overvalued) equities at IPO instead of a result of high (low) demands on IPO equities driven by high (low) investor sentiments. Last but not least, perhaps it is the case that the $RIPO$ affects returns in a lagged way. As there is a long lead time for preparing an initial public offering, high $RIPO$ will make initial public offerings attractive but can only lead to a wave of new IPOs in at least a few months' time. In this way IPO volume ($NIPO$) will lag $RIPO$ by a certain length of time. If $NIPO$ is a good proxy of investor sentiment, then $RIPO$ will also affect future returns, but only in a lagged way. We will see that our finding in next subsection is in line with this explanation.

5.2 Significance

We consider the coefficients to be statistically significant when the adjusted p-values from bootstrap simulations are below 5%.

One immediate result is that the eleven indicators perform differently. The average number (rounded to integer) of significant coefficients from three tables is 10 for $MICS$, 1 for $CEFD$, 1 for $TURN$, 6 for $NIPO$, 3 for $RIPO$, 0 for $NEIF$, 7 for $PDND$, 8 for $SENT$, 7 for $SENT^{\perp}$, 1 for $DSENT$ and 0 for $DSENT^{\perp}$. Since the literature on investor sentiment predicts that the (unobserved) investor sentiment is negatively correlated to future returns, the difference in predictive powers of eleven indicators suggests that the indicators investigated are clearly not equally informative in reflecting the investor sentiment and thus in predicting market returns.

Another general conclusion is that EWR is better or equally explained by sentiment indicators. $MICS$, $NIPO$, $PDND$, $SENT$ and $SENT^{\perp}$ all have stronger predictive power in predicting EWR to some degree. For instance, the most typical example lies in $NIPO$, which cannot predict VWR but is always significant in regressing EWR through Table 7 to 9. The only exceptions are $TURN$ and $RIPO$: the former is significant in explaining VWR but not EWR in Table 7, but the predictive power disappears after orthogonalisation; the latter cannot predict EWR but is significant in explaining VWR at longer horizons. We will discuss these two cases shortly. It has been well studied that new stocks and small stocks are generally more affected by investor sentiment. Theoretical work argues that these stocks are harder to value and more difficult to arbitrage and hence will more likely be mispriced[16]. Empirical work finds evidence with various sentiment indicators[17]. Our evidence is consistent with these existing results. As new stocks and small stocks have lower capitalisation levels, they contribute more to equal-weighted index returns than to value-weighted index returns. Therefore compared to VWR, EWR amplifies the effect of investor sentiment on new and small stocks.

As the only direct sentiment indicator, $MICS$ works well compared to the indirect indicators and even sentiment indices. The numbers of significant coefficients are 10, 10 and 11 out of 12 return-horizon combinations in Table 7 to 9 respectively. The good performance of $MICS$ becomes even more surprising given that the survey is not specifically designed for asset market investors but in fact for general consumers. Given the fact that comparatively few studies in the literature have focused on survey data, the implication from this finding might be that more attention should be paid to them in future studies. Also we expect that with better data availability in the future, surveys that focus specifically on investor sentiment should achieve even better explanatory (predictive) performance.

Level sentiment indicators $SENT$ and $SENT^{\perp}$ only reasonably predict future returns in their original forms. Only 5 and 3 coefficients out of 12 cases are significant for each index indicator respectively in Table 7, and the predictability is primarily found in only EWR. Much stronger predictive power has

[16] A good review on this literature can be found in Baker and Wurgler (2007).

[17] See e.g. Lee, Shleifer and Thaler (1991), Neal and Wheatley (1998), Kamstra et al. (2003) and Edmans et al. (2007).

been found, however, after orthogonalisation. $SENT$ becomes significant in 9 cases in both Table 8 and 9, whilst $SENT^{\perp}$ in 9 and 10 cases after two orthogonalisation methods respectively. Note that $SENT^{\perp}$ is already the first principal component of six orthogonalised indirect indicators. It might be unclear at first sight why further orthogonalisation improves predictability. Our explanation centres in the fact that Baker and Wurgler (2006, 2007) orthogonalise these indirect indicators with only the first eight fundamental variables following a consumption-based asset pricing idea. Our result suggests that by following a conditional asset pricing idea and adding four new fundamental variables, we can further exclude noise from the indirect indicators and hence improve the predictability of the index indicator.

Again since the economic intuition of the relationship between differenced sentiment indices ($DSENT$ and $DSENT^{\perp}$) with market returns is not clear, it is not surprising that predictability is hardly found in regressing either equal-weighted or value-weighted index returns on $DSENT$ or $DSENT^{\perp}$. Combining this finding with the predictive power found in e.g. level sentiment indicators $SENT$ and $SENT^{\perp}$, it is confirmed that market returns are mainly affected by absolute level of investor sentiment (bullish or bearish) and not by the direction of change in investor sentiment. As a result since changes do not necessarily represent sentiment level, it does not have any explanatory power in general.

Regarding the indirect indicators, our finding suggests that different measures are, if anything, clearly all noisy proxies of investor sentiment. It is obvious that they affect market return in quite different ways.

IPO volume ($NIPO$) consistently predicts future EWR but not VWR over all horizons. This implies that $NIPO$ might be capturing primarily the investor sentiment around small stocks in the market. As new IPO tends to be small stocks, the result is in line with long recognised view among both academics and practitioners that in practice the market has been "timed" and that firms and investment banks have been taking advantage of high sentiments when issuing new stock.

Closed-end fund discount ($CEFD$) and net equity issuance fraction ($NEIF$) have very low explanatory powers. Market turnover ($TURN$) seems to predict VWR at longer horizons in its original form, but the predictability disappears after orthogonalisation. This implies that the predictive power found in its original form comes largely from the common influence of fundamentals on both $TURN$ and VWR. In general, $CEFD$, $NEIF$ and $TURN$ seem to be at most extremely noisy indicators of investor sentiment and cannot predict market returns. While it is widely studied in the literature that $CEFD$ can predict cross-section returns and especially the size premium, our finding suggests that the predictive power cannot be extended to aggregate market return. $NEIF$ and $TURN$ are both indicators based on the hypothesis that liquidity is informative in reflecting investor sentiment. Unlike the evidence in Baker and Stein (2004), our results fail to support this hypothesis. The finding here does not necessarily invalidate the hypothesis though, but instead can be considered as supporting the argument that neither $NEIF$ or $TURN$ is a proper proxy in this application. In fact, $TURN$ may come as a result of heterogeneity in investor beliefs, which does not necessarily lead to bullish or bearish investor sentiment at an aggregate market level since it is likely that bullish belief and bearish belief may well cancel out. $NEIF$ may be determined by pure firm decisions – firms in the market may reach their debt ceilings at similar time and are therefore forced to raise fund through equity, leading to an increase in $NEIF$ and vice versa. Moreover $NEIF$ can reflect investor sentiment only to an extent that the sentiment must particularly affect equity but have no influence on the debt market.

Significant amount of predictability has been found in dividend premium ($PDND$), primarily at 1 year or longer horizons. Weaker but similar evidence is also present to show that first day return of IPO ($RIPO$) is also significant only over longer horizons. We view these finding as evidence that $PDND$ and $RIPO$ affects returns in a lagged way[18]. As most firms have rather persistent dividend policy, when investors are switching to dividend-paying firms they are not only searching for "safety" for immediate "tomorrow" but rather for "safety" in the future. Therefore the effect of the dividend premium on future returns follows a lagged pattern. Also we find that $PDND$ works only slightly better in predicting equal-weighted returns, consistent with the intuition that when investors switch between dividend payers and nonpayers they are mainly concerned about the dividend policies of the firms, and hence both large and small firms will be affected in similar ways. With regard to $RIPO$, several studies point out that it leads the volume of IPO ($NIPO$). As $NIPO$ predicts future returns, $RIPO$ might also affect future returns in a lagged way.

Of course our finding here implies that model misspecification is present in the analysis on $PDND$ and $RIPO$. One may find it counterintuitive to find insignificant relationships at short horizons and significant relationships over longer horizons in multiple horizon analysis. At first glance this pattern looks very analogue to over-rejections of the insignificant null hypothesis over long-horizons due to autocorrelated residuals. We argue that it is unlikely the case here, as there is not any special or abnormal structure in the time series of $PDND$ and $RIPO$ compared to the other indicators and therefore there should not be a particular reason why bootstrap would fail to correct the over-rejections just for these two predictors. We further show analytically in Appendix A.1 that when the true relationship is in a lagged way and the model is misspecified without being correctly lagged, exactly the same pattern as in our results should be expected whenever a highly persistent predictor is used in a single factor regression.

5.3 Robustness

The robustness checks reported here are carried out in order to verify that our findings are not an artifact of particular methodological implementation choices. We explore the robustness of the results appearing in Tables 7–9 using three different approaches: (i) by varying the bootstrap's moving block length, (ii) by employing a paired moving block resampling technique inspired by Freedman (1981, 1984), and (ii)

[18] Similar arguments about the influence of sentiment indicators on returns in a lagged way can be found in, e.g. Baker and Wurgler (2006). They argue that generally indicators that involve firm supply responses should lag behind indicators based on investor demand. Furthermore they show that indicators based on investor demand also lead changes in returns.

by combining both (i) and (ii). We first introduce each approach and then briefly discuss the associated results summaries and how they compare with those appearing in Tables 7–9. The robustness checks are documented in full in Appendices B.1–B.9, which are available upon request.

Firstly, given the way overlapping returns are constructed, it may be more appropriate to choose the block lengths in the bootstrap data generating process according to the horizon lengths. For instance, at 3 months horizon length the return in Equation 1 or 2 becomes $\frac{1}{3}\sum_{i=1}^{3} r_{t+i}$ and therefore is expected to follow the $MA(2)$ process. As a result it is likely that the residuals also follow the $MA(2)$ process. In this case choosing a block length of 3 in the moving block bootstrap will better capture the structure of the original data. By setting the block lengths equal to the horizon lengths (1, 3, 12, 24, 36 and 48 months respectively) we obtain the first set or robustness test results.

Secondly, since model misspecification in the single factor regression (Equation 1) is almost certainly present, the influence of any omitted predictor will likely be captured in the residuals. Unless all the possibly omitted predictors are independent with the sentiment indicator, there will be dependence between the regressor and the residuals in Equation 1. As discussed by Freeman (1981, 1984), when this is the case it is important to calibrate this dependence into the DGP of any bootstrap implementation in order to achieve the best asymptotic results. Freeman categorises linear models into "regression" models where regressors can be viewed as constants, and "correlation" models where regressors must be considered random. In the latter type of model, it is inappropriate to bootstrap only the residuals, since the obliteration of dependence between regressors and residuals in the pseudo data will jeopardise the ability of bootstrap method to mimic the original data. In fact, Freedman (1984) proves that the asymptotic property of assuming a joint distribution between the regressors and residuals (and instrumental variables in his study which are not relevant here) and bootstrapping them in pairs is at least as sound as the conventional asymptotic methods.

Although the most common practice in paired bootstrap is to pair the dependent variable with the regressors[19], this article is by no means the first study to pair regressors and residuals in bootstrap or to resample the pair from blocks. Li and Maddala (1997) implicitly follow this idea and combine it with a parametric DGP for the regressors. They also consider combining recursive and block bootstrap with paired bootstrap in their application. MacKinnon (2006) suggests a similar approach to that followed by Li and Maddala to be used in all cases of multivariate models. Our second robustness test is constructed by pairing regressor and residuals in the moving block resampling.

