



Chasing noise in the stock market: an inquiry into the dynamics of investor sentiment and asset pricing

Rilwan Sakariyahu¹ · Audrey Paterson²  · Eleni Chatzivgeri³ · Rodiat Lawal⁴

Accepted: 21 August 2023 / Published online: 4 October 2023
© Crown 2023

Abstract

This study explores the inclusion of sentiment measures as a risk factor in asset pricing. Using UK market data for the period January 1993 to December 2020, we create a new sentiment variable, and construct both raw and clean sentiment indices from a principal component analysis of a variety of literature-acknowledged sentiment proxies. Essentially, the model estimations are categorized into two: first, the study documents the performance of the traditional pricing models on portfolios formed on different characteristics. Second, the study augments the first category by iterating sentiment variables into the model specification. The findings reveal that sentiment-augmented asset pricing models outperform the traditional models in explaining the excess returns of the portfolios. Furthermore, using Hansen & Jagannathan (1997) non-parametric model performance technique, we observe that the sentiment-induced models produce a small distance error compared to the traditional models, thus validating the use of sentiment measures in our pricing mechanism. It is therefore opined that extant asset pricing models may not be sufficient to explain market or pricing anomalies. Investors' sentiment is an important systematic risk factor that possesses useful information, and by implication, market analysts and stakeholders must take serious cognizance of its propensities when forecasting risk-adjusted returns.

Keywords Investor sentiment · Laggards to leaders index · Noise-trading · PCA · UK

JEL Classification E21 · G02 · G12 · G14 · G15 · G17

✉ Audrey Paterson
audrey.paterson@abdn.ac.uk

Rilwan Sakariyahu
r.sakariyahu@napier.ac.uk

Eleni Chatzivgeri
eleni.chatzivgeri@ed.ac.uk

Rodiat Lawal
rlawal001@dundee.ac.uk

¹ Business School, Edinburgh Napier University, Edinburgh, UK

² Business School, University of Aberdeen, Aberdeen, UK

³ Business School, University of Edinburgh, Edinburgh, UK

⁴ The Business School, University of Dundee, Dundee, UK

1 Introduction

Over the years, research efforts on asset pricing have progressively evolved and alluded to the plausibility of Capital Asset Pricing Model (CAPM) introduced by Treynor (1961, 1962) and further independently improved by Sharpe (1964), Lintner (1965) and Mossin (1966). Building on the premier model of Markowitz (1952) on modern portfolio theory and diversification, CAPM provides a persuasive approach to estimating expected return, giving credence to market risk premium as the cardinal explanatory factor. This is in addition to the presence of a riskless asset and systematic risk factor (beta). However, successive attempts at asset pricing have questioned the validity of CAPM due to ample empirical anomalies discovered by proponents of behavioural (or sentiment) factors such as Gibbons (1982), Lee et al., (1991), Baker & Wurgler (2006), Kumar & Lee (2006), Tetlock (2007), Hyde & Sherif (2010), and Kahneman & Tversky (2013). To this end, this study explores the inclusion of sentiment measures as a risk factor in asset pricing.

The sequential findings of Fama & French (1993, 1995 & 1996) on the three-factor model, provide the foremost statistical adjustment to the CAPM, with the introduction of size and value premiums as additional risk factors. Proposing the difference in returns between high and low book to market ratio and big and small stocks, their approach shows that value and size provide explanation to the cross-sectional sensitivity of common risk factors in average stock returns. Meanwhile, the aptness of the three-factor model in finance literature has also continued to generate debates in subsequent research on asset pricing, despite the significant amount of support it has enjoyed. He et al., (1996) document that the three-factor model accounts for a small-scale proportion of the cross-sectional variation of stock returns. Griffin (2002) also criticises the three-factor model as generally country-specific with little explanatory power towards cross-country endeavour while other studies (Cakici et al. 2013; Hanauer & Linhart 2015) also condemn the inconsistency and inapplicability of the model with other markets, save the developed markets. Furthermore, Petkova (2006) observes that the three-factor model loses its ability to predict cross-section of returns when innovations are included as part of the variables in the model. More so, Carhart (1997) opines that the inability of the three-factor model to recognise the aggression phenomenon of winner-loser effect led to the development of the four-factor model. Hence, in furtherance to the size and value premiums proposed by Fama & French (1993), Carhart (1997) introduced the monthly momentum (MOM) as an additional risk factor. Chen & Fang (2009) show that the four-factor model provides increased explanation for portfolio returns than the three-factor model.

While many studies favour the use of factor models in asset pricing, a stream of research questions the over-reliance and absolute use of factor models. In their opinion, factor models may not adequately account for the discount or premium observed in prices of some classes of assets, particularly those listed in emerging markets. These studies (Lee & Gan 2006; Humpe & Macmillan 2009; Yaya & Shittu 2010; Kasman et al. 2011; Patel 2012; Ahmadi 2016) posit that apart from the factor models, macroeconomic variables such as interest rate, money supply, inflation, trade balance and unemployment also have varying significant impact on the pricing of assets. However, the outcomes of the studies have remained unclear about the most suitable fundamental proxy and the surge in number of pricing anomalies has offered reasons for additional factors. As a result, recent studies have introduced human psychology into the dynamics of stock markets. These research efforts have brought much popularity to the significance of behavioural finance, that is human sentiments, in explaining market outcomes.

Sakariyahu et al., (2021) classify sentiment-based studies into five categories. They include *market-based* sentiment studies such as Baker & Wurgler (2007), Chen (2012), He et al. (2019), Paterson et al., (2023). These studies demonstrate that market indices (e.g., trade volume, dividend, and liquidity) have informational content and provide signals to noise traders, who upon acting on these signals significantly disrupt stock price behaviour. There is also the *media-based* category of sentiments who believe that social media threads often inspire noise trading which consequently impacts price direction. Proponents of this category include Rao & Srivastava (2012), Oliveira et al. (2013), Uhl (2014), Dosumu (2023). The third category of sentiment studies use *internet-based* sentiment measures (Da et al. 2014; Zhang et al. 2017; Trichilli et al. 2020; Sakariyahu et al. 2023). These studies document that irrational investors are propelled by the outcomes of online search facilities. The penultimate category is the *non-fundamental* based sentiment measures who adopt non-economic events such as politics, weather, or religion to explain asset prices. The advocates of these measures include Levy & Yagil (2011), Bia Ikowski et al. (2012), Chang et al. (2012) and Goetzmann et al. (2014). Lastly, there is a category of researchers who use *survey-based* measures of sentiment. These studies argue that consumer or market surveys contain adequate information that highlight investors expectation of the market. The adherents include Chen (2011), Finter et al. (2012), Dalika (2014), and Salhin et al. (2016).

In contributing to the ongoing efforts on behavioural finance, we investigate whether the inclusion of sentiments as a risk factor in a model of asset returns would increase the forecast power. The introduction of a new sentiment variable creates a distinction from other sentiment studies. It is widely acknowledged in the literature that investor sentiment has no perfect or definitive measure; nonetheless, this study is motivated by the reality that persistent market anomalies largely reflect symptoms of human sentiments (Shleifer & Summers 1990; Sias et al. 2001). Thus, using laggards to leaders as a new proxy for sentiment, we propose that irrational investors express emotional apathy towards a particular stock, sector, or market when the proportion of stocks declining in value outnumbers the value-advancing stocks. This trading pattern eventually disrupts expected market standards by creating artificial price imbalance, in form of either over-escalating security prices during bullish periods or extremely suppressing them during bearish periods (Changsheng & Yongfeng 2012; Bathia & Bredin 2018). To better understand the predictive power of sentiments on stock returns, we examine the extant factor variables (such as risk premium, the size factor, the value factor, and the momentum factor) along with the index we create for investor sentiment.

Furthermore, we use a principal component analysis to form a raw sentiment index by incorporating some generic sentiment variables from the literature, such as liquidity, as measured by market turnover (Pan & Potesman 2006), dividend premium (Baker & Wurgler 2006) and consumer confidence indicator (Lemmon & Portniaguina 2006). In a bid to also distinguish between market-imposed sentiment and economic-cycle induced sentiment, this study further constructs a second sentiment index that deflates economic cycle variations, following the approach of Bathia & Bredin (2018). For instance, liquidity and ratio of lagging to leading stocks may fluctuate due to national or economic reasons. We therefore extract clean sentiment index that is free from economic cycles. Basically, we regress percentage growth in the industrial production index, consumer price index, broad money supply and base lending rate on each of the sentiment proxies. The resulting residuals generated from these regressions constitute a cleaner sentiment index. Our findings reveal that the sentiment index significantly predicts excess return in many of the portfolios formed. The inclusion of our sentiment index in the risk factors produces incremental

abnormal returns, suggesting that the sentiment-induced models forecast better than the extant risk factors.

Interestingly, despite the vast number of studies documenting sentiments, none has investigated our research direction, to the best of our knowledge. This study therefore focuses on the UK stock market by constructing and estimating alternative models to the extant asset pricing models. Our choice of the UK market is informed by its status in the assembly of developed markets. By trading volume, the UK stock market ranks first in Europe and second in the world (Federation of Exchanges, 2018). Thus, using UK data provides a yardstick to gauge the models in this study from the perspective of a developed market and further serves as alternative performance evaluation to similar research endeavour in other developed markets. Finally, to accentuate our findings, we propose a methodological substitute to the parametric tests of asset pricing, using Hansen & Jagannathan (1997) non-parametric model performance technique. This technique assesses the suitability of the models used in this study, thus providing answers to issues surrounding robustness of our findings.

The other parts of this study are organized as follows. Section 2 provides literature review while Sect. 3 explains the methodology. Section 4 shows the empirical models and assumptions underpinning the study's objectives. Section 5 presents the outcome of the preliminary analysis. Section 6 shows the empirical findings and Sect. 7 concludes the study with recommendations.

2 Literature review

An increasing number of research shows that stock price variation can be generally attributed to the activities of two kinds of investors; arbitrageurs (rational investors) and noise traders (not fully-rational investors) (Ross 1976; Shleifer & Summers 1990; Sias et al. 2001; Gagnon & Karolyi 2010; Ramiah et al. 2015). Arbitrageurs are tactical market participants who specialize in taking advantage of market inefficiencies by wielding different investment strategies. They are highly knowledgeable in diagnosing arbitrage opportunities and because they often have access to huge resources, they take immediate actions, which significantly impact on stock price direction (Ross 1976). Essentially, arbitrageurs are research-oriented traders and are versed in technical or fundamental analysis. Due to the complexity of their strategies and the speed at which efficient markets readjust stock prices back to equilibrium, the goal of arbitrage is to expeditiously trade in assets whose prices do not reflect their true fundamental values, thus exploiting the discrepancies to earn risk-free profits. Although, arbitrage is a fallout of market mispricing; nevertheless, arbitrageurs are important traders whose investment strategies provide liquidity to the stock market.

