SENTIMENT IN ASSET PRICING: A META-ANALYSIS OF SENTIMENT INDICES

Louis Burgess
Under the supervision of Dr Xi Chen

A dissertation submitted for the degree of
Master of Science

September 2020

(Word Count: 6927)

Abstract

We question the dominant position of the Baker-Wurgler (BW) index as the mainstream indirect sentiment index, and measure its explanatory power on aggregate market returns. Using a vector autoregression model, we first find that the BW index does not explain returns well. We then find that with a few trivial alterations, the explanatory power of indirect sentiment indices can be greatly increased. Our results indicate that the careful selection of sentiment proxies is the most important consideration of a PCA-based index construction. We outline that more careful study of the principal components may lead to a better understanding of sentiment effect groupings. Following this we provide some evidence that more than one principal component may represent sentiment. Overall, given the evidence, we note that researchers and practitioners should not take the BW index for granted.

Contents

1 Introduction

The importance of sentiment in asset pricing cannot be overstated, yet the mechanisms by which it operates remain unclear. Furthermore, there is no universal consensus on the definition of sentiment (Aggarwal, 2019), without which one cannot utilise this concept consistently and robustly. Given the diverging opinions on sentiment, it is important to make obvious one's definition of sentiment. Throughout this report, sentiment is treated in line with the work of Baker and Wurgler (2006) as investors' speculative beliefs about asset prices and the optimism or pessimism they feel towards a particular market. This may not be the best description of sentiment but it at least serves as a way to describe the phenomena occurring in this research. Nevertheless, this report may aid in determining a more robust definition of sentiment. Despite initially being considered as an effect of irrationality in early research (see Barberis et al., 1998), literature tends to separate the concept of rationality and sentiment. Following this, we are hesitant to discuss any relationship between the two concepts; regardless, well extracted subjective views should be of great use in asset pricing. How these views are extracted however, is of the utmost importance. Generally, there are two ways to quantify sentiment: via financial and accounting measures forming a composite index (e.g. Baker and Wurgler, 2006; Huang et al., 2014), or directly via textual analysis (e.g. Mullen and Collier, 2004) or surveys (e.g. Shefrin, 2015). Here, the focus is on indirect sentiment measures, i.e. sentiment indices composed of various observable financial quantities. Therein lies the question – how does one best represent sentiment using these quantities?

Literature on sentiment is vast, and yet there are few indirect indices which aim to represent sentiment. In fact, the majority of literature utilises the Baker and Wurgler (2006) (BW) index as *the* measure of sentiment. While this index has proven a useful tool, the field – for the most part – has given up in seeking better and more refined ways to represent sentiment. Examples such as Huang et al. (2014) and Chen et al. (2019) have at least expanded on the methodology used by Baker and Wurgler (2006), however, Concetto and Ravazzolo (2019) find that out of the indirect sentiment indices found in the literature, the original BW index had the most predictive power with respect to US market returns,[1] and so will be the focus in the empirical tests. In any case, there are few other indirect sentiment indices, and certainly none that are in common use. There seems to be an assumption that the BW is the pinnacle of sentiment representation, or maybe that it is not worth finding something better. Additionally, many authors impetuously use the BW index as sentiment without considering other possibilities. With more and more authors using the BW index without question, the problem compounds.

To partially remedy this issue, this report seeks to further examine the BW index and other individual proxies, forming a new index which will be compared against the BW index. In doing so, the extent of the BW index's renowned merit shall be tested. This is not to say that the BW index is particularly flawed or problematic, simply that further investigation into sentiment indices is warranted.