In the third approach, we make both changes mentioned above to the approach described in the methodology section.

In all three approaches the pseudo series of the dependent variable (future EWR or VWR) is still generated by Equation 3. The hypothesis test is still based on the bootstrap distribution obtained from

[19] The asymptotic property of pairing the dependent variable with the regressors has also been shown in Freedman (1981, 1984).

Equation 5.

As the slope coefficients are still recorded from regressing Equation 1, the values and hence the signs do not change after the robustness tests. The tests focus instead on the bootstrap method used to general the empirical p-values of coefficients under the null and investigate whether the observed patterns in predictive powers of sentiment indicators are robust to different methods. In what follows we briefly discuss some evidence of robust results. In general, all the observed patterns stay robust through all three different approaches.

For instance, using original indicator data through the three additional approaches respectively, compared to the results in Table 7 the number of significant coefficients changes from 10 to 9, 8, 9 for $MICS$; from 1 to 2, 3, 0 for $CEFD$; from 3 to 5, 4, 4 for $RIPO$; from 1 to 2, 1, 0 for $NEIF$; from 8 to 8, 7, 7 for $PDND$; from 5 to 6, 6, 5 for $SENT$; from 3 to 4, 3, 2 for $SENT^{\perp}$. Note that the changes in numbers do not suggest extremely sensitive bootstrap distributions through different methods, but instead are primarily accompanied with p-values relatively close to 5% threshold from all four methods. The number of significant coefficients does not change for $TURN$, $NIPO$, $DSENT$ or $DSENT^{\perp}$ through any approach in robustness test and stay as 4, 6, 1 and 0. Stronger predictability is still found in EWR than in VWR with $MICS$, $NIPO$, $PDND$, $SENT$ and $SENT^{\perp}$, evidence in line with theoretical prediction that small stocks are more affected by investor sentiment. Although the performance of $MICS$ becomes slightly weaker after all three approaches, the survey data indicator still shows stronger predictive power than other indicators. Level sentiment indices ($SENT$ and $SENT^{\perp}$) still only show reasonable explanatory power in their original forms, while lack of predictability remains with differenced sentiment indices ($DSENT$ and $DSENT^{\perp}$). $NIPO$ remains significant in explaining EWR but not VWR. $NEIF$ is still significant in only very few cases and the significance seems to be random. $CEFD$ works slightly better but the predictive power remains weak in general. Performance of $RIPO$ also only slightly improves through the three additional approaches. The conclusions regarding $TURN$ and $PDND$ stay as before too. The patterns in coefficients of $PDND$ and $RIPO$ are still consistent with the analytical prediction under model misspecification regarding lag in Appendix A.1.

Similar statements regarding strong robustness can be made when orthogonalised data by either orthogonalisation method are used. Like the change from Table 7 to Table 8 and 9, the predictive power in $TURN$ disappears after the influence of fundamental factors are excluded. The performances of $SENT$ and $SENT^{\perp}$ significantly improve after orthogonalisation as shown earlier.

To further look into the similarities and differences between different approaches, we report the number of significant coefficients across four approaches for all indicator-return-horizon combinations in Table 10. For each combination the number of p-values under 5% is reported, ranging from 0 for all insignificant coefficients to 4 for all significant coefficients. Perfectly robust results would mean that different methods must generate the same conclusion about whether insignificance can be rejected. We consider values 4 and

0 to suggest strongest robustness in the indicator-return-horizon combination and the value 2 to suggest least robust cases. As before the number after indicator names represent the choice of indicator series – 1 standing for original data; 2 standing for orthogonalised data by all twelve fundamental variables; 3 standing for orthogonalised data by only significant fundamental variables. Table 10 shows that in most cases the four approaches lead to the same conclusion for hypothesis tests (343 out of 396 indicator-return-horizon combinations), while value 2 only shows up 20 time out of 396 total combinations.

Table 10: Number of significant coefficients across four approaches in bootstrap

	EWR						VWR					
	1-m	3-m	12-m	24-m	36-m	48-m	1-m	3-m	12-m	24-m	36-m	48-m
$MICS1$	4	4	4	4	4	4	2	4	0	0	2	4
$MICS2$	4	4	4	4	4	4	1	4	0	1	2	4
$MICS3$	4	4	4	4	4	4	4	4	0	1	2	4
$CEFD1$	0	0	0	1	1	3	0	0	0	0	0	0
$CEFD2$	4	0	2	1	1	3	0	0	0	0	0	0
$CEFD3$	0	1	2	1	1	3	0	0	0	0	0	0
$TURN1$	1	2	0	0	0	0	0	0	2	4	4	4
$TURN2$	0	0	0	0	0	0	0	0	0	0	0	0
$TURN3$	0	0	0	0	0	0	0	0	0	0	0	0
$NIPO1$	4	4	4	4	4	4	0	0	0	0	0	0
$NIPO2$	4	4	4	4	4	4	0	0	0	0	0	0
$NIPO3$	4	4	4	4	4	4	0	0	0	0	0	0
$RIPO1$	0	0	0	1	0	0	0	0	3	4	4	4
$RIPO2$	0	0	0	0	0	0	0	0	0	4	4	4
$RIPO3$	0	0	0	0	0	0	0	0	3	4	4	4
$NEIF1$	0	0	0	0	0	0	0	0	0	0	2	3
$NEIF2$	0	0	0	0	0	0	0	2	2	0	0	0
$NEIF3$	0	0	0	0	0	0	0	2	2	0	0	0
$PDND1$	0	4	4	4	4	3	0	0	3	4	4	0
$PDND2$	0	0	4	4	4	3	0	0	4	4	4	0
$PDND3$	0	0	4	4	4	4	0	0	4	4	4	0
$SENT1$	4	4	0	4	3	4	1	3	0	0	0	0
$SENT2$	4	4	4	4	4	4	4	4	3	0	0	1
$SENT3$	4	4	4	4	4	4	4	4	2	0	0	1
$SENT^{\perp}1$	3	0	0	3	2	0	2	2	1	0	0	0
$SENT^{\perp}2$	4	4	4	4	3	0	4	4	4	4	2	2
$SENT^{\perp}3$	4	4	4	4	4	3	4	4	4	4	0	1
$DSENT1$	4	0	0	0	0	0	0	0	0	0	0	0
$DSENT2$	4	0	0	0	0	0	0	0	0	0	0	0
$DSENT3$	4	0	0	0	0	0	0	0	0	0	0	0
$DSENT^{\perp}1$	0	0	0	0	0	0	0	0	0	0	0	0
$DSENT^{\perp}2$	0	0	0	0	0	2	0	0	0	0	0	0
$DSENT^{\perp}3$	0	0	0	0	0	0	0	0	0	0	0	0

This table shows the number of significant coefficients of each sentiment indicator as the only regressor to explain EWR and VWR across different horizons - 1 3 12 24 36 and 48 months. For each indicator-return-horizon combination four approaches are used seperately in bootstrap to general the empirical p-values, including the standard approach in earlier analysis and three robustness test approaches. Perfectly robust results would mean that all the numbers must be either 4 or 0, while we consider the numbers of value 2 as least robust cases. As before the number 1 after indicator names stands for original data, and 2 and 3 stand for orthogonalised data by two methods.

We can further investigate the worst scenario indicator ($SENT^{\perp}$) which generates the least robust results across approaches in Table 10 (especially in $SENT^{\perp}1$, i.e. original form). Figure 2 plots the p-values for the coefficients of $SENT^{\perp}$ in all return-approach combinations as horizon increases. As before $SENT^{\perp}1$ stands for original data while $SENT^{\perp}2$ and $SENT^{\perp}3$ stand for orthogonalised data for $SENT^{\perp}$. Approaches are denoted as A B C $D - A$ stands for standard approach used in Table 7 while B C D represent robustness tests 1 to 3. For instance, the upper-left corner figure shows the p-values of four approaches when original $SENT^{\perp}$ are used to explain equal-weighted returns. We can see that the typical difference in p-values from different approaches is of reasonable magnitude[20]. Moreover, the values 1 3 and 2 for $SENT^{\perp}1$ in Table 10 mainly come from the fact that when p-values are near 5% threshold the conclusion is very sensitive to approaches, even though different approaches only lead to small fluctuations in p-values. Similar figures for other indicators are reported in Appendix A.2.

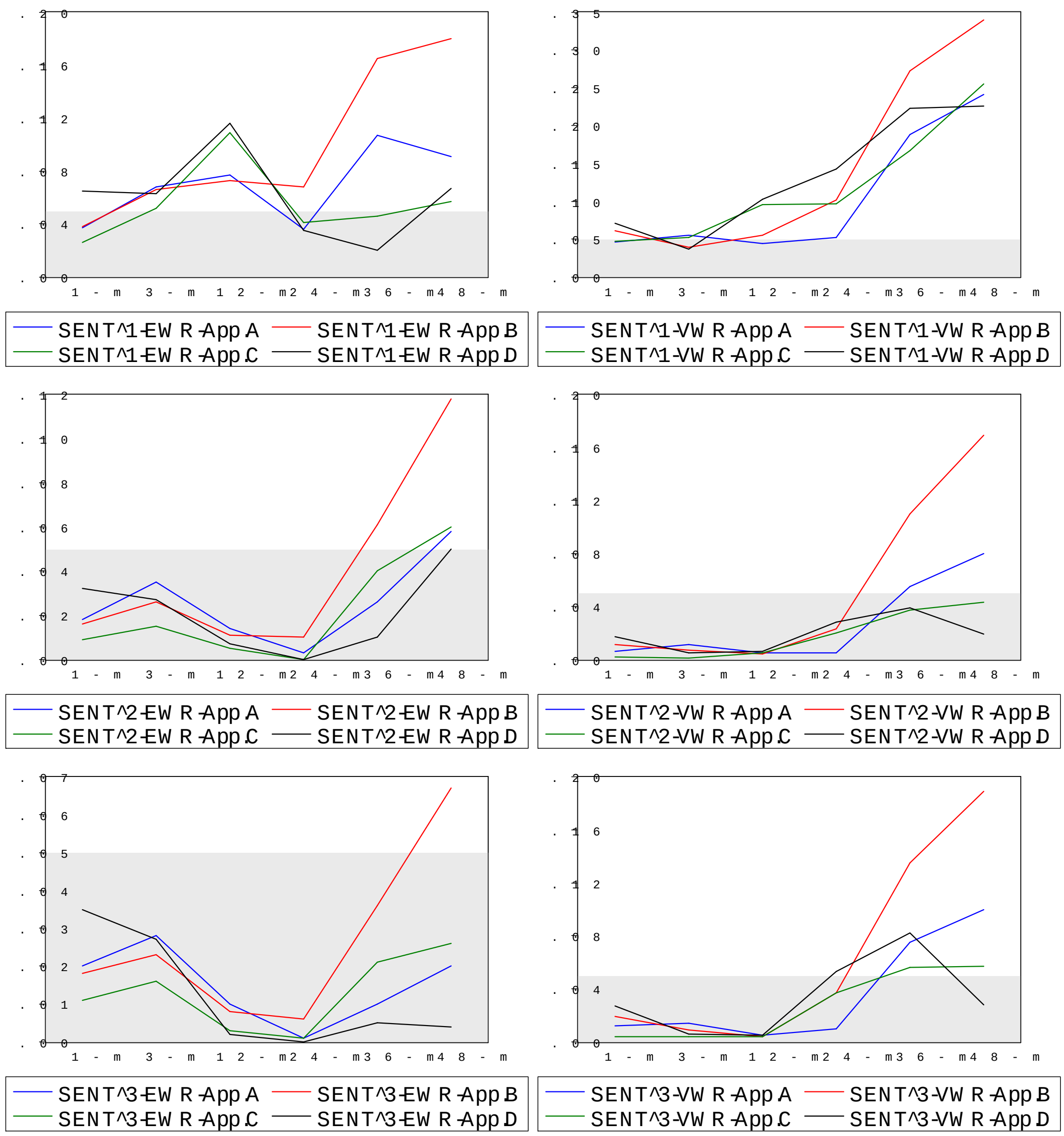


Figure 2: This figure plot the p-values for the coefficient of $SENT^{\perp}$ in all return-approach combinations as horizon increases. As before $SENT^{\perp}1$ stands for original while $SENT^{\perp}2$ and $SENT^{\perp}3$ stand for orthogonalised data for $SENT^{\perp}$. Approaches are denoted as A B C D.