While arbitrageurs are typically skilled at their practice (Gagnon & Karolyi 2010), noise-traders, on the other hand, are not fully rational investors, often without experience or professional knowledge of trading but whose motive for trading is premised on illogical and invalid information (Sias et al. 2001). As the name suggests, a noise trader is a typical market novice who reacts to noise from the market, particularly high trade volume, and causes the market to digress from normal expected trading patterns, notwithstanding the actions of other rational investors. Although the impact of noise trading on price movement is not elusive in the finance literature (De Long et al. 1991; Black 1986; Sias et al. 2001), the severity of the impact on stock markets still remains a subject of debate.

The advent of online stock trading apps (e.g., Robinhood) has also contributed to the recent surge in noise-trading. A large category of traders using these apps are those who follow volume and price signals, and they form the bulk of aggregate trading for the day (Kim et al. 2020). Considering the incremental proportion of irrational trading to well-informed trading, studies have shown that previous stock market crises could not have been unconnected to the roles of noise traders (De Long et al. 1989; Lee & Rui 2002; Scruggs 2007). Hence, in a market where noise-traders outnumber arbitrageurs, it is possible for arbitrage opportunities to exist as prices in the market may not reflect all available information. Conversely, an efficient market with equal proportion of arbitrageurs and noise-traders will prevent arbitrage opportunities because the rivalry between them will quickly return stock prices to their intrinsic values. However, given the funding restrictions usually faced by arbitrageurs, that could potentially have been used to explore price inefficiencies, there is a likelihood for arbitrage opportunities not to disappear immediately. This is often referred to as limits to arbitrage (Shleifer & Vishny 1997) and tends to ruin the plausibility of the efficient market hypothesis. Hence, more research are motivated to model asset prices from a behavioural perspective, given that traditional models and fundamental factors do not exhaustively explain erratic movements in stock prices.

The theory of noise trading introduced by Black (1986) and subsequently developed by Trueman (1988) set the pace for the recognition of behavioural finance (or investors' sentiments) in asset pricing. The theory emphasises that sentiments which generate from the noise around financial markets make market observations imperfect. Black (1986) specifically notes that noise trading which can be adduced to market uncertainty provides information to investors on market liquidity and surreptitiously creates an arbitrage opportunity for moving prices of risky assets back to their fundamental values. Although, Fama (1998) opines that pricing anomalies only exist by chance, largely because of arbitraging and methodological imperfections. Nevertheless, Brown & Cliff (2004) and Baker & Wurgler (2006) have demonstrated that the motives of market arbitrageurs are quite dissimilar to other stock market participants and by extension, the unpredictability of changes in investors' sentiments has a crucial tendency to create adverse cross-sectional variations in stock returns, thus limiting the activities of arbitrageurs. This continuous debate motivates numerous studies to channel attention towards the key drivers of stock prices.

Testing for the time-horizon (short and long term) impact of sentiment on assets in two different environments, Ling et al. (2010) examine the relation between investor sentiment and returns in public and private markets in the US. Applying vector autoregressive (VAR) models to capture the short-run dynamics, they show that investor sentiment has a positive relationship with returns in the short run, with a large magnitude of returns evident and consistent in public markets than private markets. On the long horizon, they find that a negative relationship exists between investor sentiment and returns, with a consistent price reversion to its fundamental values. Their study concludes that private markets appear to be persistently characterised by sentiment-induced mispricing than in public markets. In another related study, Brown & Cliff (2005) explore the link between asset valuation and investor sentiment in US using survey-based sentiment proxies. Their study documents that market mispricing errors are positively related to investor sentiment and that asset valuation is affected by sentiment. They provide evidence that in a multi-year horizon, future returns are negatively related to sentiment.

Also using survey data for sentiment, Ho & Hung (2009) examine the importance of investor sentiment in asset-pricing. Using the monthly equity data of the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) from the Centre for Research in Security Price (CRSP) and COMPUSTAT datasets for the period from July

1964 to December 2005, they assess whether sentiment proxies could improve the impacts of the risk-factor models (size, value, liquidity, and momentum effects) on risk-adjusted returns of individual stocks. Their result shows that in the conditional CAPM, size effect becomes less significant and further insignificant in other models. Meanwhile, they also reveal that sentiment-augmented models improve and outperform the extant factor models in capturing stock anomalies and explaining the dynamics of expected stock returns, thus concluding that investor sentiment plays a drastic role in asset pricing.

In another similar research, Bathia & Bredin (2018) explain the importance of investor sentiment measures on conditional asset pricing model. Using monthly data for the period January 1980 to December 2014, they specifically test whether incorporating sentiment measures such as IPO first day returns, IPO volume, closed-end fund discount, equity fund flow, equity put-call ratio, dividend premium and change in margin debt, could improve the performance of risk factors on risk-adjusted returns of U.S. individual stocks. Their findings disclose significant impact of these measures on the asset pricing models.

Assessing the impact of investor sentiment in the oil and gas industry, Zhu et al., (2020) document significant cross-sectional effects of sentiments on returns in the industry. Using financial statement data of common stocks listed on the NYSE, AMEX, and NASDAQ, they show that several anomalies (specifically 13 out of 15) are inherent in the oil and gas industry. Furthermore, their analysis reveals that investor sentiment has significantly positive impact on four of these prominent capital market anomalies. They therefore conclude that investor sentiment is a reality for better explanation of asset valuation. Liang et al. (2017) develop a framework to explain the effect of sentiment on asset pricing in the Chinese stock market. Using sentiment proxies such as investor's limited attention, anchoring, and other macroeconomic variables, they find and conclude that retail investors are often forced to pay more cognitive loss due to their insensitivity to market sentiment. They provide evidence that a bullish market due to higher level of investor rationality increases stock demand and thus push prices higher.

In another Chinese related study, Xu & Green (2013) adopt monthly data from the Shanghai and Shenzhen stock markets from January 1997 to December 2007 to study the impact of investor sentiment on stock returns in China. Using the three-factor Fama–French model as a benchmark, their study distinguish between positive and normal sentiments to explain the mispricing of returns. They show that the inclusion of sentiment factors such as turnover, the advances/declines ratio and the dividend premium, reduces the impact of the benchmark factors, suggesting that the Fama–French three-factor model does not comprehensively explain asset-pricing in China. Their findings further reveal that there is a difference of effect between the positive and normal sentiments and that sentiment appears to affect smaller companies than the larger ones. They therefore conclude that investor sentiment is a vital factor for explaining pricing anomalies in the Chinese market. Da et al. (2014) use information about households in the US to construct sentiment proxies. Splitting the information obtained into fear and economic attitudes, they predict the impact of household sentiments on short term return reversals and volatility. Their results show that the sentiment proxies broadly predict aggregate market returns with a mean reversal. They conclude that the market exhibits sentiment-induced temporary mispricing with a large effect on stocks that are susceptible to investor sentiments.

Decomposing investor sentiment into call and put, Yang & Zhang (2013) provide evidence of sentiment on asset pricing. By juxtaposing the two sentiment measures into a conditional framework, their study shows that sentiment-augmented asset pricing model could explain some anomalies in the stock market. Das et al. (2015) also explore the role of investor sentiment on the institutional trading behaviour in the REIT market and its

subsequent impact on asset pricing. Splitting the data period into pre global financial crisis period, crisis period and post crisis period, they test two alternative theories—flight to liquidity and style investing theory—evaluate how sentiment induces trading behaviour in the real estate market. For the three time-categories, their findings show that investor sentiment plays a major role in the movement of capital around the real estate market, thus lending credence to the two theories they adopt.

Like prior studies have noted, investor sentiment can influence stock market behaviour and by extension asset pricing. Despite the expansive research on sentiments, the findings have always varied across sentiment measures and stock markets. This therefore activates compelling efforts to further examine the role of sentiments in pricing anomalies. Introducing a fresh sentiment proxy in this study is a novelty in literature; we anticipate that our measure would be subsequently adopted as a systematic risk factor in asset pricing.

3 Data and variables

This study assesses the economic importance of noise trading vis a vis, investors' sentiments on asset pricing. To establish the significance of the sentiment factor and other risk factors, we use data from different sources and cover the period from January 1993 to December 2020. We download data relating to price and market capitalisation of all UK listed firms (FTSE All-Share) from the London Stock Exchange (LSE) and www.investing.com, while data relating to book values are downloaded from Datastream and Bloomberg. In constructing the portfolios, we used only non-financial companies with positive book to market values, thus arriving at a total of 325 companies within the sample period. For quality control, before merging the data, we examined the measures and standards adopted by these sources for data classification. For instance, we observe that the three sources have similar data designs for market and macroeconomic indices, thus entrenching consistency in our data structure. The data composition and sources are described in the next section.

3.1 Risk factors

Data on risk free rate (R_{ft}), monthly portfolio returns, (R_{pt}), market return (R_{mt}) and market risk premium ($R_{pt} - R_{ft}$) are sourced from Datastream and Bloomberg. Also, based on specific characteristics such as size (SMB_t), value (HML_t) and momentum (UMD_t), we also obtained data on both equally-weighted (EW) and value-weighted (VW) test portfolios from the Xfi Centre for Finance and Investment at the University of Exeter, which also brings up to date, the observations in Gregory et al. (2013). To give in-depth clarification, we use the excess returns of 5 book-to-market and 5 size portfolios, then the excess returns of portfolios formed using 10 book-to-market and 10 size.

3.2 Sentiment factors

Several indices have been adopted for investor sentiment in the literature, with some indirect observations. In lieu of the extant measures, we therefore propose the ratio of lagging to leading stocks as a new sentiment measure. This proxy, in addition to other sentiment measures such as liquidity (Baker & Wurgler 2007), dividend premium (Simões Vieira 2011) and consumer confidence indicator (Zouaoui et al. 2011), are used to form two composite indices of sentiment. The new sentiment proxy (the ratio of lagging to leading stock)

is first introduced as a risk factor into the pricing model; other sentiment variables are later introduced into the model as a composite risk index, after constructing both raw and clean sentiment indices. Data for these variables are obtained from Datastream.

3.3 Macroeconomic factors

Basically, we observe that the extant sentiment indices suffer from the vicissitudes of the economy, as they are highly correlated with macroeconomic cycles. Hence, in constructing the clean sentiment index mentioned earlier, we obtain data on macroeconomic variables and regress on each of the above sentiment proxies to generate residuals. The residuals of these sentiment proxies are later used to construct the second principal component analysis which we refer to as the clean sentiment index. Data for macroeconomic variables such as industrial production index (*IPI*), consumer price index (*CPI*), broad money supply (M_3) and base lending rate (*BLR*) are obtained from Bloomberg and Datastream.

4 Empirical models

In examining the significance of sentiment measures in asset pricing, we begin our model specification with single-factor model (capital asset pricing model), then we introduce the Fama-French three factor model and the Carhart four factor model. Finally, we integrate sentiment factors into the four-factor model; starting with the basic sentiment variable (ratio of laggards to leaders), then the raw and clean sentiment indices which collectively capture liquidity, dividend premium and consumer confidence indicator, after controlling for macroeconomic variation. An empirical expression of the various models used in this study is shown in the next section.