When forming a representation of an unobservable quantity, one must be able to test its efficacy in order to assure that it is at least somewhat replicating the desired quantity. For example, the implied

[1] A common metric in sentiment literature.

volatility of options given by the Black and Scholes (1973) model can be compared to realised market volatility to test the assumptions of the model. One of the quantitative checks taken to ensure that the BW index measures sentiment is its capacity to explain returns.[2] This measure must be used cautiously else one risks the chance of labelling a non-sentiment effect as sentiment. For instance, the book-to-market ratio is a good predictor of returns (see Fama and French, 1993), yet is not thought to be a sentiment proxy.[3]

A similar problem arises from the method of sentiment extraction. The basis for the BW index construction is a principal component analysis (PCA). The idea behind this is that if one selects several imperfect proxies which are theorised to contain a sentiment component, then the first principal component should track the systematic commonality between the proxies, i.e. sentiment. The issue here is twofold. One must be sure that each of the proxies does, in fact, contain a substantial sentiment component, such that the PCA extracts the common sentiment component. If the proxies do not contain substantial sentiment components, then the first principal component will not be related to sentiment, regardless of whether or not there is a component which does relate to sentiment. Furthermore, it is difficult to know if the principal component extracted does actually represent sentiment, or some other commonality between proxies. Seeking the explanatory power with respect to returns does not necessarily indicate whether or not the component represents sentiment, simply whether or not the commonality is an import factor in explaining returns.

If one follows the logic of Baker and Wurgler (2006), and concludes that their method of representing and measuring sentiment is logical and correct, one can see that there is still room for further investigation. For instance, why stop at six proxies? Why use the specific proxies that have been chosen rather than others identified by the literature? There is a whole host of alterations that could be made to the BW index, and yet little work has been done in this regard. Looking at individual proxies in the literature may give insight as to how to improve the BW index.

[2] Qualitatively, Baker and Wurgler (2006) give an anecdotal explanation of market events in line with their index and find that it aligns with their interpretation of sentiment.

[3] Hahn et al. (2010) find that the book-to-market ratio component which correlates with sentiment has far less explanatory power on returns than the component which correlates with systematic risk. However, this study uses the BW index as sentiment, further showing that the BW index is taken for granted.

2 Sentiment indices: A literature review

The earliest literature on sentiment in finance can be traced back to Barberis et al. (1998) who define it as the over- or under-reaction of the investors when forming their expectations, *i.e.* a form of investor irrationality.[4] More recent literature seems to move away from this definition and instead, highlight the subjectivity of the investors' opinion when defining sentiment. For instance, Baker and Wurgler (2006) define sentiment as the propensity for investors to speculate based on subjective asset valuations, and the optimism, or pessimism, investors feel towards a certain market. Shefrin (2008) defines sentiment as the investors' subjective beliefs in contrast to fundamental values. Glaser et al. (2009) set their definition of sentiment as the asset price expectation based on subjective principals. Zhou (2018) define sentiment as how far an asset value deviates from its economic fundamentals.

A universal definition of sentiment is still missing (see Aggarwal, 2019), but the literature seems to gradually settle around Baker and Wurgler (2006)'s view of sentiment and method of constructing a sentiment index: they select the closed-end fund discount, dividend premium, number of IPOs, first day IPO returns, equity share in new issues, and turnover ratio as the sentiment proxies, and extract a sentiment factor through a principal component analysis (PCA) on these measures. Several studies utilise this sentiment measure and apply it to their research. For instance, Yu and Yuan (2011) find that expected excess market return is positively related to the market's conditional volatility in periods of low sentiment; Stambaugh et al. (2012) find that that high sentiment periods often lead to overpricing due to existing short-sale impediments in the market; and Yang and Zhang (2014) find that sentiment affects the market equilibrium of stocks. There are many more studies which use the BW index as a sentiment measure (e.g. Berger and Turtle, 2012; Chang et al., 2019; Asness et al., 2020), as it is by far the most used in the literature, with Baker and Wurgler (2006) having amassed more than 4,500 citations according to Google Scholar.