[20] It is common than as horizon increases the differences also goes up, since with more overlapped returns and therefore both heavier autocorrelation and smaller sample size, bootstrap performs less well and different approaches generate more distinct empirical distributions. The result could be even more unstable if say, Hansen-Hodrick or Newey-West standard errors were used instead of bootstrap.

6 Double factor analysis

Tables 11 to 13 present the sentiment indicator coefficients from regression Equation 2. Table 11 is based on original investor sentiment indicators, whereas Table 12 is based on orthogonalised sentiment indicators with twelve fundamental variables and Table 13 on orthogonalised sentiment indicators with only significant fundamental variables for each indicator. In all three tables coefficient estimates $\widehat{\beta^{(k)}}$ are reported, with the adjusted p-values from bootstrap distributions in the parenthesis. Each p-value below 5% is denoted by a star (*) following the value in the parenthesis. We discuss the results in comparison to those in Section 5.

Table 11: Coefficients of original sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000319	-0.000628	-0.000167	-0.000348	-0.000030	-0.000066	-0.000008	-0.000013	-0.000015	-0.000017	-0.000015	-0.000008
	(0.037*)	(0.000*)	(0.014*)	(0.000*)	(0.127)	(0.040*)	(0.235)	(0.212)	(0.038*)	(0.099)	(0.011*)	(0.188)
CEFD	0.000095	0.000058	-0.000047	-0.000069	0.000003	0.000015	-0.000023	-0.000001	-0.000020	-0.000001	-0.000006	0.000018
	(0.406)	(0.438)	(0.377)	(0.340)	(0.469)	(0.416)	(0.171)	(0.483)	(0.120)	(0.470)	(0.322)	(0.135)
TURN	-0.008007	-0.008202	-0.002672	-0.003032	-0.001308	-0.001317	-0.000404	-0.000291	-0.000594	-0.000439	-0.000419	-0.000180
	(0.100)	(0.137)	(0.178)	(0.196)	(0.120)	(0.181)	(0.284)	(0.367)	(0.136)	(0.251)	(0.152)	(0.348)
NIPO	-0.000083	-0.000271	-0.000066	-0.000118	-0.000009	-0.000035	0.000003	-0.000010	-0.000001	-0.000016	-0.000002	-0.000013
	(0.183)	(0.005*)	(0.047*)	(0.009*)	(0.245)	(0.019*)	(0.296)	(0.108)	(0.428)	(0.006*)	(0.198)	(0.002*)
RIPO	-0.000025	-0.000041	-0.000078	-0.000135	-0.000063	-0.000068	-0.000027	-0.000021	-0.000025	-0.000021	-0.000013	-0.000010
	(0.390)	(0.377)	(0.061)	(0.020*)	(0.000*)	(0.000*)	(0.000*)	(0.004*)	(0.000*)	(0.000*)	(0.001*)	(0.007*)
NEIF	-0.016088	-0.017546	-0.004367	-0.004884	0.001494	-0.000745	0.001683	0.000730	0.001489	0.000169	0.000660	-0.000447
	(0.214)	(0.215)	(0.296)	(0.314)	(0.300)	(0.436)	(0.094)	(0.330)	(0.059)	(0.448)	(0.179)	(0.287)
PDND	0.000298	0.000556	0.000233	0.000421	0.000097	0.000130	0.000038	0.000054	0.000015	0.000028	0.000011	0.000020
	(0.086)	(0.018*)	(0.008*)	(0.001*)	(0.001*)	(0.001*)	(0.003*)	(0.001*)	(0.074)	(0.023*)	(0.057)	(0.009*)
SENT	-0.005129	-0.006325	-0.000726	-0.001010	0.000101	0.000277	0.000236	0.000198	0.000216	0.000176	0.000091	0.000033
	(0.043*)	(0.031*)	(0.286)	(0.272)	(0.400)	(0.333)	(0.114)	(0.222)	(0.065)	(0.175)	(0.192)	(0.415)
$\text{SENT}^{\perp}$	-0.005093	-0.005428	-0.001000	-0.000951	-0.000179	-0.000020	0.000113	0.000060	0.000181	0.000154	0.000071	0.000008
	(0.043*)	(0.043*)	(0.208)	(0.286)	(0.332)	(0.490)	(0.275)	(0.399)	(0.097)	(0.187)	(0.240)	(0.471)
DSENT	0.002060	0.000780	-0.004115	-0.007153	-0.002023	-0.002786	-0.000683	-0.000936	-0.000647	-0.000907	-0.000378	-0.000563
	(0.214)	(0.402)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)
$\text{DSENT}^{\perp}$	-0.003606	-0.008906	-0.002419	-0.004879	-0.001441	-0.002163	-0.000628	-0.000897	-0.000465	-0.000702	-0.000355	-0.000544
	(0.067)	(0.002*)	(0.005*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)

This table shows the coefficients of original sentiment indicators in regressions at six horizon lengths. Each indicator is used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where the residual is resampled in moving block bootstrap with block length fixed to 10.

6.1 Signs

The signs of coefficients in Table 11 to 13 are generally consistent with theoretical predictions and empirical findings in the literature. However slightly greater inconsistency has been found, compared to the results from single factor regressions. The direct indicator $MICS$ still has negative coefficients in all 36 regressions through three tables. The six indirect indicators ($CEFD$, $TURN$, $NIPO$, $RIPO$, $NEIF$ and $PDND$) also have expected signs in most cases. The fraction of expected signs is 58/72 in Table 11, 52/72 in Table 12 and 55/72 in Table 13. Compared to the fractions from single factor regressions, the decreases mainly come from two indicators – $CEFD$ and $NEIF$: $CEFD$ has only 5, 3 and 3 coefficients with expected positive signs through Table 11 to 13 respectively while $NEIF$ has only 6 coefficients with expected negative sign in each table. The positive coefficients of $RIPO$ that found at shorter horizons in Section 5 are less evident here, with only two cases at 1-months horizon in Table 12.

Using level sentiment indices ($SENT$ and $SENT^{\perp}$) as an additional predictor on top of lagged returns

Table 12: Coefficients of orthogonalised sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000411	-0.000859	-0.000199	-0.000419	-0.000045	-0.000068	-0.000024	-0.000023	-0.000018	-0.000013	-0.000027	-0.000022
	(0.052)	(0.001*)	(0.030*)	(0.001*)	(0.108)	(0.080)	(0.077)	(0.151)	(0.084)	(0.221)	(0.004*)	(0.042*)
CEFD	0.000480	0.000897	-0.000267	-0.000100	-0.000070	-0.000024	-0.000071	-0.000046	-0.000051	-0.000028	-0.000021	0.000008
	(0.179)	(0.054)	(0.108)	(0.362)	(0.154)	(0.401)	(0.009*)	(0.131)	(0.012*)	(0.176)	(0.106)	(0.325)
TURN	-0.005375	-0.010880	-0.000802	-0.004093	-0.000706	-0.001402	0.000012	-0.000403	-0.000204	-0.000742	-0.000115	-0.000364
	(0.334)	(0.208)	(0.430)	(0.270)	(0.351)	(0.291)	(0.498)	(0.382)	(0.385)	(0.224)	(0.422)	(0.292)
NIPO	-0.000087	-0.000267	-0.000064	-0.000086	-0.000010	-0.000026	0.000001	-0.000009	0.000001	-0.000010	-0.000003	-0.000012
	(0.201)	(0.008*)	(0.080)	(0.066)	(0.226)	(0.069)	(0.425)	(0.145)	(0.406)	(0.048*)	(0.177)	(0.005*)
RIPO	0.000047	0.000077	-0.000087	-0.000117	-0.000070	-0.000068	-0.000035	-0.000025	-0.000028	-0.000023	-0.000016	-0.000013
	(0.645)	(0.298)	(0.047*)	(0.040*)	(0.000*)	(0.000*)	(0.000*)	(0.003*)	(0.000*)	(0.000*)	(0.000*)	(0.003*)
NEIF	-0.028281	-0.015986	-0.018495	-0.006639	0.000670	-0.000964	0.001622	0.001256	0.001920	0.000908	0.000202	-0.000743
	(0.193)	(0.326)	(0.093)	(0.359)	(0.440)	(0.429)	(0.196)	(0.289)	(0.081)	(0.307)	(0.420)	(0.271)
PDND	0.000210	0.000409	0.000331	0.000489	0.000156	0.000180	0.000074	0.000095	0.000024	0.000044	0.000020	0.000036
	(0.207)	(0.092)	(0.002*)	(0.001*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.031*)	(0.004*)	(0.009*)	(0.000*)
SENT	-0.008237	-0.010455	-0.001537	-0.002624	-0.000125	-0.000237	0.000273	0.000071	0.000260	0.000120	0.000082	-0.000006
	(0.014*)	(0.004*)	(0.163)	(0.085)	(0.420)	(0.369)	(0.126)	(0.422)	(0.065)	(0.310)	(0.256)	(0.484)
$SENT^{\perp}$	-0.009114	-0.009374	-0.002323	-0.002429	-0.000574	-0.000483	0.000084	-0.000126	0.000215	0.000126	0.000049	-0.000023
	(0.008*)	(0.012*)	(0.074)	(0.115)	(0.153)	(0.253)	(0.373)	(0.353)	(0.117)	(0.303)	(0.357)	(0.445)
DSENT	0.002998	0.001388	-0.004109	-0.006994	-0.002019	-0.002651	-0.000711	-0.000896	-0.00668	-0.000889	-0.000397	-0.000557
	(0.117)	(0.335)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)
$DSENT^{\perp}$	-0.001673	-0.007313	-0.003673	-0.006482	-0.001648	-0.002249	-0.000830	-0.001082	-0.000583	-0.000848	-0.000450	-0.000653
	(0.285)	(0.013*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)

This table shows the coefficients of orthogonalised sentiment indicators in regressions at six horizon lengths. Each orthogonalised indicator is calculated from the corresponding original indicator orthogonalised with 12 fundamental control variables. Each new orthogonalised indicator is then used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where the residual is resampled in moving block bootstrap with block length fixed to 10.