4.1 The risk factor models

Our empirical outlook is founded on the standard capital asset pricing model. The model is mathematically expressed as:

$$R_{pt} - R_{ft} = R_{ft} + \beta R_{mt} - R_{ft} + \epsilon_{pt} \quad (1)$$

R_{pt} is the return on portfolio p for month t ; R_{ft} is the risk-free rate of return; the difference between the two is the excess return α ; R_{mt} is the return of a well-diversified market index; $R_{mt} - R_{ft}$ is referred to as the market risk premium and ϵ_{pt} is the residual term.

In addition to the single factor model, Fama & French (1993) initiated two risk factors to capture size and value. The equation is expressed as follows:

$$R_p - R_f = R_{ft} + \beta R_{mt} - R_{ft} + \beta_{p,SMB} SMB_t + \beta_{p,HML} HML_t + \epsilon_{pt} \quad (2)$$

SMB_t and HML_t refer to the size and value factors of the portfolio respectively, at a particular month.

Augmenting the three-factor model above, Carhart (1997) proposed an additional risk factor, the 'winner minus loser' factor, to capture momentum effect. The four-factor model is expressed as:

$$R_p - R_f = R_f t + \beta R_m t - R_f t + \beta_{p,SMB} SMB_t + \beta_{p,HML} HML_t + \beta_{p,UMD} UMD_t + \epsilon_p \quad (3)$$

UMD_t refers to the momentum factor at a particular month.

For the sake of simplicity, we wrap up the above risk factors into a single composite equation called risk factor variables (RFV). This is shown below:

$$R_p - R_f = R_f + \beta RFV_t + \epsilon_p \quad (4)$$

where RFV_t encloses each of the above risk factors at a particular month.

4.2 Sentiment-augmented risk factors

The next set of models incorporates our sentiment proxies into the Carhart (1997) four factor model distinctly. These proxies are the basic sentiment variable, the raw sentiment index generated from the principal component analysis (PCA) and the clean sentiment index, also generated from PCA after controlling for macroeconomic factors.

The model which embeds the basic sentiment variable BSV is shown below:

$$R_p - R_f = R_f + \beta RFV_t + \beta_{p,BSV} BSV_t + \epsilon_p \quad (5)$$

where BSV_t refers to the basic sentiment factor at a particular month.

4.3 Constructing sentiment index using principal component analysis (PCA)

Principal Component Analysis (PCA) is an intuitive statistical technique that transforms linearly correlated set of observations using orthogonal dimensions and still preserve the inherent features of the observations. Following Baker & Wurgler (2006); Chen & Sherif (2016), this study computes a principal component analysis (PCA) that generates a first stage sentiment index capturing the common component in basic sentiment variable (BSV), consumer confidence indicator (CCI), dividend premium (DP) and liquidity (LIQ). Furthermore, both the Eigen values and the variances of the four components are calculated.

$$PCA_1 = \beta_1(X_1) + \beta_2(X_2) + \beta_1 P(X_p) \quad (6)$$

where

$$PCA_1 = \beta^T(X) \quad (7)$$

On the first principal component, PCA_1 represents the subject's score; $\beta^T(X)$ stands for the regression coefficient for the observed variable P and X_p represents the subject's score on the observed variable P . In calculating the first principal component, focus is centred around the highest variance in the observation. To prevent the creation of potential large values for the weights of β_1 and β_2 , which consequently affects the variance of PCA_1 , this study restricts the weights to ensure that their sum of squares is 1.

$$PCA_2 = \beta_{21}(X_1) + \beta_{22}(X_2) + \beta_{2p}(X_p) \quad (8)$$

To eliminate macroeconomic variations that may be rooted in the raw sentiment proxies, this study further generates a clean sentiment index by obtaining residuals of each sentiment proxy after regressing the macroeconomic variables (IPI , CPI , $M3$ and BLR) on

them. Again, this study calculates both the Eigen values and the variances of the four components for the clean sentiment index. In a similar pattern, the second principal component is produced with an orthogonal transformation that prevents it from being correlated with the first principal component but must produce the next greatest possible variance. The outcomes of the principal component analysis are shown in Table 1 below.

4.3.1 Raw sentiment-index model

The results of the PCA show that the first two principal components are well suited to construct a raw sentiment index. Their results show Eigen values of 1.8685 and 1.2775 respectively and variances of 0.4671 and 0.3194 respectively. Hence, the first two principal components explain about 79percent of the overall variance for all variables. The output of the raw sentiment index is provided below:

$$RSI = (0.4671/0.7865) Component1 + (0.3194/0.7865) Component2 \tag{9}$$

$$Component_1 = -0.3276BSV + 0.4749CCI - 0.5874DP + 0.5675LIQ \tag{10}$$

$$Component_2 = 0.6943BSV + 0.5209CCI - 0.3271DP - 0.3736LIQ \tag{11}$$

Following the computation of the raw sentiment index, it is therefore integrated into the Carhart (1997) four factor model as follows:

$$R_{p,t} - R_{f,t} = R_{f,t} + \beta[R_{FV}_t] + \beta_{p,RSI}RSI_t + \epsilon_p t \tag{12}$$

Table 1 Results of first and second stage principal component analyses

Variables	Comp.1	Comp.2	Comp.3	Comp.4	KMO
Panel A: First stage principal component analysis (Raw sentiment index)					
<i>BSV</i>	-0.3276	0.6943	-0.5231	0.3701	#####
<i>CCI</i>	0.4749	0.5209	0.6277	0.3304	#####
<i>DP</i>	-0.5874	-0.3271	0.3822	0.6339	#####
<i>LIQ</i>	0.5675	-0.3736	-0.4316	0.5933	#####
<i>Eigen value</i>	1.86845	1.27748	0.487328	0.36674	-
<i>Variance prop</i>	0.4671	0.3194	0.1218	0.0917	-
<i>Overall KMO</i>	-	-	-	-	#####
Panel B: Second stage principal component analysis (Clean sentiment index)					
<i>BSV</i>	-0.071	0.876	0.448	0.162	#####
<i>CCI</i>	0.547	0.298	-0.652	0.432	#####
<i>DP</i>	-0.635	-0.141	-0.098	0.753	#####
<i>LIQ</i>	0.541	-0.351	0.604	0.469	#####
<i>Eigen value</i>	1.62	1.079	0.741	0.561	-
<i>Variance prop</i>	0.405	0.27	0.185	0.14	-
<i>OverallKMO</i>	-	-	-	-	#####

This table shows the output of the principal component analyses (PCA). Panel A documents output for the first stage of the PCA (raw sentiment index) while panel B represents that of the second stage of the PCA (clean sentiment index). Comp. 1, 2, 3 and 4 all represent the component analyses and KMO stands for Kaiser-Meyer-Olkin, which is a measure of the proportion of variance among variables with common variance

Here RSI_t refers to the raw sentiment factor at a particular month.

4.3.2 Clean sentiment-index model

The results of the clean sentiment index also favour the first two principal components. Their results show Eigen values of 1.620 and 1.079 respectively and variances of 0.405 and 0.270, hence, explaining about 67.5percent of the overall variance for all variables. The output of the clean sentiment index is shown below:

$$CSI = (0.405/0.675) \text{ Component1} + (0.270/0.675) \text{ Component2} \quad (13)$$

$$\text{Component}_1 = -0.071BSV + 0.547CCI - 0.635DP + 0.541LIQ \quad (14)$$

$$\text{Component}_2 = 0.876BSV + 0.298CCI - 0.141DP - 0.351LIQ \quad (15)$$

This index is incorporated in the model as follows:

$$R_p - R_f = R_f + \beta RFV_t + \beta_{p,CSI} CSI_t + \epsilon_p \quad (16)$$

where CSI_t refers to the clean sentiment factor at a particular month.

4.3.3 Performance of asset pricing models

It is widely acknowledged in literature that estimates produced by asset pricing models are approximately close to reality. Therefore, as a robustness test, it becomes empirically logical to compare the performance of the competing asset pricing models used for estimation, to find out which model provides the best estimates. Several techniques have been employed to achieve this task, however, the model proposed by Hansen & Jagannathan (1997) (H–J) has enjoyed dominance in finance literature. The H–J distance is a statistical innovation that compares the economic performance of a set of competing models, diagnosing misspecification relating to them and identifying the most suitable among the models. Hence, H–J distance can be best described as a model revealing the maximum pricing error associated with a portfolio. It is measured as:

$$\delta = \min ||m - m^*||, \quad (17)$$

where m represents the fitted values and m^* represents the actual values.

In an insightful manner, Hansen & Jagannathan (1997) also show that for a proposed asset pricing model, the random payoff of a portfolio is

$$\delta = \max |\pi(\xi) - |\pi^v(\xi)|, \quad ||\xi|| = 1 \quad (18)$$

where $\pi(\xi)$ and $\pi^v(\xi)$ are alpha of asset prices measured by respective asset pricing model. Chen & Sherif (2016) further show that HJ measures the mean square distance between the fitted and actual values. Mathematically, HJ minimum distance can be denoted as ϵ ($m - m^*$) such that:

$$m - m^* = [m(\pi)R - 1]E(RR)^{-1}(R) \quad (19)$$

We further breakdown the above equation into two parts such that:

$$[\varepsilon[m(\pi)R - 1]] = \alpha \quad (20)$$

$$E(RR)^{-1} = S \quad (21)$$

where R represents excess returns and S inverse of sum of returns.

5 Preliminary analysis

In this study, the explanatory variables are classified into three: factor variables, sentiment variables and macroeconomic variables. Tables 2, 3, 4, 5 show the descriptive statistics for each of the explanatory variables and the portfolios. The tables report the values for the mean, standard deviation, minimum and maximum. During the observed monthly periods, Table 2 which specifically shows the summaries for the explanatory variables report that among the factor variables (*CAPM*, *SMB*, *HML* *UMD*), *UMD* has the highest descriptive values. The monthly momentum effect (*UMD*) reveals a mean of (0.01), standard deviation (0.049), minimum (−0.250) and maximum (0.160) values. On the other hand, monthly reports for Value (*SMB*) and Size (*HML*) reveal similar mean (0.002) and stand deviation (0.034) but different minimum and maximum values. The results suggest that the average excess return expected from the inclusion of momentum factor is much higher compared to other factors. This, perhaps, justifies the introduction of the risk factor in the pricing model. In the case of the sentiment variables, liquidity has the highest values for the mean

Table 2 Summary statistics for the explanatory variables

Variable	Mean	Std.Dev	Min	Max	Skewness	Kurtosis	J-B
Panel A: Factor variables							
CAPM	0.004	0.039	−0.136	0.099	−0.690	3.957	34.070
SMB	0.002	0.034	−0.148	0.156	−0.226	5.998	111.037
HML	0.002	0.034	−0.186	0.123	−0.514	9.634	544.579
UMD	0.010	0.049	−0.250	0.160	−0.956	7.266	264.047
Panel B: Sentiment variables							
BSV	0.24	0.25	−0.41	0.98	0.24	3.15	3.14
CCI	−7.96	8.12	−30.10	3.60	−0.82	2.65	33.66
DP	3.33	0.62	2.04	5.37	0.10	3.37	2.13
LIQ	18.34	0.63	16.85	19.57	−0.59	2.87	17.27
RSI	−2.11	4.20	−13.73	4.14	−0.79	2.60	31.87
CSI	−0.33	4.00	−11.38	5.61	−0.79	2.62	32.30
Panel C: Macro-economic variables							
IPI	0.01	0.01	−0.06	0.04	−0.60	7.21	231.59
CPI	2.01	1.03	−0.10	5.20	0.53	3.50	16.44
M3	14.08	0.54	13.12	14.81	−0.25	1.66	24.59
BLR	3.69	2.44	0.25	7.50	−0.33	1.51	31.98