In contrast to such a vast volume of applications, limited effort is put on further developing Baker and Wurgler (2006)'s construction of the sentiment index. From the literature reviewed, only Huang et al. (2014) and Chen et al. (2019) have proposed alternative indirect sentiment indices: for instance, Huang et al. (2014) uses the same proxies as Baker and Wurgler (2006) but uses partial least-squares to maximise cross-sectional returns explanation. However, Concetto and Ravazzolo (2019) find that out of the indices featured in Huang et al. (2014), and Baker and Wurgler (2006), the original BW index held the most explanatory power with respect to market returns, and so will be the focus in our empirical tests.[5]

At this point, the sentiment literature has presented obvious gaps on its journey to construct a "comprehensive sentiment index" using financial measures. Many researchers prior to Baker and Wurgler (2006) have identified a variety of sentiment-related measures. However, the majority of these have neither been compared to, nor included in Baker and Wurgler (2006)'s index despite many proxies having been re-

[4]Perhaps the most famous piece of work on investor irrationality is the prospect theory (by Kahneman and Tversky, 1979) which identifies the investors' increasing tendency to take risks when losing more, and argues it is evidence for investor irrationality.

[5]Chen et al. (2019) uses proxies unique to the internet finance industry to form an industry specific sentiment index, and so would not be an appropriate measure to explain aggregate market returns.

viewed in prior work by Brown and Cliff (2004). See Table 1 for all the existing sentiment proxies and sentiment-related measures in the literature.

Table 1: Sentiment proxies identified from the literature. Key papers are outlined in which the proxies are mentioned or explained as sentiment proxies. The categorisation of these proxies are our creation based on the work of Brown and Cliff (2004). Note that we assign our own abbreviations to the proxies.

Category	**Proxy**	**Key Literature**
Investor outlook	IPO first day returns ($RIPO$)	Baker and Wurgler (2006, 2007)
	Dividend premium ($DIVP$)	Baker and Wurgler (2004, 2006, 2007)
	Change in margin borrowing ($\Delta DEBT$)	Brown and Cliff (2004)
	Change in short interest ($\Delta SINT$)	Brown and Cliff (2004)
	Ratio of specialists' short sales to total short sales (SS_S/SS_T)	Brown and Cliff (2004)
	Ratio of odd-lot sales to purchases ($ODLT_R$)	Brown and Cliff (2004)
	Position in futures by trader type (FUT)	Wang (2003); Brown and Cliff (2004)
	Proportion of fund assets held in cash ($CASH$)	Brown and Cliff (2004)
	CBOE put/call ratio (PCR)	Simon and Wiggins (2001); Brown and Cliff (2004); Bandopadhyaya and Jones (2011)
	Buy-sell imbalance ($IMBA$)	Kumar and Lee (2006)
	Closed-end fund discount ($CEFD$)	Neal and Wheatley (1998); Baker and Wurgler (2006, 2007)
	Mutual fund redemptions ($FUND$)	Neal and Wheatley (1998)
	Odd lot sales ($ODLT_S$)	Neal and Wheatley (1998); Brown and Cliff (2004)
Current market performance	Volatility (VIX)[6]	Simon and Wiggins (2001); Brown and Cliff (2004); Bandopadhyaya and Jones (2011); Baker and Wurgler (2007)
	Trading index ($TRIN$)	Simon and Wiggins (2001); Brown and Cliff (2004)
	Advances to declines ratio (AD)	Brown and Cliff (2004)
Firm financing decisions	Equity share in new issues ($ESNI$)	Baker and Wurgler (2000, 2006, 2007)
	Number of IPOs ($NIPO$)	Baker and Wurgler (2006, 2007)
Liquidity	Trading volume (VOL)	Baker and Stein (2004); Baker and Wurgler (2007)
	Turnover ratio ($TURN$)	Baker and Stein (2004); Baker and Wurgler (2006)

Table 1 also presents our categorisation of the sentiment-related measures, based on early review work done by Brown and Cliff (2004). Where categorisation is not clear, or not present, we categorise proxies based on underlying mechanisms found in the literature. This categorisation is important because it lays the basis for a deeper understanding about *what sentiment really is*. Sections (2.1 $\sim$ 2.4) provide further explanations of the relevance of these measures provided by the literature.

Overall, we find four groups of sentiment-related measures:

(a) Investor outlook, *i.e.* the effects of optimism and pessimism for future expected returns, and an investor's beliefs of the performance of the future market in general, rational or otherwise (in line with the definition used by Baker and Wurgler, 2006).

(b) Current performance of the market, under the assumption that it may be a driver of sentiment (see Brown and Cliff, 2004).