Table 13: Coefficients of orthogonalised sentiment indicators and p-values

	1 month		3 months		12 months		24 months		36 months		48 months	
	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR	VWR	EWR
MICS	-0.000448	-0.000907	-0.000201	-0.000418	-0.000053	-0.000080	-0.000027	-0.000028	-0.000023	-0.000020	-0.000030	-0.000025
	(0.037*)	(0.000*)	(0.035*)	(0.000*)	(0.062)	(0.044*)	(0.055)	(0.100)	(0.030*)	(0.129)	(0.001*)	(0.025*)
CEFD	0.000209	0.000541	-0.000224	-0.000102	-0.000079	-0.000056	-0.000064	-0.000047	-0.000046	-0.000027	-0.000019	0.000006
	(0.346)	(0.159)	(0.141)	(0.345)	(0.121)	(0.266)	(0.014*)	(0.114)	(0.022*)	(0.175)	(0.129)	(0.390)
TURN	-0.007827	-0.012928	-0.001280	-0.004839	-0.001596	-0.002854	-0.000209	-0.000876	-0.000603	-0.001198	-0.000235	-0.000579
	(0.243)	(0.170)	(0.402)	(0.240)	(0.205)	(0.131)	(0.407)	(0.254)	(0.204)	(0.101)	(0.332)	(0.189)
NIPO	-0.000038	-0.000212	-0.000064	-0.000090	-0.000008	-0.000024	0.000001	-0.000009	0.000001	-0.000011	-0.000003	-0.000012
	(0.341)	(0.030*)	(0.071)	(0.050*)	(0.277)	(0.084)	(0.458)	(0.147)	(0.422)	(0.042*)	(0.189)	(0.003*)
RIPO	-0.000033	-0.000047	-0.000080	-0.000137	-0.000064	-0.000068	-0.000027	-0.000021	-0.000025	-0.000021	-0.000013	-0.000010
	(0.384)	(0.365)	(0.060)	(0.018*)	(0.000*)	(0.000*)	(0.000*)	(0.005*)	(0.000*)	(0.000*)	(0.001*)	(0.011*)
NEIF	-0.024996	-0.013964	-0.016034	-0.007908	0.001599	-0.000929	0.002138	0.001345	0.002159	0.000993	0.000404	-0.000735
	(0.212)	(0.347)	(0.123)	(0.328)	(0.634)	(0.432)	(0.133)	(0.296)	(0.057)	(0.287)	(0.340)	(0.276)
PDND	0.000103	0.000310	0.000322	0.000517	0.000153	0.000192	0.000067	0.000097	0.000025	0.000048	0.000018	0.000037
	(0.327)	(0.148)	(0.002*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.022*)	(0.001*)	(0.014*)	(0.000*)
SENT	-0.006916	-0.008948	-0.001677	-0.002812	-0.000037	-0.000124	0.000248	0.000069	0.000230	0.000087	0.000068	0.000007
	(0.021*)	(0.011*)	(0.130)	(0.071)	(0.471)	(0.441)	(0.145)	(0.419)	(0.091)	(0.345)	(0.293)	(0.523)
$SENT^{\perp}$	-0.008216	-0.009062	-0.002193	-0.002770	-0.000520	-0.000595	0.000138	-0.000133	0.000193	0.000075	0.000055	-0.000030
	(0.011*)	(0.013*)	(0.078)	(0.076)	(0.176)	(0.209)	(0.293)	(0.338)	(0.131)	(0.381)	(0.336)	(0.425)
DSENT	0.002060	0.000780	-0.004115	-0.007153	-0.002023	-0.002786	-0.000683	-0.000936	-0.000647	-0.000907	-0.000378	-0.000563
	(0.211)	(0.400)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)
$DSENT^{\perp}$	-0.003485	-0.008658	-0.003557	-0.006236	-0.001658	-0.002256	-0.000790	-0.001053	-0.000557	-0.000817	-0.000426	-0.000634
	(0.106)	(0.004*)	(0.001*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)	(0.000*)

This table shows the coefficients of orthogonalised sentiment indicators in regressions at six horizon lengths. Each orthogonalised indicator is calculated from the corresponding original indicator orthogonalised with a subset of 12 fundamental control variables. Only those control variables that are significant in explaining each original indicator are included in each (different) subset. Each orthogonalised indicator is then used with first-order lag of the dependent variable as the regressors to explain both value-weighted and equal-weighted future NYSE returns at 1 month, 3 months, and 1, 2, 3, 4 years. The coefficients of sentiment indicators from the regressions are reported. The p-value for the t-statistic of each coefficient is also reported in parentheses below the coefficient value. The p-values are obtained from the empirical distributions satisfying the null hypothesis in bootstrap simulations, where the residual is resampled in moving block bootstrap with block length fixed to 10.

generates more surprising signs. Through Table 11 to 13 quite a few positive signs have been found with the coefficients for both $SENT$ and $SENT^{\perp}$. In fact in Table 11 only less than half the coefficients (10 out of 24) for these indices stay positive, contrary to theoretical expectations. The fraction only increases mildly to 15/24 and 14/24 in Table 12 and 13 respectively. All the unexpected signs show up over 1-year and longer horizons.

Both differenced sentiment indices ($DSENT$ and $DSENT^{\perp}$) tend to have negative coefficients. $DSENT$ has positive signs at only 1-month horizon in all three tables, while coefficients of $DSENT^{\perp}$ have negative signs in all 36 regressions through three tables.

Although more inconsistency between the signs of coefficients and the theoretical predictions as well as existing empirical findings is present than in Section 5, we will see in what follows that all the coefficients of unexpected signs are statistically insignificant except two cases for $CEFD$. Therefore these unexpected signs do not in general invalidate any theoretical predictions or provide conflicting evidence against any existing empirical studies in the literature.

6.2 Significance

We consider the coefficients to be statistically significant when the adjusted p-values from bootstrap simulations are below 5%.

A similar conclusion to that in Section 5 can be drawn that the eleven indicators perform differently. The average number (rounded to integer) of significant coefficients from three tables is 7 for $MICS$, 1 for $CEFD$, 0 for $TURN$, 4 for $NIPO$, 9 for $RIPO$, 0 for $NEIF$, 10 for $PDND$, 2 for $SENT$, 2 for $SENT^{\perp}$, 10 for $DSENT$ and 11 for $DSENT^{\perp}$. Since the literature on investor sentiment predicts that the (unobserved) investor sentiment may have incremental predictive power in future returns, the differences in predictive powers of eleven indicators suggest that the indicators investigated are clearly not equally informative in reflecting the investor sentiment and thus in predicting market returns.

As the null hypothesis of no incremental predictive power is being tested here, there are reasonable differences between the significant coefficients found here and in single factor analysis. Specifically, the indicators that perform well in Section 5, including $MICS$, $NIPO$, $SENT$ and $SENT^{\perp}$, now have lower predictive powers. Contrarily the performances of $RIPO$, $PDND$, $DSENT$ and $DSENT^{\perp}$ have improved. $CEFD$ and $NEIF$ still fail to predict future returns. The weak explanatory power of original $TURN$ series in single factor analysis, if any at all, disappears when lagged return is used as an additional regressor.

EWR is still better or equally explained by sentiment indicators. However the evidence is not as strong as in single factor analysis. The pattern disappears with $MICS$, $SENT$ and $SENT^{\perp}$ and can only be found with $NIPO$ and original $PDND$ data. This suggests that once taking into account the self-explanatory power in market returns, only weaker evidence is present to support the view that new stocks and small stocks are more affected by investor sentiment.

The direct sentiment indicator $MICS$ still works reasonably well. The numbers of significant coefficients are 7, 5 and 8 out of 12 return-horizon combinations in Table 11 to 13 respectively. Given the fact that the survey is not even specifically designed for asset market investors but in fact for general consumers, we stay optimistic that with better data availability in the future, surveys that focus specifically on investor sentiment should achieve even better performance.

$NIPO$ only works reasonably well in double factor analysis. Predictive power has been found over different horizons when the original series is used (6 significant coefficients in Table 11), but this power decreases once fundamental influences are excluded (3 and 4 respectively in Table 12 and 13). The argument that $NIPO$ might be capturing primarily the investor sentiment around small stocks in the market still seems valid.

Closed-end fund discount ($CEFD$) and liquidity measures – net equity issuance fraction ($NEIF$) and market turnover ($TURN$) – still have very low explanatory powers. $CEFD$ has two significant coefficients when either orthogonalised series is used. However the significance seems random and the coefficients do not even have expected positive signs as in theories and previous empirical studies. No significant coefficient is found with either $NEIF$ or $TURN$. The widely found incremental predictive power of $CEFD$ in cross-section returns cannot be extended into aggregate market returns. As discussed in Section 5, $TURN$ may come as a result of heterogeneity in investor beliefs, which does not necessarily lead to bullish or bearish investor sentiment at an aggregate market level and thus may not be a proper proxy for investor sentiment. $NEIF$ may be determined by pure firm decisions in choosing equity or debt financing, and should not be used to represent investor sentiment unless the investor sentiment affects only equity market but not debt market.

Large amount of predictability has been found in dividend premium ($PDND$), primarily at 1 year or longer horizons. Similar evidence is also present to show that first day return of IPO ($RIPO$) is also significant only over longer horizons. Like in Section 5, we view these finding as evidence that $PDND$ and $RIPO$ affects returns in a lagged way. The discussion in Section 5 on the reasons why $PDND$ and $RIPO$ affects returns in such ways is also valid here. Also the results here are consistent with the analytical predictions in Appendix A.1.

The most radical changes occur in the results from level and differenced sentiment indicators. With lagged return as an additional regressor, the incremental predictability from $SENT$ and $SENT^{\perp}$ can be found only at monthly horizon length. Oppositely, strong predictability is found with $DSENT$ and $DSENT^{\perp}$ at 3-months and longer horizons. This predictability may come from two aspects. On the one hand, if market returns are mainly affected by absolute level of investor sentiment, then changes in market returns will be highly correlated to changes in absolute levels of investor sentiment. Combining this argument with the fact that overlapping returns at 3-months and longer horizons are highly persistent due to the moving average structure, it is not surprising that significant coefficients will be found for $DSENT$ and $DSENT^{\perp}$. On the other hand, by examining the time series features of $DSENT$ and $DSENT^{\perp}$ we find that neither of them is persistent and autoregressions have essentially zero explanatory power. This is exactly the condition needed to most validate the analytical predictions in Appendix A.1 that model misspecification will lead to insignificant coefficients at shorter horizons but significant coefficients at longer horizons. To this end, it may be inappropriate to ignore the possibility of incremental predictive power of $DSENT$ and $DSENT^{\perp}$ in a lagged way without any careful consideration. We believe that this may provide an additional reason why the results are behaving so.

6.3 Robustness

We follow the same approaches to implement robustness tests, by setting the block lengths equal to the horizon lengths (1, 3, 12, 24, 36 and 48 months respectively), by adopting a paired moving block resampling technique (pairing the sentiment indicator and residuals), or by making both changes at the same time. In this subsection we briefly discuss the findings different from those in Table 11 to 13. The pseudo series of the dependent variable (future EWR or VWR) is still generated by Equation 4, with pre-sample value used to start the recursive process. The hypothesis test is still based on the bootstrap distribution obtained from Equation 6. The complete results are reported and discussed in more details in separate Appendices B.10 to B.18, which are available upon request.

As the slope coefficients are still recorded from regressing Equation 2, the values and hence the signs do not change after the robustness tests. The tests focus instead on the bootstrap method used to general the empirical distribution of t-statistic under the null and generate only differences in the p-values. Like in Section 5.3, all the observed patterns stay robust through all three different approaches.