This table shows the descriptive statistics of the explanatory variables. Data covers the period January 1993 to December 2020. Panel A shows the output for the risk factor variables, panel B shows the output for the sentiment variables and panel C documents the output for the macro-economic variables

Table 3 Summary statistics for the 5 BTM and size portfolios

Variable	Mean	Std.Dev	Min	Max
Panel A				
Equally-weighted book to market portfolio				
v1	0.98	5.20	-23.28	24.47
v2	1.02	4.46	-20.07	20.12
v3	1.21	4.55	-19.64	28.08
v4	1.25	4.88	-22.62	29.46
v5	1.64	5.31	-22.18	41.28
Value-weighted book to market portfolio				
v1	0.78	3.60	-11.63	11.19
v2	0.75	4.13	-20.38	9.62
v3	1.00	4.11	-15.92	13.18
v4	0.83	4.64	-15.49	12.09
v5	1.24	5.63	-25.64	23.30
Panel B				
Equally-weighted size portfolio				
s1	1.75	4.76	-17.91	28.64
s2	1.31	5.05	-22.76	35.06
s3	1.23	5.14	-18.89	32.10
s4	0.96	5.19	-25.39	27.76
s5	0.87	4.45	-16.98	14.43
Value-weighted size portfolio				
s1	1.35	4.64	-19.47	24.98
s2	1.33	5.02	-22.77	24.91
s3	1.25	4.87	-18.27	21.97
s4	0.98	4.91	-25.34	21.12
s5	0.75	3.71	-14.12	8.81

This table reports descriptive statistics for 5 portfolios formed on book to market ratio and size. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. V1 represents the portfolio with the lowest book to market ratio while V5 denotes the portfolio with the highest book to market ratio. S1 represents the portfolio with the lowest firm size while s5 denotes the portfolio with the highest firm size

(18.34), minimum (16.85) and maximum (19.57) while the consumer confidence indicator reports the highest standard deviation (8.12). We therefore interpret the results that on the average, market turnover in the UK is much higher than other sentiment variables used in the study. This is not unexpected given the volume of transactions executed in the UK stock market monthly. For the macroeconomic variables, broad money supply shows the highest mean, minimum and maximum values (14.08; 13.12 and 14.81) while basic lending rate reports the highest standard deviation (2.44). The results for Tables 3, 4, 5 which capture the descriptive statistics for the factor variables are also reported in similar fashion.

Table 5 shows the correlation coefficients for the explanatory variables. We report cross-relationships among the intersecting components - risk factors, sentiment measures and macroeconomic variables. Our keen interest is on the sign and size of relationships

Table 4 Summary statistics for the 10 BTM and size portfolios

Variable	Mean	Std.Dev	Min	Max
Panel A				
Equally-weighted book to market portfolio				
v1	0.92	5.59	-24.59	27.15
v2	1.03	5.00	-22.02	21.61
v3	1.02	4.53	-19.13	18.56
v4	1.01	4.59	-21.06	21.77
v5	1.27	4.81	-20.09	31.46
v6	1.15	4.54	-19.19	24.69
v7	1.17	4.94	-21.54	28.40
v8	1.32	5.07	-23.75	30.56
v9	1.38	5.03	-21.90	32.63
v10	1.95	6.21	-22.47	50.76
Value-weighted book to market portfolio				
v1	0.72	3.96	-12.29	10.94
v2	0.84	3.71	-10.77	12.53
v3	0.70	4.32	-22.55	9.41
v4	0.87	4.47	-18.14	12.77
v5	1.00	5.14	-21.34	20.32
v6	1.01	5.31	-23.63	20.05
v7	0.93	5.25	-16.40	14.48
v8	0.92	4.85	-18.65	14.62
v9	1.04	5.63	-23.69	24.19
v10	1.59	7.20	-29.61	26.52
Panel B				
Equally-weighted size portfolio				
s1	1.94	5.09	-13.26	24.07
s2	1.55	5.28	-22.32	37.14
s3	1.28	5.04	-22.48	30.16
s4	1.33	5.35	-23.04	40.05
s5	1.19	5.28	-19.60	30.61
s6	1.27	5.25	-18.14	33.73
s7	1.00	5.42	-24.86	30.91
s8	0.91	5.17	-25.90	24.72
s9	0.94	5.04	-19.59	18.09
s10	0.81	4.06	-15.35	10.80
Value-weighted size portfolio				
s1	1.07	4.78	-14.79	23.70
s2	1.44	4.92	-21.21	25.26
s3	1.27	4.82	-22.67	25.09
s4	1.35	5.33	-22.83	26.94
s5	1.25	5.10	-19.35	23.16
s6	1.25	4.90	-17.59	21.25
s7	1.06	5.06	-24.19	24.07
s8	0.94	4.95	-26.00	19.53
s9	0.99	4.76	-19.39	13.90

Table 4 (continued)

Variable	Mean	Std.Dev	Min	Max
s10	0.72	3.70	-13.98	10.14

This table reports descriptive statistics for 10 portfolios formed on book to market ratio and size. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. V1 represents the portfolio with the lowest book to market ratio while V10 denotes the portfolio with the highest book to market ratio. S1 represents the portfolio with the lowest firm size while s10 denotes the portfolio with the highest firm size

between the sentiment variables and other variables. First, we observe that across the sentiment measures (*BSV*, *CCI*, *DP* and *LIQ*) and risk factors (*CAPM*, *SMB*, *HML* and *UMD*), there is a presence of low but significant correlation coefficients. We interpret this outcome to imply that despite the significant interaction among the proxies, the size of cross-relationship is not strong enough to cause severe mutual impact. Hence, these variables can still be utilized as explanatory variables in the model without further adjustment. Interestingly, across the factor variables and macroeconomic variables, only *HML* and *CPI* reveal a significant negative relationship of 0.13. Other variables within the same intersect show very low and insignificant correlation coefficients. Meanwhile, a cursory look at the sentiment variables and macroeconomic variables shows strong and significant relationships. Including these variables in the models without further modification could produce spurious results. Hence, the outcome of these relationships led to extracting clean sentiment indices by running a regression of the sentiment measures on each of the macroeconomic variables and extracting the residuals therefrom. The residuals are further cross-correlated with the macroeconomic variables, and they report very weak relationships. Hence, the residuals are used to construct principal component analyses (*PCA*) which represents the new sentiment factors. Nevertheless, we still run the raw sentiment index along with the other models just for the sake of statistical comparison.

6 Empirical findings

We begin our empirical analysis by focusing on the regression outputs in Tables 6, 7, 8, 9. The tables show the estimates of alpha, and t-statistic obtained for each variant of the pricing models. Considering that this study employs six models, each table reports a pair of regression estimates (alpha values and t-statistic) for each of the six models against the individual portfolios. For clarity, alpha values are reported in first column while estimates of t-statistic, which are put in parenthesis, appear on the second column. Furthermore, below each regression output, the estimates of Gibbons et al. (1989) *GRS* test statistic are reported. *GRS* test ascertains the rejection (or acceptance) of the null hypothesis that all alpha values for each model are significantly close to zero; thus, the p-values of the *GRS* test are collectively reported. The p-value of the coefficient of the *GRS* must be insignificant for the model to be efficient. Additionally, the regression output is further extended to capture the mean adjusted R^2 , mean alpha value and the mean standard error of the alpha value. Following the approach of Gregory et al., (2013), it is imperative to mention that the magnitude of the mean adjusted R^2 for the six regression models are compared.

Table 5 Pairwise correlation of explanatory variables

Variable	CAPM	SMB	HML	UMD	BSV	CCI	DP	LIQ	RSI	CSI	IPI	CPI	M3	BLR
CAPM	1													
SMB	0.06	1												
HML	0.12**	-0.13**	1											
UMD	-0.21***	-0.09	-0.52***	1										
BSV	0.18***	0.23***	0.12**	-0.13**	1									
CCI	0.01	-0.01	0.10*	0.08	0.06	1								
DP	-0.09	-0.05	-0.07	-0.09	0.06	-0.35***	1							
LIQ	-0.11*	-0.02	-0.02	0.01	-0.49***	0.19***	-0.21***	1						
RSI	0.01	-0.01	0.10*	0.08	0.04	1***	-0.60***	0.24***	1					
CSI	0.01	0.01	0.10*	0.08	0.04	1***	-0.60***	0.25***	0.8**	1				
IPI	0.09	0.08	0.05	-0.04	0.07	0.09	0.07	-0.05	0.09	0.09	1			
CPI	-0.02	-0.03	-0.13**	0.01	-0.01	-0.74***	0.33***	-0.13**	-0.73***	-0.74***	-0.05	1		
M3	0.01	-0.01	-0.09	-0.03	-0.47***	-0.28***	0.05	0.6***	-0.26***	-0.26***	-0.05	0.14**	1	
BLR	-0.07	-0.09	0.04	0.07	0.28***	0.37***	-0.23***	-0.22***	0.36***	0.36***	-0.01	-0.1*	-0.86***	1

This table shows the output of the pairwise correlation analysis for the factor, sentiment and macro-economic variables. Data covers the period from January 1993 to December 2020

Table 6 Regression output for 5 book to market portfolios

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Panel A: Equally-weighted book to market portfolio												
v1	0.52	(2.62)	0.45	(4.36)	0.54	(5.02)	0.37	(5.02)	0.62	(2.25)	0.61	(5.21)
v2	0.60	(3.88)	0.48	(5.80)	0.56	(6.64)	0.36	(6.64)	0.57	(2.84)	0.57	(6.12)
v3	0.81	(4.69)	0.65	(7.42)	0.72	(7.93)	0.50	(7.93)	0.72	(3.65)	0.72	(7.26)
v4	0.84	(4.21)	0.64	(6.93)	0.67	(6.97)	0.36	(6.97)	0.78	(2.48)	0.77	(7.32)
v5	1.25	(5.20)	1.02	(7.24)	1.18	(8.16)	0.81	(8.16)	1.26	(3.71)	1.25	(7.89)
Mean alpha	0.81	-	0.65	-	0.74	-	0.49	-	0.80	-	0.79	-
GRS test-stat	2.65	-	1.64	-	2.64	-	1.52	-	2.19	-	2.22	-
p value	0.20	-	0.03	-	0.16	-	0.13	-	0.00	-	0.32	-
Mean Adj R ²	5.30	-	4.45	-	4.79	-	4.94	-	3.82	-	3.79	-
Mean S.E	0.19	-	0.10	-	0.11	-	0.16	-	0.12	-	0.12	-
Hansen-Jagannathan distance												
δ	0.032	-	0.013	-	0.017	-	0.053	-	0.039	-	0.011	-
Panel B: Value-weighted book to market portfolio												
v1	0.44	(3.68)	0.52	(6.63)	0.55	(6.66)	0.22	(6.66)	0.58	(1.79)	0.58	(6.37)
v2	0.33	(2.94)	0.34	(2.97)	0.30	(2.55)	0.22	(2.55)	0.35	(1.19)	0.34	(2.63)
v3	0.52	(3.45)	0.50	(3.36)	0.40	(2.56)	0.26	(2.56)	0.48	(1.09)	0.47	(2.75)
v4	0.41	(2.46)	0.35	(2.48)	0.47	(3.22)	0.09	(3.22)	0.42	(0.42)	0.43	(2.65)
v5	0.76	(3.46)	0.63	(3.52)	0.74	(3.96)	1.10	(3.96)	0.71	(3.84)	0.72	(3.53)
Mean alpha	0.50	-	0.47	-	0.50	-	0.38	-	0.51	-	0.51	-
GRS test-stat	1.35	-	1.43	-	1.30	-	1.53	-	1.39	-	1.29	-
p value	0.10	-	0.01	-	0.30	-	0.11	-	0.00	-	0.20	-
Mean Adj R ²	6.28	-	6.14	-	6.11	-	5.60	-	5.37	-	5.41	-
Mean S.E	0.16	-	0.13	-	0.14	-	0.21	-	0.16	-	0.15	-