(c) Firm financing decisions, under the assumption that the firm would always carefully gauge the enthusiasm of prospective investors to ensure appropriate timing and size of their financing activities

[6] We use the CBOE Volatility Index throughout this report, but other volatility indices can be used.

(see Baker and Wurgler, 2007).

(d) (Excess) liquidity, on the understanding that it reflects a period of positive sentiment, as it is created when the market has a substantial proportion of irrational investors who trade on noise (see Baker and Stein, 2004).

The grouping and categorisation of sentiment-related measures also lead us to hypothesise that sentiment may be extracted as more than one factor in a PCA on all these measures, opposing Baker and Wurgler (2006)'s approach of only using the first principal component to construct a sentiment index. So far, however, we have not found any literature that attempt to study the true meaning of factors, although some perform a PCA on sentiment proxies following Baker and Wurgler (2006)'s approach. While this is not the focus of this project, we believe this warrants future research.

2.1 Investor outlook

The change in margin borrowing is the change in the total debt of all margin accounts. A greater amount of borrowing represents optimism that investing will provide superior returns compared to the money market (Brown and Cliff, 2004). Since leverage will multiply any profits or losses, investors must be confident about the performance of stock market if they are to borrow money to invest (Brigham and Houston, 2012). It is likely that there are components of both institutional and individual investor sentiment embedded in this proxy, since both types of investor have the opportunity to utilise leveraged investment.

The change in short interest indicates the change in the number of open short positions in the market. This proxy reflects the overall pessimism in the market (Brown and Cliff, 2004) – the more short interest, the more participants believe prices will fall.

The net position in futures by trader type echoes the mechanism of short interest, as it is the difference between open and short interest on futures contracts (Wang, 2003). This indicator is separated depending on the type of trader – institutional hedger, institutional speculator or individual trader. Wang (2003) find that while the individual component has no forecasting power, institutional hedgers are weak contrarian indicators, and institutional speculators show evidence of being price continuation indicators for the SPX futures market.

The ratio of specialists' short sales to total short sales is a value which represents expert opinions in the market. Under the assumption that specialists are more informed investors, when their short selling becomes proportionately large, the market is likely to decline (Brown and Cliff, 2004). When this indicator is high, identified specialists have a much more pessimistic outlook than the rest of the market.[7]

The put-call ratio as a sentiment proxy represents pessimism in the market (Brown and Cliff, 2004). The more speculative puts there are, compared to calls, the more derivative traders believe that there will be a market decline in the future. This measure has been used to explain variations in asset prices both in (e.g. Billingsley and Chance, 1988), and out (e.g. Bandopadhyaya and Jones, 2011) of a sentiment narrative.

[7] Brown and Cliff (2004) make no mention of how specialists are identified, therefore we choose not to include this indicator in our analyses.

The proportion of fund assets held in cash is a proxy which bears a similar intuition to the change in short interest – the more cash a fund holds the more pessimistic the fund manager is with respect to future investment returns (Brown and Cliff, 2004). Being dominated by the opinion of fund managers, this sentiment proxy is of an institutional nature. Unfortunately this proxy is not as simple as initially suggested, fund managers may keep large amounts of cash for varying reasons; if a manager believes there will be high quality investment opportunities in the near future, it is safe to assume that they may hold some cash aside to invest, especially if other holdings are particularly illiquid (Simutin, 2013).

Redemptions of mutual funds signal that investors are pessimistic about future expected returns (Neal and Wheatley, 1998). Mutual funds provide pooled investment opportunity for small and individual investors (Lemke et al., 1995), so greater redemption volume of mutual fund shares represents greater pessimism, since investors are unloading their exposure to the market. This proxy therefore relates to individual investor sentiment.

IPO first day returns give an insight into market hype surrounding a firm's initial public offering. Immediate returns on IPOs reflect the market's enthusiasm for newly public companies (Baker and Wurgler, 2006), with a low return often regarded as a sign of poor market timing (see Ritter, 1991). With both individual and institutional investors participating in the purchase of newly issued stock on the secondary market, this proxy likely represents both levels of investor sentiment.