Using original indicator data through the three robustness test approaches and comparing the results to Table 11, we can confirm all the conclusions drawn in last subsection. Using the three approaches respectively, the number of significant coefficients changes from 7 to 6, 4, 7 for $MICS$; from 0 to 1, 0, 0 for $TURN$; from 6 to 6, 5, 5 for $NIPO$; from 9 to 8, 9, 9 for $RIPO$; from 0 to 0, 1, 0 for $NEIF$; from 9 to 7, 10, 9 for $PDND$; from 2 to 3, 1, 0 for $SENT$; from 2 to 3, 0, 0 for $SENT^{\perp}$. Note again that the changes in numbers do not suggest extremely sensitive bootstrap distributions through different methods, but instead are primarily accompanied with p-values relatively close to 5% threshold from all four methods. The number of significant coefficients does not change for $CEFD$, $DSENT$ or $DSENT^{\perp}$ through any approach in robustness test and stay as 0, 10 and 11. Stronger predictability is still found in EWR than in VWR with only $NIPO$ and $PDND$, showing just mild evidence in line with theoretical prediction that small stocks are more affected by investor sentiment. Although the performance of $MICS$ becomes slightly weaker in robustness tests, the survey data indicator still shows higher predictive power than most indicators. Level sentiment indices ($SENT$ and $SENT^{\perp}$) show only mild explanatory power in their original forms, while strong predictability remains with differenced sentiment indices ($DSENT$ and $DSENT^{\perp}$). $NIPO$ remains significant in explaining EWR but not VWR in general. $CEFD$, $NEIF$ and $TURN$ still show lack of predictive powers. The conclusions regarding $PDND$ and $RIPO$ stay as before too. The patterns in coefficients of $PDND$ and $RIPO$ and possibly $DSENT$ and $DSENT^{\perp}$ are still consistent with the analytical prediction under model misspecification regarding lag in Appendix A.1. Similar statements can be made when either orthogonalised data are used.

We report Table 14 in the same way as in Table 10. Table 14 shows that results from double factor analysis in Table 11 to 13 are even more robust to different approaches than those from single factor analysis, with 358 out of 396 indicator-return-horizon combinations leading to consistent conclusions in hypothesis tests across approaches now while the faction of least robust cases (in which value 2 shows up) further declines to 7 out of 396 total combinations. To provide more detailed comparisons among results from four approaches, we also report in Appendix A.3 the figures that plot the p-values for the coefficients of each indicator in all return-approach combinations as horizon increases. Like in Appendix A.2, typical difference in p-values from different approaches is of reasonable magnitude. Most inconsistent results come from p-values that are near 5% threshold and make the conclusions to hypothesis tests sensitive.

7 Testing more complex dynamics

Whilst most studies on the predictive power of investor sentiment focuses on explaining asset prices and returns with sentiment indicators, the possibility that there might be more complex dynamics between asset prices or returns and investor sentiment does not go unnoticed. For example, Brown and Cliff (2004) use a VAR model to find that market returns clearly Granger cause future changes in sentiment while very limited evidence supports that sentiment causes subsequent returns. Wang et al. (2006) use two trading ratios and two survey results as sentiment indicators and show that they are mostly caused by returns and volatility rather than vice versa. Given the predictive power of sentiment indicators widely found in the literature and shown in this article, it becomes natural to consider the possibility of more complex dynamics involving market returns and investor sentiment indicators. We show illustrative evidence on the presence of such dynamics through Granger causality tests.

Similar to Brown and Cliff (2004) and Wang et al. (2006), we conduct Granger causality tests through bivariate VAR models involving monthly return and each sentiment indicator. We rely on information criteria such as Akaike Information Criterion (AIC) and Schwarz/Bayesian Criterion (BIC) in selection of lag orders included in the VAR model. Among all combinations of both equal-weighted and value-weighted index returns with all three measures (original and two orthogonalised series) of eleven sentiment indicators, the lag order 1-1 yields the best values for both AIC and BIC except for original $MICS$ with EWR and original $SENT$ with EWR. Both exceptions suggest that a lag order 1-2 (1 for EWR and

Table 14: Number of significant coefficients across four approaches in bootstrap

	EWR						VWR					
	1-m	3-m	12-m	24-m	36-m	48-m	1-m	3-m	12-m	24-m	36-m	48-m
$MICS1$	4	4	2	0	1	0	3	3	0	0	3	4
$MICS2$	4	4	0	0	0	3	1	1	0	0	0	4
$MICS3$	4	4	4	0	0	4	4	2	0	0	2	4
$CEFD1$	0	0	0	0	0	1	0	0	0	0	0	0
$CEFD2$	1	0	0	0	0	0	0	0	0	4	4	0
$CEFD3$	0	0	0	0	0	0	0	0	0	3	4	0
$TURN1$	1	0	0	0	0	0	0	0	0	0	0	0
$TURN2$	0	0	0	0	0	0	0	0	0	0	0	0
$TURN3$	0	0	0	0	0	0	0	0	0	0	0	0
$NIPO1$	4	4	4	0	4	4	0	1	0	0	0	0
$NIPO2$	4	4	4	4	4	4	0	0	0	0	0	0
$NIPO3$	4	0	0	0	4	4	0	0	0	0	0	0
$RIPO1$	0	2	4	4	4	4	0	1	4	4	4	4
$RIPO2$	0	1	4	4	4	4	0	3	4	4	4	4
$RIPO3$	0	3	4	4	4	4	0	1	4	4	4	4
$NEIF1$	0	0	0	0	0	0	0	0	0	0	1	0
$NEIF2$	0	0	0	0	0	0	0	0	0	0	1	0
$NEIF3$	0	0	0	0	0	0	0	0	0	1	1	0
$PDND1$	4	4	4	4	3	3	0	3	4	4	1	1
$PDND2$	0	4	4	4	4	4	0	4	4	4	4	4
$PDND3$	0	4	4	4	4	4	0	4	4	4	4	4
$SENT1$	1	0	0	0	0	0	1	0	0	0	2	0
$SENT2$	3	0	0	0	0	0	4	0	0	0	0	0
$SENT3$	3	0	0	0	0	0	4	0	0	0	0	0
$SENT^{\perp}1$	2	0	0	0	0	0	2	0	0	0	0	0
$SENT^{\perp}2$	4	0	0	0	0	0	4	0	0	0	0	0
$SENT^{\perp}3$	3	0	0	0	0	0	4	0	0	0	0	0
$DSENT1$	0	4	4	4	4	4	0	4	4	4	4	4
$DSENT2$	0	4	4	4	4	4	0	4	4	4	4	4
$DSENT3$	0	4	4	4	4	4	0	4	4	4	4	4
$DSENT^{\perp}1$	4	4	4	4	4	4	0	4	4	4	4	4
$DSENT^{\perp}2$	4	4	4	4	4	4	0	4	4	4	4	4
$DSENT^{\perp}3$	4	4	4	4	4	4	0	4	4	4	4	4

This table shows the number of significant coefficients of each sentiment indicator as an additional regressor with lagged returns to explain EWR and VWR across different horizons - 1 3 12 24 36 and 48 months. For each indicator-return-horizon combination four approaches are used in bootstrap to general empirical p-values, including standard approach in earlier analysis and three robustness test approaches. Perfectly robust results would mean that all the numbers must be either 4 or 0, while we consider the numbers of value 2 as least robust cases. As before the number 1 after indicator names stands for original data, and 2 and 3 stand for orthogonalised data by two methods.

2 for $MICS$ or $SENT$) will provide slight improvement. For comparison reasons, we ignore these two exceptions and set the lag order to 1-1 in all return-indicator combinations.

It is worth pointing out the similarity between this approach and the double factor analysis in Section 6, at least at 1-month horizon. These two methods share the same idea of testing for incremental predictability and the evidence at 1-moth horizon in Section 6 also implies causality. The difference is that bootstrap is used for inference in Section 6 while here we draw conclusions based on standard t-statistics. On the one hand in this section we explicitly discuss about causality from investor sentiment indicators to market returns; on the other hand by following VAR we also test for causality in the opposite direction – from market returns to sentiment indicators. The p-values of rejecting the null hypothesis that no Granger causality is present are reported in Table 15. Arrows represent directions of Granger causality. Significant p-values are denoted by stars (*, ** and *** representing 10%, 5% and 1% significance levels respectively).

Our results show that the dynamics between sentiment indicators and market returns do not follow a uniform pattern. We find Granger causality at neither, either, or both directions for different indicators.

Table 15: p-values in Granger causality tests

	Original	Orthogonalised(all)	Orthogonalised(significant)		Original	Orthogonalised(all)	Orthogonalised(significant)
				Direct indicator			
$MICS \rightarrow EWR$	0.006***	0.006***	0.003***				
$EWR \rightarrow MICS$	0.000***	0.237	0.441				
$MICS \rightarrow VWR$	0.091*	0.110	0.080*				
$VWR \rightarrow MICS$	0.000***	0.019**	0.090*				
				Indirect indicators			
$CEFD \rightarrow EWR$	0.903	0.159	0.381	$TURN \rightarrow EWR$	0.296	0.459	0.367
$EWR \rightarrow CEFD$	0.695	0.560	0.632	$EWR \rightarrow TURN$	0.907	0.923	0.817
$CEFD \rightarrow VWR$	0.810	0.362	0.670	$TURN \rightarrow VWR$	0.246	0.720	0.568
$VWR \rightarrow CEFD$	0.464	0.102	0.264	$VWR \rightarrow TURN$	0.975	0.695	0.960
$NIPO \rightarrow EWR$	0.016**	0.031**	0.079*	$RIPO \rightarrow EWR$	0.772	0.609	0.740
$EWR \rightarrow NIPO$	0.000***	0.041**	0.032**	$EWR \rightarrow RIPO$	0.123	0.420	0.126
$NIPO \rightarrow VWR$	0.381	0.413	0.717	$RIPO \rightarrow VWR$	0.830	0.710	0.782
$VWR \rightarrow NIPO$	0.020**	0.113	0.135	$VWR \rightarrow RIPO$	0.082*	0.577	0.085*
$NEIF \rightarrow EWR$	0.485	0.686	0.720	$PDND \rightarrow EWR$	0.039**	0.213	0.336
$EWR \rightarrow NEIF$	0.000***	0.002***	0.001***	$EWR \rightarrow PDND$	0.095*	0.555	0.340
$NEIF \rightarrow VWR$	0.448	0.400	0.452	$PDND \rightarrow VWR$	0.189	0.440	0.700
$VWR \rightarrow NEIF$	0.000***	0.098*	0.019**	$VWR \rightarrow PDND$	0.061*	0.917	0.234
				Index indicators			
$SENT \rightarrow EWR$	0.112	0.025**	0.049**	$SENT^{\perp} \rightarrow EWR$	0.153	0.050**	0.055*
$EWR \rightarrow SENT$	0.577	0.483	0.348	$EWR \rightarrow SENT^{\perp}$	0.772	0.421	0.375
$SENT \rightarrow VWR$	0.125	0.036**	0.069*	$SENT^{\perp} \rightarrow VWR$	0.111	0.024**	0.039**
$VWR \rightarrow SENT$	0.763	0.499	0.605	$VWR \rightarrow SENT^{\perp}$	0.915	0.928	0.706
$DSENT \rightarrow EWR$	0.805	0.667	0.805	$DSENT^{\perp} \rightarrow EWR$	0.003***	0.036**	0.011**
$EWR \rightarrow DSENT$	0.083*	0.405	0.083*	$EWR \rightarrow DSENT^{\perp}$	0.005***	0.173	0.015**
$DSENT \rightarrow VWR$	0.414	0.249	0.414	$DSENT^{\perp} \rightarrow VWR$	0.144	0.563	0.210
$VWR \rightarrow DSENT$	0.164	0.498	0.164	$VWR \rightarrow DSENT^{\perp}$	0.016**	0.174	0.044**

This table shows the p-values of rejecting the null hypothesis of Granger non-causality between sentiment indicators and market return. Original and orthogonalised series are used for each sentiment indicator. Both equal-weighted and value-weighted NYSE index returns are used too. The lag orders in the tests are all set to 1-1, based on information creteria such as AIC and BIC for model selection. Significant p-values are denoted by *, ** and *** representing 0.1, 0.05 and 0.01 significance levels respectively. Arrows represent directions of Granger causality.