Table 6 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Hansen-Jagannathan distance												
δ	0.010	-	0.11	-	0.065	-	0.09	-	0.01	-	0.018	-

This table reports the regression output for 5 portfolios formed on book to market ratio. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. V1 represents the portfolio with the lowest book to market ratio while V5 denotes the portfolio with the highest book to market ratio. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

Furthermore, as stated earlier, apart from the use of adjusted R^2 , several measures are employed in literature to ascertain the most suitable model for asset pricing. However, the technique proposed by Hansen & Jagannathan (1997) (H–J) focuses on the behavioural patterns of prediction and improves the accuracy of asset pricing models. Hence, to better understand the strength of the models, this study adopts H–J test to also evaluate the economic performance of the five models.¹

We start our interpretation with the regression outputs on Table 6 showing the 5 BTM portfolios. The results for the equally and value-weighted portfolios reveal similar indicators. It is observed that the Fama-French 3-factor model and raw sentiment index failed the GRS test for the weighted portfolios. They show significant p values, thus a rejection of the null hypothesis that the two models (3F and RSI) have alpha estimates jointly close to zero.

CAPM, 4F, 4F_BSV and 4F_CSI passed the GRS test in both the equally and value weighted portfolios. Furthermore, the results show that the inclusion of the basic sentiment variable in the model gradually magnifies the mean alpha, thus implying that excess return is improved by including the sentiment variable in the model specification. Furthermore, the results of the adjusted R^2 show a reduction from 5.30 for the CAPM model to 3.70 for the 4F CSI model. Also, the findings show that H–J error decreases from 0.032 for the CAPM to 0.011 for the clean sentiment index. We therefore infer from these findings that our sentiment measure improves excess returns in this portfolio and based on the outcomes of the H–J estimate, the model with clean sentiment index performs better. Similar result patterns are also revealed in the value-weighted book to market portfolio.

Akin to the findings on 5 BTM portfolios, Table 7 shows the regression outputs for the 5 size portfolios. At least, five models passed the GRS test; the only exception is the 4F RSI which has a significant p value and thus fails the GRS test statistic. The five other models: CAPM, 3F, 4F, 4F BSV and 4F CSI passed the GRS test suggesting that the alpha estimates can be jointly distinguished from zero. Surprisingly, we observe that for both the equally and value-weighted portfolios, mean alphas of the five separate models moved in a reverting approach. For instance, in the equally weighted portfolios, mean alpha for CAPM is 0.82, reduced to 0.66 using the 3F model and later increased to 0.80 when the clean sentiment index was introduced. Also, it is observed that a drastic change occurred in the mean adjusted R^2 of the two sets of portfolios when the sentiment variable is included. To illustrate this, the CAPM produced a mean adjusted R^2 of 5.37 for the equally-weighted portfolio and reduced to 4.47 using the 4F CSI. In addition, for the value-weighted size portfolio, the CAPM produced a mean adjusted R^2 of 5.06 but was significantly increased to 5.01 under the 4F CSI. Also, the H–J distance error declines from 0.031 to 0.011 under the equally-weighted portfolio and from 0.026 to 0.060 under the value-weighted portfolio. By implication, the use of both raw and clean sentiment indices in the model specification produced efficient results, with an excess return higher than those of other models. Table 8 covers the regression outputs for the 10 BTM portfolios. The study observes that for the 10 BTM portfolios, using the equally and value weighted constructs, the outcomes reveal about three models (3F, 4F BSV and 4F CSI) that passed the GRS test but with insignificant p values. The introduction of the sentiment variable in the model does not significantly improve the estimates of the mean alpha but the adjusted R^2 of the sentiment-augmented models decreased steadily from a value of 5.66 for CAPM to 5.23 for 4F CSI

¹ We acknowledge that for many of the estimations, H–J distance value decreased except for the value-weighted portfolios in tables 6 and 7. However, in some estimations, the H–J does not follow a priori expectations. This can possibly be due to sampling error from the portfolio construction.

Table 7 Regression output for 5 size portfolios

Variable	CAPM	3F	4F	4F_BSV	4F_RSI	4F_CSI	(t-stat)	(t-stat)	(t-stat)	(t-stat)	
Panel A: Equally-weighted size portfolio											
s1	1.45	1.27	1.30	0.83	1.59	1.54	(6.17)	(7.83)	(3.31)	(8.57)	(8.67)
s2	0.94	0.73	0.80	0.81	0.87	0.86	(4.08)	(6.36)	(4.21)	(6.07)	(6.20)
s3	0.80	0.61	0.69	0.43	0.69	0.69	(3.84)	(7.19)	(2.98)	(6.30)	(6.54)
s4	0.48	0.32	0.44	0.29	0.41	0.42	(2.62)	(4.91)	(2.14)	(4.05)	(4.26)
s5	0.39	0.34	0.47	0.10	0.46	0.46	(4.12)	(5.99)	(0.89)	(5.16)	(5.37)
Mean alpha	0.82	0.66	0.74	0.50	0.81	0.80	-	-	-	-	-
GRS test-stat	1.73	1.92	2.28	4.55	2.12	2.02	-	-	-	-	-
p value	0.20	0.11	0.19	0.17	0.03	0.12	-	-	-	-	-
Mean Adj R ²	5.37	4.99	5.33	5.55	5.53	4.47	-	-	-	-	-
Mean S.E	0.19	0.11	0.11	0.17	0.13	0.12	-	-	-	-	-
Hansen-Jagannathan distance											
δ	0.031	0.012	0.096	0.11	0.048	0.011	-	-	-	-	-
Panel B: Value-weighted size portfolio											
s1	1.04	0.86	0.86	0.47	1.10	1.07	(4.68)	(5.84)	(2.09)	(6.61)	(6.63)
s2	0.97	0.80	0.70	0.70	0.86	0.84	(4.19)	(5.48)	(3.57)	(6.04)	(6.05)
s3	0.84	0.67	0.62	0.31	0.67	0.66	(4.26)	(6.76)	(2.27)	(6.47)	(6.63)
s4	0.54	0.39	0.35	0.24	0.39	0.38	(3.04)	(4.15)	(1.90)	(4.02)	(4.11)
s5	0.34	0.38	0.37	0.26	0.41	0.41	(5.70)	(7.31)	(3.35)	(7.07)	(7.22)
Mean alpha	0.75	0.62	0.58	0.40	0.69	0.67	-	-	-	-	-
GRS test-stat	1.18	1.67	1.64	3.88	1.66	1.02	-	-	-	-	-
p value	0.22	0.13	0.10	0.07	0.01	0.20	-	-	-	-	-
Mean Adj R ²	5.06	5.68	5.76	5.69	5.57	5.01	-	-	-	-	-
Mean S.E	0.18	0.10	0.10	0.15	0.11	0.11	-	-	-	-	-

Table 7 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Hansen–Jagannathan distance												
δ	0.026	–	0.012	–	0.048	–	0.018	–	0.001	–	0.060	–

This table reports the regression output for 5 portfolios formed on size. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. s1 represents the portfolio with the lowest firm size while s5 denotes the portfolio with the highest firm size. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

Table 8 Regression output for 10 book to market portfolios

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Panel A: Equally-weighted BTM portfolio												
v1	0.46	(2.01)	0.40	(3.03)	0.52	(3.80)	0.34	(1.64)	0.60	(3.87)	0.59	(3.95)
v2	0.58	(3.13)	0.51	(4.86)	0.56	(5.21)	0.40	(2.42)	0.64	(5.22)	0.63	(5.32)
v3	0.61	(3.80)	0.49	(4.91)	0.61	(5.90)	0.47	(2.97)	0.59	(5.00)	0.59	(5.21)
v4	0.60	(3.51)	0.46	(4.69)	0.51	(5.01)	0.26	(1.68)	0.55	(4.76)	0.55	(4.87)
v5	0.86	(4.54)	0.69	(6.28)	0.76	(6.66)	0.70	(3.99)	0.79	(6.06)	0.79	(6.23)
v6	0.76	(4.27)	0.60	(5.80)	0.67	(6.25)	0.29	(1.83)	0.65	(5.33)	0.65	(5.56)
v7	0.76	(3.72)	0.56	(5.37)	0.59	(5.41)	0.22	(1.37)	0.67	(5.39)	0.65	(5.48)
v8	0.92	(4.26)	0.71	(5.86)	0.75	(5.90)	0.50	(2.60)	0.90	(6.27)	0.88	(6.32)
v9	0.99	(4.50)	0.78	(5.87)	0.95	(7.01)	0.43	(2.13)	0.91	(5.95)	0.92	(6.22)
v10	1.56	(5.09)	1.30	(6.04)	1.44	(6.45)	1.25	(3.64)	1.65	(6.51)	1.62	(6.60)
Mean alpha	0.81	-	0.65	-	0.74	-	0.49	-	0.80	-	0.79	-
GRS test-stat	3.24	-	1.64	-	1.63	-	3.03	-	1.35	-	1.85	-
<i>p</i> value	0.01	-	0.10	-	0.00	-	0.21	-	0.04	-	0.11	-
Mean Adj <i>R</i> ²	5.66	-	5.93	-	5.26	-	5.42	-	5.29	-	5.23	-
Mean S.E	0.21	-	0.12	-	0.13	-	0.19	-	0.14	-	0.14	-
Hansen-Jagannathan distance												
δ	0.037	-	0.025	-	0.022	-	0.015	-	0.031	-	0.041	-
Panel B: Value-weighted BTM portfolio												
v1	0.38	(2.49)	0.48	(4.39)	0.56	(4.91)	0.21	(1.24)	0.59	(4.54)	0.59	(4.67)
v2	0.49	(3.83)	0.53	(4.98)	0.46	(4.19)	0.15	(0.92)	0.49	(3.92)	0.49	(4.03)
v3	0.29	(1.99)	0.29	(2.02)	0.26	(1.73)	0.05	(0.23)	0.26	(1.53)	0.26	(1.59)
v4	0.43	(3.14)	0.43	(3.08)	0.40	(2.79)	0.41	(1.84)	0.52	(3.15)	0.50	(3.14)
v5	0.51	(2.96)	0.51	(2.92)	0.28	(1.61)	0.06	(0.22)	0.37	(1.84)	0.36	(1.84)
v6	0.53	(2.72)	0.46	(2.51)	0.52	(2.70)	0.46	(1.58)	0.58	(2.65)	0.57	(2.69)
v7	0.49	(2.35)	0.42	(2.24)	0.57	(2.92)	0.14	(0.49)	0.60	(2.69)	0.59	(2.77)