Dividend premium is defined as the average difference between book-to-market ratios of dividend paying and non-paying firms (Baker and Wurgler, 2006). This essentially outlines the extra amount market participants are willing to pay for dividend income. Since dividend payers are usually established large-cap companies, they are seen as stable and safe, but will have fewer growth opportunities (Karpavičius and Yu, 2018). Increased demand for dividends indicate that investors are seeking fewer growth opportunities, possibly because they believe the market is unlikely to grow.

Since closed-end fund shares are not generally redeemable with the issuer, they are traded on the secondary market (Lemke et al., 1995). This means that the shares are priced by the market and so the price is not necessarily equal to the net asset value (NAV) of the fund share, thus there is a discount or premium attached (Lee et al., 1991). Generally the market price is less than the NAV, which suggests a pessimistic outlook of market participants, who believe that shares are performing worse than usual; the larger the closed-end fund discount, the more pessimistic the outlook. Lee et al. (1991) also state that closed-end fund shares tend to be held by individuals, such that this proxy reflects individual investor sentiment.

Odd-lots are defined as trades of less than 100 shares. They depict the trades of small investors and give an insight into their behaviour. Since small investors are generally regarded as less informed (Han and Chung, 2013), odd-lot movements are often taken as contrarian trading signals. The ratio of odd-lot sales to purchases is therefore a pessimistic individual sentiment proxy (Brown and Cliff, 2004).

Buy-sell imbalance is constructed by taking a ratio of the difference between volume inflow and outflow and the total volume of a stock or index (Kumar and Lee, 2006). According to Kumar and Lee (2006), this proxy relates to the time-varying preferences of retail investors, and they label this 'sentiment'. Thus, the buy-sell imbalance is treated as an individual investor sentiment proxy.

2.2 Current market performance

The trading index (TRIN) reflects the volume weighted number of shares with increasing value, against shares with decreasing value (Brown and Cliff, 2004), essentially measuring the proportion of the market that is going up in value. It is therefore an optimistic sentiment proxy. The non-volume weighted value can also be used and is known as the advance-decline (AD) ratio (Brown and Cliff, 2004). This has a similar interpretation to that of the TRIN. High levels of the TRIN often indicate market bottoms, as the combination of a low number of advancing issues with low volume and high number of declining issues with high volume shows that sellers may have finished selling (Simon and Wiggins, 2001).

The ratio of new highs to new lows is designed to capture the relative strength of the current market, and does so in a similar way to the TRIN (Brown and Cliff, 2004). Simply put, this indicator reports the number of new highs in the market to the number of new lows, per unit time. It is an optimistic sentiment proxy and, as pointed out by Huddart et al. (2009), attracts a lot of investor attention since stock highs and lows are widely reported in business publications. In fact, while abnormal returns can be achieved for up to six trading days after a new low, excess turnover persists for up to two weeks (Mizrach and Weerts, 2009).

Measuring the implied volatility of option prices is a common method of extracting the 'perceived' variation of share prices, and is usually calculated using the Black and Scholes (1973) model. During periods of turmoil, the implied (and realised) volatility are usually much higher than in tranquil periods (Bae et al., 2007). This is thought to be due to the asymmetric volatility response, noted often as the leverage effect (Black, 1976), seen where share price decrease induces a higher future volatility increase than that of a share price increase. Given this asymmetric increase in volatility when prices decline, implied volatility is seen as a pessimistic sentiment proxy – the VIX is often noted as the "fear index".

2.3 Firm financing decisions

The decisions firms make to raise capital are complex and may give an indication as to their sentiment. One factor in the decision process is of course investor demand – why go public if the demand for shares in your firm is low? A firm must carefully gauge the enthusiasm of prospective investors to ensure appropriate IPO timing (Brau and Fawcett, 2006). Therefore, the number of IPOs reflect the outlook of investors to some degree, and this is likely to be an optimistic sentiment proxy.[8] Baker and Wurgler (2007) note that the demand for IPOs is said to be very sensitive to investor sentiment.