For instance, strong evidence has been found to reject that $MICS$ does not Granger cause EWR at 1% significance level, with all three measures of $MICS$. It can also be rejected at 1% significance level that EWR does not Granger cause $MICS$. We can also strongly reject Granger noncausality from VWR to $MICS$ and mildly reject noncausality from $MICS$ to VWR.

There is also strong evidence to reject Granger noncausality on both directions between $NIPO$ and EWR. Relatively much weaker evidence can be found regarding the dynamics between $NIPO$ and VWR, with only noncausality from VWR to original $NIPO$ data rejected.

Granger noncausality has been rejected from both EWR and VWR to $NEIF$. However the rejection cannot be made in the reverse direction, as it cannot be rejected that $NEIF$ does not Granger cause either index return.

We also find evidence to reject noncausality from original $PDND$ to EWR. Noncausality in the opposite direction can only be weakly rejected from both EWR and VWR to $PDND$.

There is generally little evidence supporting causalities between $RIPO$ and market returns. The only weak evidence found is the rejection of noncausality from VWR to $RIPO$ at 10% significance level.

We fail to reject the null that there is no Granger causality between $CEFD$ or $TURN$ with market returns. Noncausality cannot be rejected with either measure of these indicators and with either index return.

We cannot reject the hypothesis that original $SENT$ and $SENT^{\perp}$ do not Granger cause market returns. Nevertheless after orthogonalisation stronger evidence of rejection has been found with both level index indicators. Neither EWR nor VWR seems to Granger cause any of the three measures of $SENT$ and $SENT^{\perp}$.

Only weak evidence has been found supporting EWR Granger causing $DSENT$. Strong evidence is present to suggest that $DSENT^{\perp}$ Granger causes EWR. However we cannot extend the causality to VWR. Evidence has also been found to support that market returns also Granger cause $DSENT^{\perp}$.

The findings here confirm our earlier statement that different sentiment indicators are not all equally informative. Intuitively most of these results are consistent with the findings from Section 5 and 6, e.g. (i) $MICS$ tends to predict returns and in particular EWR; (ii) $NIPO$ in general consistently performs well in predicting EWR; (iii) $CEFD$, $TURN$ and $NEIF$ fail to show significant explanatory power etc. In general, we believe that $MICS$, $NIPO$ and index indicators are generally better proxies for investor sentiment while $CEFD$, $TURN$, $NEIF$ and possibly $RIPO$ are relatively noisy. As discussed earlier, the economic meaning of what $PDND$ captures and possible lags before its influence shows up in returns require careful thinking in selecting the most appropriate model specification in any application. To conclude, the complex and non-uniform dynamics between market returns and different sentiment indicators suggest that careful consideration be taken when future studies face such decision on selection among different available indicators.

8 Conclusion

In this article we examine market return predictability by using investor sentiment indicators as (i) the only predictor and (ii) an additional predictor on top of lagged returns. We do so by conducting a comprehensive investigation on eleven major investor sentiment indicators in the existing asset pricing literature, in a unified framework within the same sample period. Equal-weighted and value-weighted index returns of NYSE are both analysed. The investor sentiment indicators studied include direct sentiment measures, indirect sentiment measures and first principal component index measures. We conduct long-horizon regressions at time lengths of 1, 3, 12, 24, 36 and 48 months. Parallel studies are implemented using both original indicator data and orthogonalised data according to two methods, with respect to twelve macroeconomic fundamental variables and only a significant subset of the twelve respectively. Moving block bootstrap has been used to general empirical p-values for hypothesis tests. We find signs of coefficients mostly consistent with the predictions of theories and existing empirical evidence. Some indicators predict market returns significantly while others do not show much predictive power. Results are also consistent with the argument in the literature that some indicators affect returns in a lagged way. All these findings show that different indicators are not equally informative in reflecting investor sentiment.

We further search for more complex dynamics between market returns and investor sentiment indicators by implementing Granger causality tests through bivariate VAR models. Information criteria including AIC and BIC are chosen in selection of lag orders. The results show that there are complex and non-uniform dynamics between market return and different sentiment indicators.

References

[1] Amihud, Y., and Mendelson, H. (1986). Asset Pricing and the bid-ask spread, *Journal of Financial Economics*, 17, 223-249

[2] Baker, M. and Stein, J.C. (2004). Market Liquidity as a Sentiment Indicator, *Journal of Financial Markets*, 7, 271-299

[3] Baker, M. and Wurgler, J. (2000). The Equity Share in New Issues and Aggregate Stock Returns, *Journal of Finance*, 55, 2219-2257

[4] Baker, M. and Wurgler, J. (2004a). A Catering Theory of Dividends, Journal of Finance, 59, 1125-1165

[5] Baker, M. and Wurgler, J. (2004b). Appearing and Disappearing Dividends: The Link to Catering Incentives, *Journal of Financial Economics*, 73, 271-288

[6] Baker, M. and Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns, *Journal of Finance*, 61, 1645-1690

[7] Baker, M. and Wurgler, J. (2007). Investor Sentiment in the Stock Market, *Journal of Economic Perspectives*, 21, 129-151

[8] Barberis, N., Shleifer, A., and Vishny, R. (1998). A Model of Investor Sentiment. *Journal of Financial Economics*, 49(3), 307-343

[9] Bernard, V.L. and Thomas, J.K. (1990). Evidence That Stock Prices Do Not Fully Reflect The Implications of Current Earnings for Future Earnings, *Journal of Accounting and Economics*, 13(4), 305-340

[10] Black, F. (1986). Noise, *Journal of Finance*, 41(3), 529-543

[11] Brennan, M., Chordia, T. and Subrahmanyam, A. (1998). Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics*, 49, 345-373

[12] Brennan, M., and Subrahmanyam, A. (1996). Market microstructure and asset pricing: on the compensation for illiquidity in stock returns, *Journal of Financial Economics*, 41, 441-464

[13] Brickley, J.A. and Schallheim, J.S. (1985). Lifting the Lid on Closed-End Investment Companies: A Case of Abnormal Returns, *Journal of Financial and Quantitative Analysis*, 20, 107-117

[14] Brown, G.W., and Cliff, M.T. (2004). Investor Sentiment and the Near-Term Stock Market, *Journal of Empirical Finance*, 11, 1-27

[15] Brown, G.W., and Cliff, M.T. (2005). Investor Sentiment and Asset Valuation, *Journal of Business*, 78, 405-440

[16] Campbell, J.Y. (2001). Why Long Horizons? A Study of Power against Persistent Alternatives, Journal of Empirical Finance, 8, 459-491

[17] Campbell, J.Y. and Kyle, A.S. (1993). Smart Money, Noise Trading and Stock Price Behaviour, *Review of Economic Studies*, 60(1), 1-34

[18] Campbell, J.Y. and Shiller, R.J. (1988). The dividend-price Ratio and Expectations of Future Dividends and Discount Factors, *Review of Financial Studies*, 1, 195-228

[19] DeBondt, W.F.M., and Thaler, R. (1985). Does the Stock Market Overreact?, Journal of Finance, 30(3), 793-805

[20] DeLong, J.B., Shleifer, A., Summers, H., and Waldmann, R. (1990). Noise Trading Risk in Financial Markets, *Journal of Political Economy*, 98, 703-738

[21] Edmans, A., Garcia, D., and Norli, O. (2007). Sports Sentiment and Stock Returns, *Journal of Finance*, 62(4), 1967-1998

[22] Fama, E.F. and French, K.R. (1988a).Permanent and Temporary Components of Stock Prices, *Journal of Political Economy*, 96(2), 246-273

[23] Fama, E.F. and French, K.R. (1988b). Dividend Yields and Expected Stock Returns, *Journal of Financial Economics*, 22, 3-25

[24] Fama, E.F. and French, K.R. (1989). Business Conditions and Expected Stock Returns, *Journal of Financial Economics*, 25, 23-49

[25] Fama, E.F. and French, K.R. (2001). Disappearing Dividends: changing firm characteristics or lower propensity to pay?, *Journal of Financial Economics*, 60(1), 3-43

[26] Fisher, K.L. and Statman, M. (2000). Investor Sentiment and Stock Returns, *Financial Analysts Journal*, 56, 16-23

[27] Freedman, D.A. (1981). Bootstrapping Regression Models, *The Annals of Statistics*, 9(6), 1218-1228

[28] Freedman, D.A. (1984). On Bootstrapping Two-stage Least-square Estimates in Stationary Linear Models, *The Annals of Statistics*, 12(3), 827-842

[29] Goetzmann, W. and Jorion, P. (1993). Testing the Predictive Power of Dividend Yields, *Journal of Finance*, 48, 663-679

[30] Jegadeesh, N. (1991). Seasonality in Stock Price Mean Reversion: Evidence from the US and the UK, *Journal of Finance*, 46(4), 1427-1444

[31] Hansen, L.P. and Hodrick, R.J. (1980). Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Empirical Analysis, *Journal of Political Economy*, 88, 829-853

[32] Herzfeld, T.J. (1980). *The Investor's Guide to Closed-End Funds*, McGraw-Hill, New York, NY

[33] Hodrick, R.J. (1991). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement, *Review of Financial Studies*, 5(3), 357-386

[34] Ibbotson, R., Sindelar, J., and Ritter, J. (1994). The Market's Problems with the Pricing of Initial Public Offerings, *Journal of Applied Corporate Finance*, 7, 66-74

[35] Kamstra, M.J., Kramer, L.A., and Levi, M.D. (2003). Winter Blues: a Sad Stock Market Value, *American Economic Review*, 93(1), 324-343

[36] Lahiri, S.N. (1999). Theoretical Comparisons of Block Bootstraps Methods, *Annals of Statistics*, 27, 386-404

[37] Lakonishok, J., Shleifer, A., and Vishny, R.W. (1994). Contrarian Investment, Extrapolation, and Risk, *Journal of Finance*, 49(5), 1541-1578

[38] Lee, C., Shleifer, A., and Thaler, R.H. (1991). Investor Sentiment and the Closed-End Fund Puzzle, *Journal of Finance*, 46(1), 75-109

[39] Lemmon, M. and Portniaguina, E. (2006). Consumer Confidence and Asset Prices: Some Empirical Evidence, *Review of Financial Studies*, 19(4), 1499-1529

[40] Li, H. and Maddala, G.S. (1997). Bootstrapping Cointegrating Regressions, *Journal of Econometrics*, 80, 297-318

[41] Ljungqvis, A., Nanda, V., and Singh, R. (2006). Hot Markets, Investor Sentiment, and IPO Pricing, *Journal of Business*, 79(4), 1667-1702

[42] Lo, A.W. and MacKinlay, A.C. (1988). Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies*, 1(1), 41-66

[43] Mackinnon, J. (2006). Bootstrap Methods in Econometrics, *Queen's Economics Department Working Paper No.1028*