Table 8 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
v8	0.53	(2.60)	0.44	(2.44)	0.53	(2.79)	0.02	(0.09)	0.45	(2.10)	0.47	(2.24)
v9	0.55	(2.52)	0.43	(2.38)	0.51	(2.70)	0.76	(2.63)	0.47	(2.20)	0.48	(2.30)
v10	1.08	(3.25)	0.89	(3.02)	1.01	(3.24)	0.99	(2.07)	1.02	(2.89)	1.02	(3.03)
Mean alpha	0.53	-	0.49	-	0.51	-	0.33	-	0.54	-	0.54	-
GRS test-stat	1.62	-	1.71	-	1.32	-	1.81	-	1.04	-	1.45	-
p value	0.03	-	0.20	-	0.01	-	0.16	-	0.00	-	0.10	-
Mean Adj R ²	5.57	-	6.74	-	6.10	-	6.24	-	6.03	-	5.05	-
Mean S.E	0.19	-	0.17	-	0.18	-	0.27	-	0.20	-	0.20	-
Hansen-Jagannathan distance												
δ	0.024	-	0.015	-	0.017	-	0.033	-	0.024	-	0.010	-

This table reports the regression output for 10 portfolios formed on book to market. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. v1 represents the portfolio of firms with lowest book to market while v10 stands for portfolio of firms with highest book to market value. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

Table 9 Regression output for 10 size-portfolios

Variable	CAPM	3F	4F	4F BSV	4F RSI	4F_CSI	(t-stat)
Panel A: Equally-weighted size portfolio							
s1	1.70	1.54	1.50	0.84	1.85	1.80	(6.99)
s2	1.20	0.98	1.08	0.82	1.32	1.28	(6.94)
s3	0.93	0.73	0.77	0.68	0.88	0.86	(5.12)
s4	0.93	0.73	0.82	0.94	0.85	0.84	(5.57)
s5	0.77	0.57	0.67	0.36	0.69	0.69	(5.11)
s6	0.83	0.64	0.71	0.52	0.69	0.69	(5.54)
s7	0.53	0.35	0.49	0.32	0.42	0.44	(3.73)
s8	0.43	0.29	0.38	0.26	0.39	0.39	(3.35)
s9	0.42	0.32	0.48	0.12	0.46	0.46	(4.27)
s10	0.37	0.35	0.46	0.08	0.46	0.47	(5.20)
Mean alpha	0.81	0.66	0.74	0.50	0.80	0.80	-
GRS test-stat	7.09	9.64	11.78	2.58	10.95	11.48	-
p value	0.00	0.00	0.00	0.10	0.00	0.00	-
Mean Adj R ²	5.51	5.12	5.46	5.67	5.63	5.04	-
Mean S.E	0.21	0.13	0.13	0.20	0.15	0.15	-
Hansen-Jagannathan distance							
δ	0.023	0.011	0.025	0.031	0.022	0.008	-
Panel B: Value-weighted size portfolio							
s1	0.81	0.66	0.64	0.02	0.91	0.87	(3.84)
s2	1.11	0.92	0.93	0.62	1.17	1.13	(6.58)
s3	0.94	0.76	0.71	0.58	0.83	0.81	(5.03)
s4	0.97	0.81	0.68	0.75	0.87	0.84	(5.43)
s5	0.84	0.67	0.63	0.30	0.69	0.68	(5.06)
s6	0.84	0.68	0.62	0.33	0.67	0.66	(5.79)
s7	0.62	0.46	0.46	0.28	0.47	0.47	(4.39)

Table 9 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F BSV	(t-stat)	4F RSI	(t-stat)	4F_CSI	(t-stat)
s8	0.49	(2.77)	0.36	(3.75)	0.29	(2.92)	0.22	(1.44)	0.34	(3.01)	0.33	(3.04)
s9	0.50	(3.94)	0.42	(4.48)	0.44	(4.50)	0.10	(0.72)	0.47	(4.22)	0.46	(4.33)
s10	0.32	(4.52)	0.37	(7.25)	0.37	(6.84)	0.28	(3.44)	0.41	(6.72)	0.41	(6.84)
Mean alpha	0.75	-	0.61	-	0.58	-	0.35	-	0.69	-	0.67	-
GRS test-stat	8.70	-	9.93	-	8.80	-	2.45	-	8.96	-	9.23	-
p val	0.00	-	0.00	-	0.00	-	0.11	-	0.00	-	0.00	-
Mean Adj R ²	5.54	-	5.42	-	5.47	-	5.65	-	5.15	-	5.08	-
Mean S.E	0.20	-	0.12	-	0.12	-	0.19	-	0.14	-	0.14	-
Hansen-Jaganathan distance												
δ	0.021	-	0.034	-	0.015	-	0.031	-	0.017	-	0.011	-

This table reports the regression output for 10 portfolios formed on size. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. S1 represents the portfolio of firms with lowest size while s10 stands for portfolio of firms with highest size value. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

Table 10 Regression output for 6 size and book to market portfolios using NYSE composite index

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Panel A: Equally-weighted size and BTM portfolio												
sl	0.46	(1.90)	0.35	(2.73)	0.48	(3.63)	0.33	(1.63)	0.58	(3.88)	0.56	(3.94)
sm	0.77	(3.93)	0.59	(6.88)	0.71	(8.04)	0.59	(4.41)	0.71	(7.07)	0.71	(7.31)
sh	1.08	(4.99)	0.85	(8.52)	0.92	(8.81)	0.66	(4.15)	1.03	(8.71)	1.02	(8.87)
bl	0.32	(2.11)	0.32	(3.00)	0.46	(4.29)	0.21	(1.26)	0.44	(3.56)	0.44	(3.73)
bm	0.41	(3.72)	0.34	(3.61)	0.40	(4.05)	0.12	(0.83)	0.38	(3.40)	0.39	(3.57)
bh	0.52	(2.79)	0.36	(2.81)	0.59	(4.62)	0.09	(0.49)	0.55	(3.79)	0.56	(3.97)
Mean alpha	0.60	-	0.47	-	0.60	-	0.34	-	0.62	-	0.62	-
GRS test-stat	5.61	-	15.07	-	20.87	-	4.35	-	17.51	-	18.42	-
<i>p</i> value	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
Mean Adj <i>R</i> ²	6.30	-	6.61	-	6.29	-	6.16	-	6.09	-	6.01	-
Mean S.E	0.18	-	0.11	-	0.11	-	0.17	-	0.13	-	0.12	-
Hansen-Jagannathan distance												
δ	0.034	-	0.012	-	0.047	-	0.021	-	0.018	-	0.009	-
Panel B: Value-weighted size and BTM portfolio												
sl	0.36	(1.53)	0.29	(2.76)	0.35	(3.21)	0.35	(2.09)	0.33	(2.65)	0.34	(2.78)
sm	0.63	(3.58)	0.47	(6.22)	0.49	(6.22)	0.38	(3.15)	0.46	(5.19)	0.47	(5.42)
sh	0.68	(3.53)	0.47	(6.67)	0.46	(6.33)	0.13	(1.21)	0.54	(6.58)	0.53	(6.65)
bl	0.42	(3.56)	0.52	(6.41)	0.55	(6.55)	0.18	(1.43)	0.58	(6.00)	0.57	(6.18)
bm	0.34	(2.94)	0.36	(3.26)	0.31	(2.67)	0.28	(1.58)	0.40	(3.02)	0.38	(3.02)
bh	0.39	(2.57)	0.34	(3.27)	0.44	(4.08)	0.40	(2.41)	0.36	(2.96)	0.38	(3.17)
Mean alpha	0.47	-	0.41	-	0.44	-	0.29	-	0.45	-	0.45	-
GRS test-stat	14.98	-	14.57	-	14.97	-	2.66	-	12.48	-	13.14	-
<i>p</i> value	0.00	-	0.00	-	0.00	-	0.002	-	0.00	-	0.00	-
Mean Adj <i>R</i> ²	5.24	-	5.94	-	5.05	-	5.18	-	5.06	-	5.00	-
Mean S.E	0.17	-	0.09	-	0.10	-	0.15	-	0.11	-	0.11	-

Table 10 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Hansen–Jagannathan distance												
δ	0.022	–	0.018	–	0.013	–	0.011	–	0.032	–	0.021	–

This table reports the regression output for 6 portfolios formed on size and book to market. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the NYSE composite index. To interpret the variables in both panels, s represents small, l stands for low, m represents medium, h stands for high and b represents big. For instance, in panel A, sl is interpreted as the portfolio with the small firm size and low book to market. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

model. This is equally the same for the H–J error. The output for the 10way size portfolio in Table shows almost identical pattern of results but with only 4F BSV passing the GRS test for both the equally and value weighted portfolios. It produced a higher p value indicating its alpha estimates are jointly distinguishable from zero. Summarily, the regression outputs suggest that the inclusion of sentiment measures as risk factors, significantly improve the extant models.

7 Robustness: NYSE composite index

The previous estimations were based on portfolios formed within the UK All Share Index. To further test the accuracy of our model, we formed portfolios using the NYSE composite index. Essentially, we use six portfolios formed based on size and BTM. Table 10 conveys the ability of the six asset pricing models to describe the cross-sectional excess returns of these portfolios. For both the equally and value-weighted portfolios, the findings depict similar patterns of previous outputs. It is shown that sentiment variable marginally increases the mean alphas and gradually reduces adjusted R^2 of the six separate models. Despite the significant impact of the sentiment variable, these outcomes become deflated as all the models failed to pass the GRS test because they produced significant p values (Table 11, 12).

8 Robustness: alternative sentiment proxies

We further consider alternative sentiment proxies and create a new set of raw and clean sentiment indices for model estimation. We use turnover ratio (calculated as trading volume divided by market value) (Baker & Wurgler 2007) and investors intelligence index (Brown & Cliff 2005). Repeating the same methodological approach as shown above and testing the model on 6 portfolios formed based on size and momentum within the UK All Share Index, our findings in table 10 show that the sentiment variable gradually magnifies the mean alpha. Overall, the estimation results reveal that the sentiment measures perform better than the traditional factor models. Additionally, we use VIX as an alternative sentiment measure and apply two stage least squares regression. Our findings show that the results do not differ significantly from the main estimation, suggesting that the sentiment-induced models perform better.