Again, the decision a firm makes in what method of financing it undertakes may give an insight into the firm management's mindset and how they gauge investor enthusiasm. Baker and Wurgler (2000) find that a high ratio of equity issues to all total new issues signals low stock market returns, making the measure a pessimistic sentiment indicator. However, they also state that the exact mechanism under which this proxy operates is yet unknown.

[8] The accuracy of this may be dictated by how firm managers' perceive, or measure, investor enthusiasm.

2.4 Liquidity

Whilst the rationality of investors is unassumed for the previous sentiment proxies, liquidity effects here are that which are explained by irrational investor behaviour. There are several measures of liquidity; regardless of which measure one selects, the theory is that there is a significant group of irrational investors who underreact to market information and will therefore make uninformed trades (according to Baker and Stein, 2004), and these trades will be more frequent, thus boosting liquidity. Baker and Stein (2004) also suggest that when liquidity is unusually high, irrational traders have positive sentiment. However, Gebka and Wohar (2013) argue that higher volume does not imply high levels of noise trading. They argue that their empirical evidence is in line with the model by Campbell et al. (1993) in which both informed and uninformed trading causes increases in volume but the aggregate returns on that day are related to how informed the investors are. Because there is no clear consensus on the use of liquidity as a sentiment index, we elect to utilise it cautiously.

To measure liquidity as a sentiment proxy one can use quantities such as trading volume or turnover ratio, however, data from Jeffery Wurgler's (of Baker and Wurgler, 2006, 2007) website indicates that the turnover ratio is no longer used in their sentiment indicator because of extreme levels of high-frequency trading boosting liquidity greatly and introducing trading patterns which are harder to interpret, especially in a sentiment context.

3 Constructing sentiment indices

3.1 Data and proxy generation

Margin debt data was sourced from Thompson Reuters Eikon (monthly, 1960–2017); short interest from Wharton Research Data Services (daily, 2005–2020); CBOE exchange put-call ratio and VIX from CBOE (monthly, 2010–2019 and 2004–2019, respectively); dividend premium, closed-end fund discount, first-day IPO returns, number of IPOs, and equity share in new issues from Jeffery Wurgler's website (monthly, 1960–2018); and odd-lot transaction ratio from the SEC (daily, 2012–2020). Suitable datasets for the ratio of specialists' short sales, ratio of fund assets held in cash, odd-lot sales, mutual fund redemptions, futures positions by trader type, and buy-sell imbalance could not be obtained nor generated.

A list of S&P 500 constituents was obtained from a Wikipedia article to serve as a representative of US market movements.[9] Daily data from 2005 to 2020 on the adjusted close and volume of the index constituents (as well as the index as a whole using the `^GSPC` ticker) was obtained from Yahoo Finance, using the `yfinance` Python module. For the volume monthly time series, the sum of volume for all stocks downloaded for each month was taken. The S&P 500 returns time series was calculated by taking log returns on the index value time series. Short interest data was converted to monthly data by summing values per month. Generating a monthly time series for AD ratio and TRIN was performed using the methodology of Brown and Cliff (2004). The two quantities are defined as

$$AD = \frac{\text{Advancing issues}}{\text{Declining issues}}, \tag{1}$$

$$TRIN = \frac{\text{Advancing issues / Advancing volume}}{\text{Declining issues / Declining volume}}. \tag{2}$$

The number and volume of advancing and declining issues are calculated daily, and summed for each month. The quantities described by (1) and (2) can then be calculated on a monthly basis. Finally, the new highs to new lows proxy was calculated by taking the number of daily index constituents which surpass their 250 day highs (and lows), and summing them per month. The ratio between the number of highs and lows is then taken, and the a logarithm is taken to attenuate extreme values i.e.

$$NHNL = \ln\left(\frac{\#250 \text{ day highs in month}}{\#250 \text{ day lows in month}}\right). \tag{3}$$

3.2 Index construction

In this section, we first review the construction of the Baker and Wurgler (2006) sentiment index. We then construct our new comprehensive sentiment index to run against the BW index. Finally, we construct a short-term BW index variant, to test whether the period in which the index is constructed plays an important role in its explanatory power.