[44] Malkiel, B.G. (1977). The valuation of Closed-End Investment-Company Shares, Journal of Finance, 32(3), 847-859

[45] Mehra, R. and Prescott, E.C. (1985). The Equity Premium, *Journal of Monetary Economics*, 15, 145-161

[46] Mishkin, F. (1992). Is the Fisher Effect for Real?: A Reexamination of the Relationship between Inflation and Interest Rates, *Journal of Monetary Economics*, 30(2), 195-215

[47] Neal, R. and Wheatley, S.M. (1998). Do Measures of Investor Sentiment Predict Returns?, *Journal of Financial and Quantitative Analysis*, 33, 523-547

[48] Newey, W.K. and West, K.D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708

[49] Poterba, J.M. and Summers, L.H. (1988). Mean Reversion in Stock Prices, *Journal of Financial Economics*, 22(1), 27-59

[50] Qiu, L. and Welch, I. (2006). Investor Sentiment Measures, *NBER Working Paper,* 10794

[51] Scheinkman, J.A. and Xiong, W. (2003). Overconfidence and Speculative Bubbles, *Journal of Political Economy*, 111(6), 1183-1219

[52] Shiller, R.J. (1981a). Do Stock Price Move Too Much to Be Justified by Subsequent Changes in Dividends?, *American Economics Review*, 71(3), 421-436

[53] Shiller, R.J. (1981b). The Use of Volatility Measures in Assessing Market Efficiency, *Journal of Finance*, 36(2), 291-304

[54] Shiller, R.J. (1999). Measuring Bubble Expectations and Investor Confidence, NBER Working Paper, 7008

[55] Shiller, R.J., Fischer, S., and Friedman, B.M. (1984). Stock Prices and Social Dynamics, *Brookings Papers on Economic Activity*, 2, 457-510

[56] Stambaugh, R. (1999). Predictive Regressions, *Journal of Financial Economics*, 54, 375-421

[57] Summers, L.H. (1986). Does the Stock Market Rationally Reflect Fundamental Values?, *Journal of Finance*, 41(3), 591-601

[58] Swaminathan, B. (1996). Time-varying Expected Small Firm Returns and Closed-end Fund Discounts, *Review of Financial Studies*, 9(3), 845-887

[59] Kahneman, D. and Tversky, A. (1984). Choices, Values, and Frames, *American Psychologist*, 39(4), 341-350

[60] Valkanov, R. (2003). Long-Horizon Regressions: Theoretical Results and Applicatiosn, *Journal of Financial Economics*, 68, 201-232

[61] Wang, Y., keswani, A., and Taylor, S. (2005). The Relationships between Sentiment, Returns and Volatility, *International Journal of Forecasting*, 22, 109-123

[62] Zweig, M.E. (1973). An Investor Expectation Stock Price Predictive Model Using Closed-End Fund Premiums, *Journal of Finance*, 28(1), 67-78

A Appendix

A.1 Discussion on model misspecification without correct lag

Consider an $AR(1)$ process X following the DGP

$$X_t = \rho X_{t-1} + \mu_t \qquad \mu_t \sim N(0, \sigma_\mu^2) \tag{A1}$$

where $Cov(X_{t-1}\mu_t) = 0$.

If there is a variable Y following the DGP

$$Y_t = \alpha_0 + \beta_0 X_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma_\epsilon^2) \tag{A2}$$

where $Cov(X_{t-1}\varepsilon_t) = 0$ and for simplicity ε_t is independent with μ_t, then regressing Y_t on X_{t-1} using OLS in a finite sample yields the estimates with the following values for estimated value, standard deviation and t statistic (null hypothesis valued 0):

$$\begin{aligned}
\widehat{\beta_0} &= \frac{Cov(X_{t-1}Y_t)}{Var(X_{t-1})} \\
S.D._{\widehat{\beta_0}} &= \frac{\sigma_\epsilon}{\sqrt{Var(X_{t-1})}} \\
t_0 &= \frac{Cov(X_{t-1}Y_t)}{\sigma_\epsilon \sqrt{Var(X_{t-1})}}
\end{aligned}$$

However, if the OLS model is misspecified without the correct lag, but instead follows the form

$$Y_t = \alpha_1 + \beta_1 X_t + e_t \tag{A3}$$

then in infinite samples we should have the relationship

$$\begin{aligned}
\alpha_1 &= \alpha_0 \\
\beta_1 &= \frac{\beta_0}{\rho} \\
e_t &= \varepsilon_t - \frac{\beta_0}{\rho}\mu_t \sim N(0, \sigma_\epsilon^2 + \frac{\beta_0^2}{\rho^2}\sigma_\mu^2)
\end{aligned}$$

while in a finite sample the estimates will have the following values for estimated value, standard deviation and t statistic (null hypothesis valued 0):

$$
\begin{aligned}
\widehat{\beta_1} &= \frac{Cov(X_tY_t)}{Var(X_t)} \\
S.D._{\widehat{\beta_1}} &= \frac{\sigma_e}{\sqrt{Var(X_t)}} \\
t_1 &= \frac{Cov(X_tY_t)}{\sigma_e\sqrt{Var(X_t)}}
\end{aligned}
\tag{A4}
$$

Note that from Equation A1 we can derive

$$
\begin{aligned}
Cov(X_tY_t) &= \rho Cov(X_{t-1}Y_t) \\
Var(X_t) &= \rho^2 Var(X_{t-1}) + \sigma_\mu^2
\end{aligned}
$$

where if $\sigma_\mu^2 = (1-\rho^2)Var(X_t)$ then homoskedasticity is obtained and $\sigma_\mu^2 \neq (1-\rho^2)Var(X_t)$ implies heteroskedasticity in X. Now we can rewrite Equations A4 into

$$
\begin{aligned}
\widehat{\beta_1} &= \frac{\rho Cov(X_{t-1}Y_t)}{\rho^2 Var(X_{t-1}) + \sigma_\mu^2} \\
S.D._{\widehat{\beta_1}} &= \frac{\sqrt{\sigma_\epsilon^2 + \frac{\beta_0^2}{\rho^2}\sigma_\mu^2}}{\sqrt{\rho^2 Var(X_{t-1}) + \sigma_\mu^2}} \\
t_1 &= \frac{Cov(X_{t-1}Y_t)}{\sqrt{\sigma_\epsilon^2 + \frac{\beta_0^2}{\rho^2}\sigma_\mu^2}\sqrt{Var(X_{t-1}) + \frac{\sigma_\mu^2}{\rho^2}}}
\end{aligned}
\tag{A5}
$$

It is obvious that $t_1 < t_0$ is always true. This suggests that model misspecification will possibly lead to under-rejection of the null hypothesis of zero slope. Moreover, once given ρ, the larger σ_μ^2 the bigger difference between t_1 and t_0. Intuitively, if we are testing Equation A3 when the true alternative follows Equation A2, the power of the t-test depends on how noisy X_t is as a proxy of X_{t-1}. In other words, larger σ_μ^2 means that a higher fraction of variation in X_t comes from noise, and therefore the t-test will be less likely to reject the null that the variation in Y_t cannot be predicted by the variation in X_t.

However, if we are conducting a long-horizon analysis then it is possible to find significant relationships. For instance, if we run the following regression

$$
Y_{t+1} + Y_t = \alpha_2 + \beta_2 X_t + \epsilon_t \tag{A6}
$$

then the true linear relationship between Y_{t+1} and X_t will be captured and thus significant slopes are likely to be found. Following a proof similar to that above, it is easy to show that $t_2 > t_1$ is always true, where t_2 represents the t statistic from Equation A6. Nevertheless as the residuals are autocorrelated,

t_2 will tend to over-reject the null as well. Keeping this in mind, we only interpret the condition that $t_2 > t_1$ is always true as a preliminary conclusion and do not document the proof here. More precise analytical proof should involve adjusted standard errors proposed by Hansen-Hodrick or Newey-West and will of course become extremely complex. Perhaps a more appropriate approach is to use Monte Carlo simulations for a more straightforward demonstration. It is nevertheless certainly beyond the scope of this appendix.

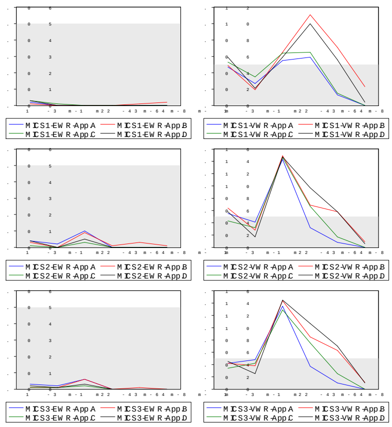
The one-period lag between Y and X in the linear relationship assumed above is only illustrative and can be extended to any length. It is easy to show that the general conclusion still holds with different lag lengths.

Our results in Section 5.2 that slope coefficients are insignificant at shorter horizons and turn significant over longer horizons for sentiment indicators $RIPO$ and $PDND$ confirm the analytical conclusions above. For example, $RIPO$ has a first-order autoregressive slope of 0.68, 0.63 and 0.67 for the original and two orthogonalised series used in Table 7 to 9 respectively. The R^2 of the first-order autoregression is only 0.46, 0.40 and 0.46 respectively, implying high σ_μ^2 in Equation A1 compared to the variation in X. $PDND$ has a first-order autoregressive slope of 0.94, 0.83 and 0.85 for the original and two orthogonalised series used in Table 7 to 9 respectively. The R^2 of the first-order autoregression is 0.89, 0.69 and 0.72 respectively, consistent with that the documented fact is only mild in Table 7 but more announced in Table 8 and 9. The argument can also be used for the results regarding $RIPO$ and $PDND$ in Section 6.2.

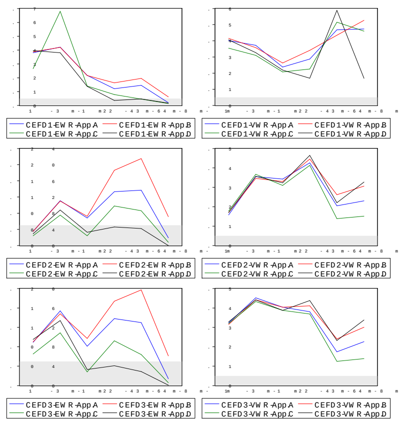
Perhaps a more extreme example comes with $DSENT$ and $DSENT^\perp$ in Section 6.2. In all Tables 11 to 13, the coefficients of $DSENT$ stay highly insignificant at 1-month horizon and turn extremely significant at 3-months and longer horizons, with all p-values from bootstrap distribution equal to zero. The same situation is present for $DSENT^\perp$ in explaining VWR. We confirm that $DSENT$ has an insignificant first-order autoregressive slope of 0.13, 0.08 and 0.13 for the original and the two orthogonalised series. The R^2 of the first-order autoregression is extremely low, only at values of 0.02, 0.01 and 0.02 respectively, showing that the variation in $DSENT$ essentially all comes from a high σ_μ^2 in Equation A1. $DSENT^\perp$ has an insignificant first-order autoregressive slope of -0.08, -0.05 and -0.07 for the original and the two orthogonalised series. The R^2 of the first-order autoregression is also extremely low, only at values of 0.01, 0.00 and 0.01 respectively, showing that the variation in $DSENT^\perp$ essentially all comes from a high σ_μ^2 in Equation A1 too. While we cannot exclude the possibility that bootstrap fails to correct the size biases in hypothesis tests due to autocorrelated residuals for $DSENT$ and $DSENT^\perp$, our analytical prediction here provides an additional possible reason why the results are behaving so.