9 Conclusion and policy implications

In recent times, studies have revealed linkage between the activities of noise traders with stock price disruption. The motivation stems from the evidence that sentiments which permeate stock markets reflect investors' illusion about market information. Moreover, ample evidence from the use of experimental psychology in finance (behavioural finance) show that human biases and cognition affect the behaviour of stock prices and essentially, could trigger prolonged market anomalies. More so, the absence of single definitive sentiment measure has increased research interest in employing various tools for predictive purposes. Motivated by these growing efforts and the fact that traditional pricing methods do not aptly account for erratic stock price movements, we contribute to the current literature by

Table 11 Regression Output for 6 Size and Momentum using alternative sentiment proxies

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Panel A: Equally-weighted size and momentum portfolio												
sl	0.45	(1.59)	0.19	(1.15)	0.67	(4.58)	0.42	(1.87)	0.77	(4.60)	0.75	(4.68)
sm	0.82	(5.12)	0.66	(9.38)	0.65	(8.75)	0.43	(3.82)	0.72	(8.62)	0.71	(8.78)
sh	1.51	(7.48)	1.39	(12.70)	1.04	(11.52)	0.76	(5.57)	1.07	(10.36)	1.06	(10.82)
bl	0.00	(-0.33)	-0.20	(-1.31)	(3.76)	0.12	(0.70)	0.47	(3.57)	0.46	(3.69)	
bm	0.56	(5.18)	0.50	(5.48)	0.49	(5.12)	0.14	(0.99)	0.48	(4.49)	0.49	(4.72)
bh	0.87	(4.72)	0.88	(5.92)	0.33	(3.05)	0.02	(0.14)	0.36	(2.93)	0.36	(3.00)
Mean alpha	0.69	-	0.57	-	0.61	-	0.32	-	0.65	-	0.64	-
GRS test-stat	12.80	-	38.99	-	31.59	-	6.36	-	26.61	-	28.21	-
p value	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
Mean Adj R ²	5.53	-	5.90	-	5.51	-	5.52	-	5.31	-	5.40	-
Mean S.E	0.19	-	0.13	-	0.11	-	0.16	-	0.12	-	0.12	-
Hansen-Jagannathan Distance												
δ	0.021	-	0.015	-	0.041	-	0.031	-	0.025	-	0.032	-
Panel B: Value-weighted size and momentum portfolio												
sl	0.00	(-0.28)	-0.30	(-1.98)	0.34	(3.46)	0.09	(0.63)	0.38	(3.44)	0.38	(3.50)
sm	0.60	(3.67)	0.44	(5.91)	0.44	(5.68)	0.25	(2.11)	0.46	(5.22)	0.46	(5.38)
sh	1.29	(6.17)	1.19	(10.18)	0.75	(8.85)	0.60	(4.61)	0.74	(7.62)	0.74	(7.93)
bl	-0.10	(-0.69)	-0.20	(-1.09)	0.58	(4.95)	0.45	(2.51)	0.62	(4.61)	0.61	(4.72)
bm	0.54	(5.68)	0.56	(6.04)	0.50	(5.13)	0.50	(3.36)	0.46	(4.22)	0.47	(4.43)
bh	0.70	(3.93)	0.77	(4.88)	0.16	(1.51)	0.00	(-0.31)	0.26	(2.09)	0.25	(2.04)
Mean alpha	0.49	-	0.41	-	0.47	-	0.31	-	0.49	-	0.49	-
GRS test-stat	15.40	-	28.37	-	21.90	-	5.93	-	16.95	-	18.06	-
p value	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
Mean Adj R ²	6.38	-	7.87	-	7.83	-	7.87	-	8.82	-	6.04	-
Mean S.E	0.19	-	0.13	-	0.10	-	0.15	-	0.11	-	0.11	-

Table 11 (continued)

Variable	CAPM	(t-stat)	3F	(t-stat)	4F	(t-stat)	4F_BSV	(t-stat)	4F_RSI	(t-stat)	4F_CSI	(t-stat)
Hansen–Jagannathan distance												
δ	0.013	–	0.054	–	0.033	–	0.024	–	0.002	–	0.035	–

This table reports the regression output for 6 portfolios formed on size and momentum. Data covers the period from January 1993 to December 2020. Equally and value weighted portfolios are formed using the firms in the UK stock market. To interpret the variables in both panels, s represents small, l stands for low, m represents medium, h stands for high and b represents big. For instance, in panel A, sl is interpreted as the portfolio with the small firm size and low momentum. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI)

T-statistics for the coefficients of the models are shown in parenthesis

Table 12 Robustness: 2SLS regression output for 6 size and momentum using VIX as an alternative sentiment proxy

Variable	CAPM	3F	(t-stat)	4F	(t-stat)	4F BSV	(t-stat)	4F RSI	(t-stat)	4F CSI	(t-stat)	4F VIX	(t-stat)	
First stage regression														
sl	0.23*	(0.10)	0.19	(0.05)	0.45**	(0.11)	0.26*	(0.17)	0.33*	(0.02)	0.58	(1.22)	0.33***	(0.16)
sm	0.36**	(0.01)	0.44*	(1.02)	0.22*	(1.50)	0.33*	(0.16)	0.29**	(1.35)	0.10*	(0.09)	0.08*	(0.11)
sh	0.24**	(0.11)	0.27**	(1.08)	0.34	(1.21)	0.64*	(0.07)	0.72*	(0.16)	0.56**	(0.20)	0.34*	(0.06)
bl	0.50*	(1.88)	0.49*	(0.80)	1.56*	(0.43)	0.61	(0.41)	0.79*	(0.93)	0.71	(0.22)	0.33**	(0.10)
bm	0.81*	(0.13)	0.58*	(0.23)	0.25**	(0.34)	0.67	(0.11)	0.34*	(0.13)	0.44**	(0.19)	0.58	(1.22)
bh	0.55*	(0.09)	0.47**	(0.36)	0.17*	(0.22)	0.65**	(2.48)	0.83*	(0.24)	0.75	(0.21)	0.56**	(0.20)
Mean alpha	0.16	-	0.27	-	0.31	-	0.26	-	0.15	-	0.19	-	0.21	-
Adj R ²	0.33	-	0.39	-	0.41	-	0.32	-	0.48	-	0.42	-	0.41	-
Second stage regression														
sl	0.29*	(0.15)	0.11**	(0.08)	0.35*	(0.06)	0.29	(0.10)	0.43**	(0.06)	0.31	(0.11)	0.23*	(0.09)
sm	0.26*	(0.11)	0.30**	(0.12)	0.32*	(0.09)	0.26**	(0.11)	0.28	(0.05)	0.30*	(0.12)	0.33**	(0.01)
sh	0.30*	(0.11)	0.22*	(0.12)	0.27	(0.13)	0.44*	(0.02)	0.32**	(0.16)	0.29	(0.08)	0.22	(0.09)
bl	0.29*	(0.16)	0.36	(0.10)	0.56*	(0.01)	0.40	(0.12)	0.39**	(0.13)	0.31	(0.12)	0.37*	(0.07)
bm	0.31**	(0.10)	0.38	(0.03)	0.45*	(0.04)	0.37	(0.06)	0.41***	(0.10)	0.24*	(0.05)	0.38*	(0.22)
bh	0.35**	(0.05)	0.37*	(0.16)	0.29	(0.02)	0.35**	(0.48)	0.43***	(0.14)	0.25	(0.01)	0.26	(0.09)
Mean alpha	0.11	-	0.20	-	0.15	-	0.10	-	0.25	-	0.12	-	0.09	-
Adj R ²	0.22	-	0.38	-	0.29	-	0.42	-	0.34	-	0.29	-	0.31	-

This table reports the two stage least squares (2SLS) regression output for 6 portfolios formed on size and momentum. Data covers the period from January 1993 to December 2020. The models tested are the capital asset pricing model (CAPM), Fama & French (1993) three-factor model (3F), Carhart (1997) four-factor model (4F), four-factor model plus the basic sentiment variable (4F-BSV), four-factor model plus the raw sentiment index (4F-RSI) and four-factor model plus the clean sentiment index (4F-CSI). T-statistics for the coefficients of the models are shown in parenthesis

conceptualising what noise trading might represent, and thereafter investigate the dynamics of noise trading and investors' sentiments in asset pricing. In estimating the predictive ability of sentiment constructs, as an alternative or complimentary risk factor, first, we document the impact of extant asset pricing models, such as capital asset pricing model (CAPM), Fama-french 3-factor model and Carhart 4-factor model, on the excess returns of a variety of portfolios formed in different patterns and extracted from firms within the UK. Secondly, the study extends its focus to incorporate sentiment measures into the erstwhile models with a view to discerning the influence of such measures to increase or decrease excess returns. The major concern was to generate intercept and coefficient for each portfolio, with the aim of examining if the intercept terms for each model are jointly significantly different from zero.

To generate convincing estimates, the study conducted correlation analysis of variables to verify commonality of components. Hence, instead of using only the basic sentiment model we constructed, the study also includes, in different model specifications, raw and clean sentiment indices derived from a principal component analysis (PCA) of variables. The first PCA explains about 79% of the overall variance for all variables while the second PCA provides explanation to about 67.5% variation. Further to the inclusion of the iterated sentiment variable in different pricing models, the findings indicate that sentiment-augmented estimates provide statistical significance in explaining the excess returns of majority of the portfolios, in fact, outperforming the traditional asset pricing models. Our results are further corroborated with the use of Hansen & Jagannathan (1997) non-parametric test which substantiates the inclusion of sentiment measures in our asset pricing models. More importantly, our results are consistent and conform with the positions of past studies on investor sentiments, such as Baker & Wurgler (2007), Joseph et al. (2011) and Bathia & Bredin (2018). Hence, our basic sentiment measure, the ratio of laggards to leaders, communicates strong market (or price) signal to noise traders. The ratio can therefore serve as a valid proxy for investor sentiment, which in turn can be used to predict excess or abnormal returns.