[9] The accuracy of this list is not of great importance; the subsequent companies serve as a sample of large-cap firms: https://en.wikipedia.org/wiki/List_of_S%26P_500_companies.

Benchmark: Baker and Wurgler (2006) sentiment index (BW)

Baker and Wurgler (2006) originally construct their index using $CEFD$, $DIVP$, $RIPO$, $NIPO$, $ESNI$, and $TURN$. They then perform a PCA on this data (and subsequent one month lags) over the period 1962–2001, and retrieve the first principal component as a preliminary sentiment index. The lead or lag for each variable which correlates less with the principal component is then discarded, leaving an index with six variables. This index is then scaled to have unit variance and produces the following loadings

$$\begin{aligned} BW_t = &-0.241\,CEFD_t + 0.242\,TURN_{t-1} + 0.253\,NIPO_t \\ &+ 0.257\,RIPO_{t-1} + 0.112\,ESNI_t - 0.283\,DIVP_{t-1}. \end{aligned} \tag{4}$$

In the sentiment data that is updated by Jeffery Wurgler, the $TURN$ factor is removed (see Section 2.4 for explanation), and the new index loadings are not made clear. However, the final index data is provided.

A comprehensive sentiment index (CSI)

We repeat the same procedure as in Baker and Wurgler (2006) but instead using the twelve sentiment proxies for which we have sufficient data, and over the period 2006–2018. We first standardise our data for the PCA, and then the index is constructed, however, there was a significant number of missing data points in 2008 for the $RIPO$ data, and so the mean $RIPO$ value was inserted. Next, the sign on the loadings was reversed to make the new index positively correlated with the BW index (Pearson coefficient = 0.62). Finally, the index was scaled to have unit variance, yielding

$$\begin{aligned} CSI_t = &\ 0.074\,RIPO_t + 0.229\,NIPO_t - 0.311\,CEFD_t - 0.414\,ESNI_t \\ &+ 0.082\,\Delta VOL_t - 0.333\,DIVP_{t-1} - 0.038\,TRIN_{t-1} + 0.060\,\Delta SINT_{t-1} \\ &- 0.390\,VIX_{t-1} + 0.132\,\Delta DEBT_{t-1} + 0.083\,NHNL_{t-1} + 0.038\,AD_{t-1}, \end{aligned} \tag{5}$$

where the variables are defined as in Table 1. One stark contrast to the BW index, seen in Figure 1, is how much more the CSI reacts to the 2007-2008 crisis. The reason for this higher level of reactivity is so far unclear, but suggests that the CSI may be more useful, as it incorporates information faster than the BW index.

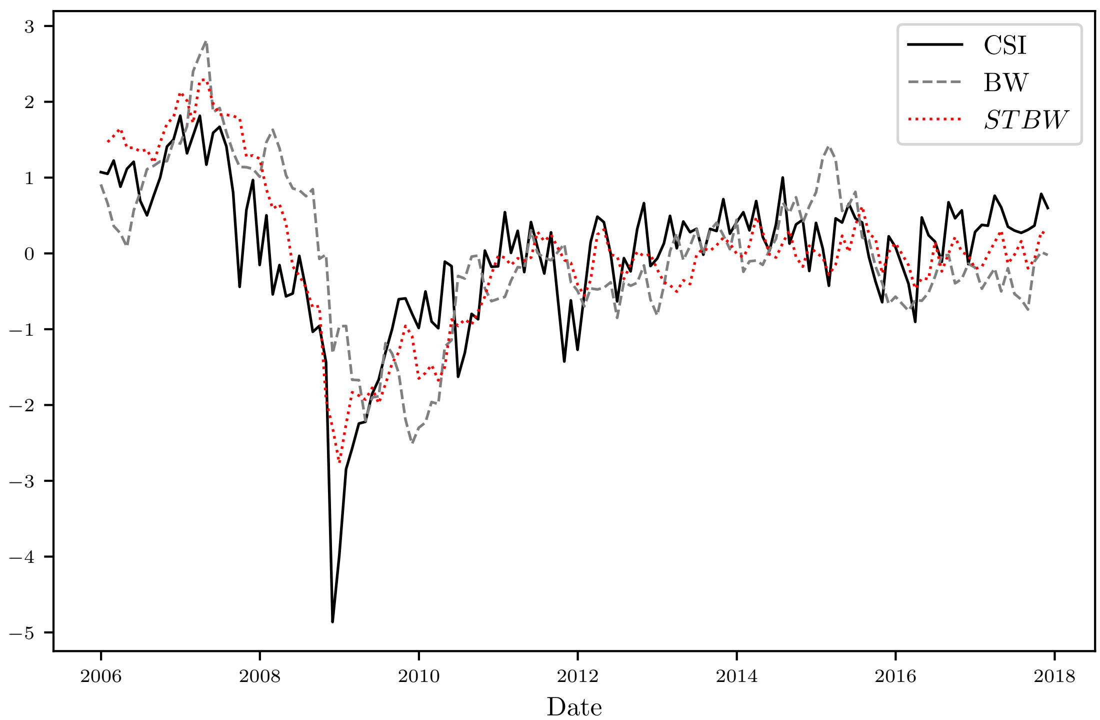
The short-term BW index variant ($STBW$)