A.2 Plot of p-values in single factor models across approaches

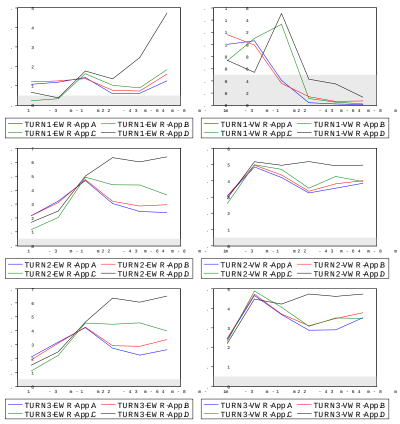
MICS



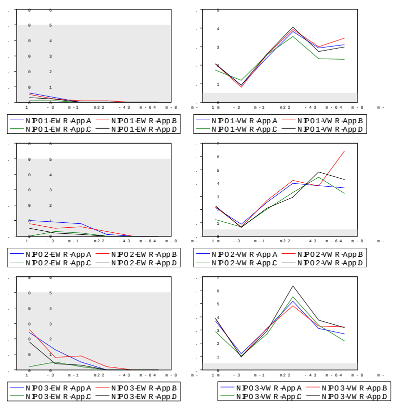
CEFD



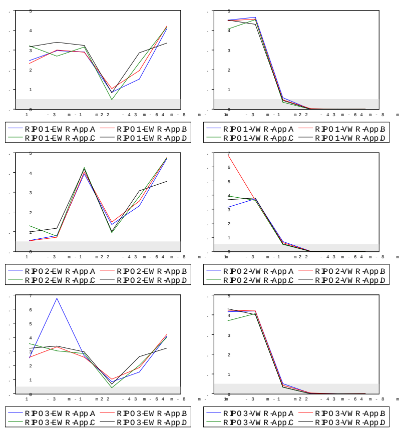
TURN



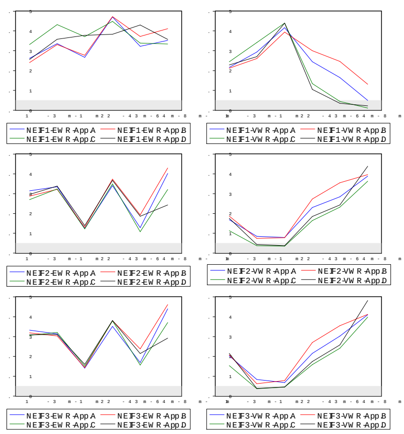
NIPO



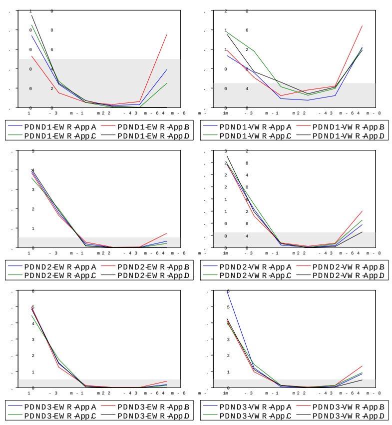
RIPO



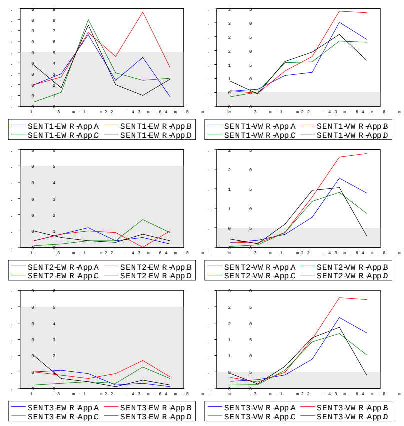
NEIF



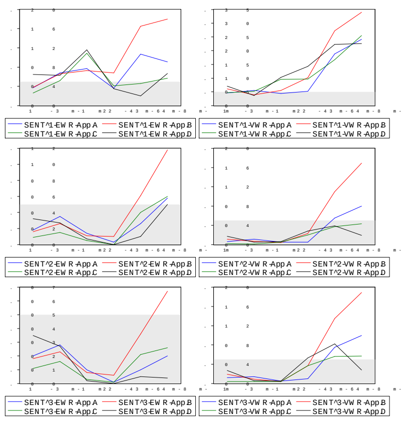
$PDND$



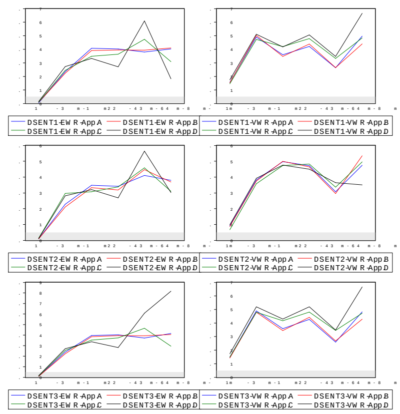
$SENT$



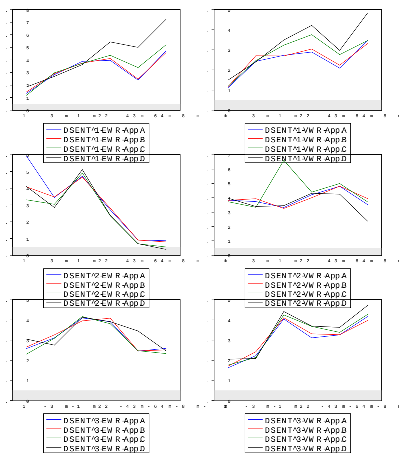
$SENT^{\perp}$



$DSENT$

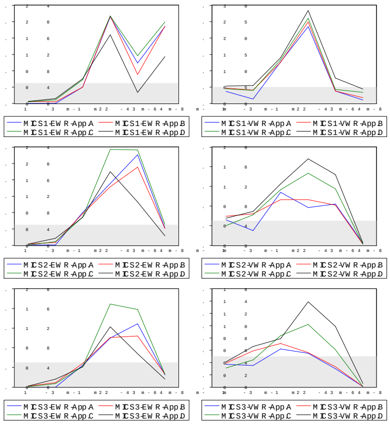


$DSENT^{\perp}$

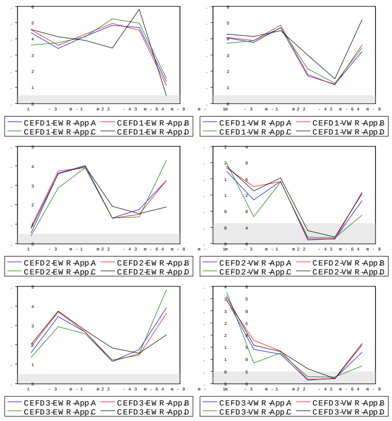


A.3 Plot of p-values in double factor models across approaches

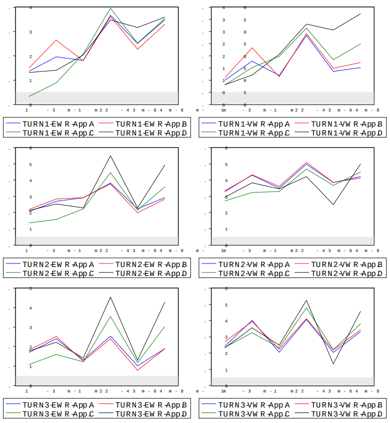
MICS



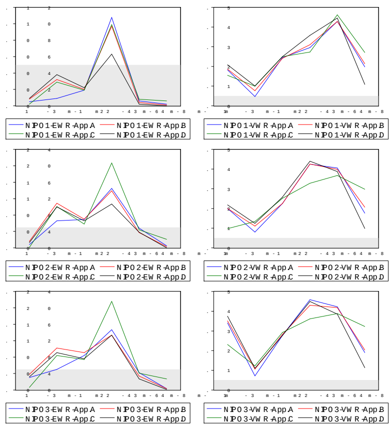
CEFD



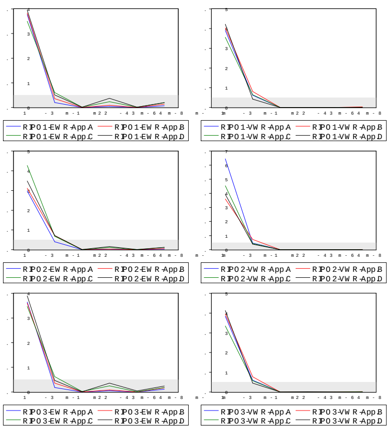
TURN



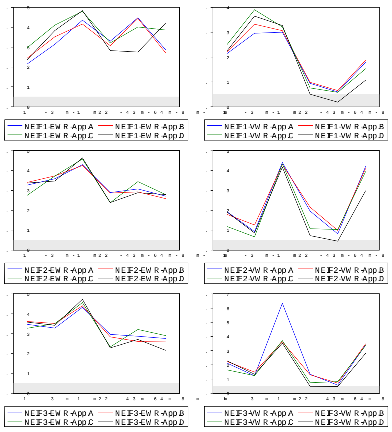
NIPO



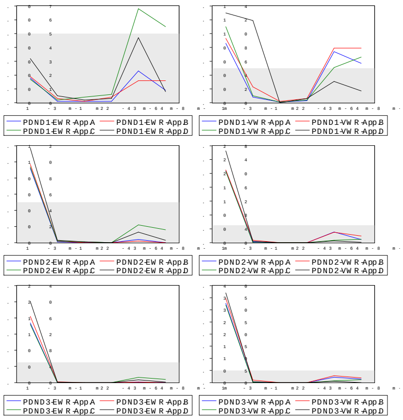
RIPO



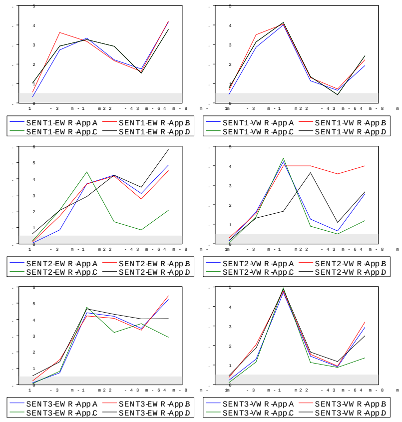
NEIF



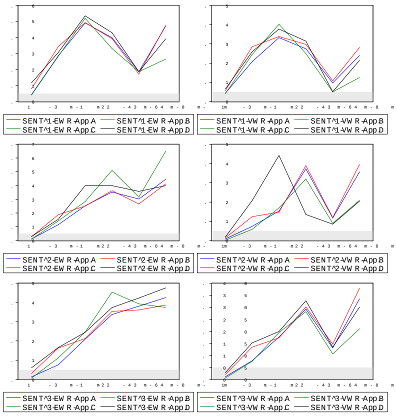
$PDND$



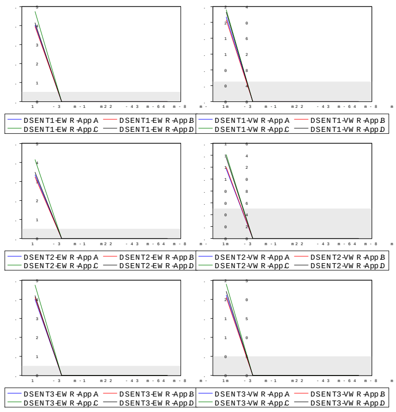
$SENT$



$SENT^{\perp}$



$DSENT$



$DSENT^{\perp}$