Our findings provide significant indications for future academic and practical attempts at asset pricing. The study shows that sentiment is a vital systematic risk factor that must not be ignored by analysts, regulators and other stakeholders when analysing the return characteristics of any asset or portfolio. We show that four factor model is inadequate and needs a sentiment factor to improve asset pricing. The study therefore concludes that extant asset pricing models may not be sufficient to capture market or price anomalies as the activities of noise traders, exemplified by sentiments, have significant implications on the behaviour of stock markets.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmadi F (2016) The relationship between inflation and stock index in the last ten years, in Tehran stock exchange. *World Sci News* 40:34–57
- Baker M, Wurgler J (2006) Investor sentiment and the cross-section of stock returns. *J Financ* 61(4):1645–1680
- Baker M, Wurgler J (2007) Investor sentiment in the stock market. *J Econ Perspect* 21(2):129–152
- Bathia D, Bredin D (2018) Investor sentiment: does it augment the performance of asset pricing models? *Int Rev Financ Anal* 59:290–303
- Bia-likowski J, Etebari A, Wisniewski TP (2012) Fast profits: Investor sentiment and stock returns during Ramadan. *J Bank Financ* 36(3):835–845
- Black F (1986) Noise. *J Financ* 41(3):528–543
- Brown GW, Cliff MT (2004) Investor sentiment and the near-term stock market. *J Empir Financ* 11(1):1–27
- Brown GW, Cliff MT (2005) Investor sentiment and asset valuation. *J Bus* 78(2):405–440
- Cakici N, Fabozzi FJ, Tan S (2013) Size, value, and momentum in emerging market stock returns. *Emerg Mark Rev* 16:46–65
- Carhart MM (1997) On persistence in mutual fund performance. *J Financ* 52(1):57–82
- Chang SC, Chen SS, Chou RK, Lin YH (2012) Local sports sentiment and returns of locally headquartered stocks: a firm-level analysis. *J Empir Financ* 19(3):309–318
- Changsheng H, Yongfeng W (2012) Investor sentiment and assets valuation. *Syst Eng Proc* 3:166–171
- Chen SS (2011) Lack of consumer confidence and stock returns. *J Empir Financ* 18(2):225–236
- Chen SS (2012) Revisiting the empirical linkages between stock returns and trading volume. *J Bank Financ* 36(6):1781–1788
- Chen AS, Fang SC (2009) Uniform testing and portfolio strategies for single and multifactor asset pricing models in the Pacific Basin markets. *Appl Econ* 41(15):1951–1963
- Chen J, Sherif M (2016) Illiquidity premium and expected stock returns in the UK: a new approach. *Phys A* 458:52–66
- Da Z, Engelberg J, Gao P (2014) The sum of all FEARS: investor sentiment and asset prices. *Rev Financ Stud* 28(1):1–32
- Dalika NK (2014) Sentiment and returns: analysis of investor sentiment in the South African market. *Invest Manag Financ Innov* 12(2):267–276
- Das PK, Freybote J, Marcato G (2015) An investigation into sentiment-induced institutional trading behavior and asset pricing in the reit market. *J Real Estate Financ Econ* 51(2):160–189
- De Long JB, Shleifer A, Summers LH, Waldmann RJ (1989) The size and incidence of the losses from noise trading. *J Financ* 44(3):681–696
- De Long JB, Shleifer A, Summers LH, Waldmann RJ (1991) The survival of noise traders in financial markets. *J Bus* 34(5):1–19
- Dosumu OE, Sakariyahu R, Oyekola O, Lawal R (2023) Panic bank runs, global market contagion and the financial consequences of social media. *Econ Lett* 26:111170
- Fama EF (1998) Market efficiency, long-term returns, and behavioural finance. *J Financ Econ* 49(3):283–306
- Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J Financ Econ* 33(1):3–56
- Fama EF, French KR (1995) Size and book-to-market factors in earnings and returns. *J Financ* 50(1):131–155
- Fama EF, French KR (1996) Multifactor explanations of asset pricing anomalies. *J Financ* 51(1):55–84
- Finter P, Niessen-Ruenzi A, Runzi S (2012) The impact of investor sentiment on the German stock market. *Z Betrieb* 82(2):133–163
- Gagnon L, Karolyi GA (2010) Multi-market trading and arbitrage. *J Financ Econ* 97(1):53–80
- Gibbons MR (1982) Multivariate tests of financial models: a new approach. *J Financ Econ* 10(1):3–27
- Gibbons MR, Ross SA, Shanken J (1989) A test of the efficiency of a given portfolio. *Econometrica* 57:1121–1152
- Goetzmann WN, Kim D, Kumar A, Wang Q (2014) Weather-induced mood, institutional investors, and stock returns. *Rev Financ Stud* 28(1):73–111
- Gregory A, Tharyan R, Christidis A (2013) Constructing and testing alternative versions of the Fama-French and Carhart models in the UK. *J Bus Financ Acc* 40(1–2):172–214
- Griffin JM (2002) Are the Fama and French factors global or country specific? *Rev Financ Stud* 15(3):783–803
- Hanauer MX, Linhart M (2015) Size, value, and momentum in emerging market stock returns: integrated or segmented pricing? *Asia Pac J Financ Stud* 44(2):175–214

- Hansen LP, Jagannathan R (1997) Assessing specification errors in stochastic discount factor models. *J Financ* 52(2):557–590
- He J, Kan R, Ng L, Zhang C (1996) Tests of the relations among market wide factors, firm-specific variables, and stock returns using a conditional asset pricing model. *J Financ* 51(5):1891–1908
- He Z, He L, Wen F (2019) Risk compensation and market returns: The role of investor sentiment in the stock market. *Emerg Mark Financ Trade* 55(3):704–718
- Ho C, Hung C-H (2009) Investor sentiment as conditioning information in asset pricing. *J Bank Financ* 33(5):892–903
- Humpe A, Macmillan P (2009) Can macroeconomic variables explain long-term stock market movements? a comparison of the US and Japan. *Appl Financ Econ* 19(2):111–119
- Hyde S, Sherif M (2010) Tests of the conditional asset pricing model: further evidence from the cross-section of stock returns. *Int J Financ Econ* 15(2):198–211
- Joseph K, Wintoki MB, Zhang Z (2011) Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *Int J Forecast* 27(4):1116–1127
- Kahneman D and Tversky A (2013) Choices, values, and frames. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, pp. 269–278.
- Kasman S, Vardar G (2011) The impact of interest rate and exchange rate volatility on banks' stock returns and volatility: evidence from Turkey. *Econ Modell* 28(3):1328–1334
- Kim D, Park K, Lee DJ, Ahn Y (2020) Predicting mobile trading system discontinuance: the role of attention. *Electron Commer Res Appl* 44:101008
- Kumar A, Lee CM (2006) Retail investor sentiment and return c0-movements. *J Financ* 61(5):2451–2486
- Lee M, Gan C (2006) Macroeconomic variables and stock market interactions: New Zealand evidence. *Invest Manag Financ Innov* 3(4):89–101
- Lee BS, Rui OM (2002) The dynamic relationship between stock returns and trading volume: domestic and cross-country evidence. *J Bank Financ* 26(1):51–78
- Lee CM, Shleifer A, Thaler RH (1991) Investor sentiment and the closed-end fund puzzle. *J Financ* 46(1):75–109
- Lemmon M, Portniaguina E (2006) Consumer confidence and asset prices: some empirical evidence. *Rev Financ Stud* 19(4):1499–1529
- Levy T, Yagil J (2011) Air pollution and stock returns in the US. *J Econ Psychol* 32(3):374–383
- Liang H, Yang C, Zhang R, Cai C (2017) Bounded rationality, anchoring-and- adjustment sentiment, and asset pricing. *North Am J Econ Financ* 40:85–102
- Ling DC, Naranjo A, and Scheick B (2010) Investor sentiment and asset pricing in public and private markets. *RERI WP*, 170.
- Lintner J (1965) Security prices, risk, and maximal gains from diversification. *J Financ* 20(4):587–615
- Markowitz H (1952) Portfolio selection. *J Financ* 7(1):77–91
- Mossin J (1966) Equilibrium in a capital asset market. *Econ J Econ Soc* 56:768–783
- Oliveira N, Cortez P and Areal N (2013) On the predictability of stock market behaviour using stocktwits sentiment and posting volume. In *Portuguese Conference on Artificial Intelligence*. Springer, Singapore, pp. 355–365.
- Pan J, Poteshman AM (2006) The information in option volume for future stock prices. *Rev Financ Stud* 19(3):871–908
- Patel S (2012) The effect of macroeconomic determinants on the performance of the Indian stock market. *NMIMS Manag Rev* 22(1):117–127
- Paterson A, Sakariyahu R, Lawal R, Alabi A (2023) The Impact of government policy responses to the COVID-19 pandemic and brexit on the UK financial market: a behavioural perspective. *Brit J Manag*. <https://doi.org/10.1111/1467-8551.12702>
- Petkova R (2006) Do the Fama-French factors proxy for innovations in predictive variables? *J Financ* 61(2):581–612
- Ramiah V, Xu X, Moosa IA (2015) Neoclassical finance, behavioural finance, and noise traders: a review and assessment of the literature. *Int Rev Financ Anal* 41:89–100
- Rao, T. & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis. In *Proceedings of the 2012 international conference on advances in social networks analysis and mining (ASONAM 2012)* (pp. 119–123): IEEE Computer Society.
- Ross SA (1976) Arbitrage theory of capital asset pricing. *J Econ Theory* 13:341–360
- Sakariyahu R, Sherif M, Paterson A, Chatzivgeri E (2021) Sentiment-apt investors and UK sector returns. *Int J Financ Econ* 26(3):3321–3351
- Sakariyahu R, Lawal R, Oyekola O, Dosumu OE, Adigun R (2023) Natural disasters, investor sentiments and stock market reactions: evidence from Turkey-Syria earthquakes. *Econ Lett* 228:111153

- Salhin A, Sherif M, Jones E (2016) Managerial sentiment, consumer confidence and sector returns. *Int Rev Financ Anal* 47:24–38
- Scruggs JT (2007) Noise trader risk: evidence from the Siamese twins. *J Financ Markets* 10(1):76–105
- Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. *J Financ* 19(3):425–442
- Shleifer A, Summers LH (1990) The noise trader approach to finance. *J Econ Perspect* 4(2):19–33
- Shleifer A, Vishny RW (1997) The limits of arbitrage. *J Financ* 52(1):35–55
- Simoes Vieira E (2011) Investor sentiment and the market reaction to dividend news: European evidence. *Manag Financ* 37(12):1213–1245
- Simoes Vieira E (2011) Investor sentiment and the market reaction to dividend news: European evidence. *Manag Financ* 37(12):1213–1245
- Tetlock PC (2007) Giving content to investor sentiment: the role of media in the stock market. *J Financ* 62(3):1139–1168
- Treynor JL (1961) Market value, time, and risk. *Time, and Risk* (August 8, 1961).
- Treynor JL (1962) Toward a theory of market value of risky assets. *Available at SSRN 628187*.
- Trichilli Y, Abbes MB, Masmoudi A (2020) Predicting the effect of Googling investor sentiment on Islamic stock market returns. *Int J Islam Middle East Financ Manag* 13(2):165–193
- Trueman B (1988) A theory of noise trading in securities markets. *J Financ* 43(1):83–95
- Uhl MW (2014) Reuters sentiment and stock returns. *J Behav Financ* 15(4):287–298
- Xu Y, Green CJ (2013) Asset pricing with investor sentiment: evidence from Chinese stock markets. *Manch Sch* 81(1):1–32
- Yang C, Zhang R (2013) Sentiment asset pricing model with consumption. *Econ Model* 30:462–467
- Yaya OS, Shittu OI (2010) On the impact of inflation and exchange rate on conditional stock market volatility: a re-assessment. *Am J Sci Indust Res* 1(2):115–117
- Zhang Y, Zhang Y, Shen D, Zhang W (2017) Investor sentiment and stock returns: evidence from provincial TV audience rating in China. *Phys A* 466:288–294
- Zhu Z, Ji Q, Sun L, Zhai P (2020) Oil price shocks, investor sentiment, and asset pricing anomalies in the oil and gas industry. *Int Rev Financ Anal* 70:101516
- Zouaoui M, Nouyrigat G, Beer F (2011) How does investor sentiment affect stock market crises? *Evid Panel Data Financ Rev* 46(4):723–747

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.