To construct the short-term BW index, the same process as in Baker and Wurgler (2006) is performed, but in this case the index is constructed using data from the period 2006–2018. As above, the resulting index is scaled to have unit variance:

$$\begin{aligned} STBW_t = &-0.634\,CEFD_{t-1} + 0.289\,NIPO_{t-1} \\ &+ 0.124\,RIPO_t - 0.675\,ESNI_t - 0.630\,DIVP_{t-1}. \end{aligned} \tag{6}$$

Interestingly, the sign on the equity share of new issues is reversed when compared to the original BW index, and its loading is comparatively higher, possibly indicating a higher importance of this sentiment proxy during the more recent period. Figure 1 shows that this new short-term index is very similar to the original BW index, with one interesting difference; the short-term BW index tends to lead the original BW index, i.e. it is more reactive to changes in sentiment.

Figure 1: Plot of BW index, CSI, and short-term BW index for the period 2006–2018. All three indices have been centred and have unit variance for the sake of a clear comparison. The two indices which are constructed using more recent data tend to lead the original BW index.



4 Modelling

4.1 Baseline model

Baker and Wurgler (2006) use the fact that their index can account for cross-sectional stock returns to partly justify that their index represents sentiment. If their index has explanatory power with respect to individual stocks, then it should be able to explain aggregate market returns (represented by the S&P 500 index returns). We therefore select the explanatory power on returns as the metric for comparing indices. In doing so, the superior proxy or index will be the one which is better able to explain market returns. To measure the extent to which sentiment can be used to do this, we employ a vector autoregression (VAR) model with a similar specification to that seen in Kelly and Ahmad (2018),[10] specified as

$$\begin{aligned} r_t = \beta_0 + \sum_{i=1}^{N} \left(\sum_{\tau=0}^{M} \beta_{i,t-\tau} p_{i,t-\tau} \right) + \sum_{\tau=0}^{M} \beta_{N+1,t-\tau} BW_{t-\tau} + \sum_{\tau=0}^{M} \beta_{N+2,t-\tau} CSI_{t-\tau} \\ + \sum_{i=N+3}^{C} \left(\sum_{\tau=0}^{M} \beta_{i,t-\tau} c_{i,t-\tau} \right) + \epsilon_t, \end{aligned} \tag{7}$$

where r_t is the aggregate market return at time t, all $\beta_{n,m}$ are the linear coefficients, where n is the coefficient index, and m is the corresponding time of the variable that the coefficient multiplies, $p_{i,t}$ are individual sentiment proxies, and all $c_{i,t}$ are macroeconomic control variables.

To make sure that the proxies and indices only account for sentiment components in the regression, several financial and economic control variables are added to the regression model, along with their lags. Some readily available variables were included in data provided by Jeffery Wurgler. These quantities are not chosen for any particular reason, they are simply useful factors to capture external non-sentiment effects and include an industrial production index (indpro), nominal durables consumption (consdur), nominal non-durables consumption (consnon), nominal services consumption (consserv), the consumer price index (cpi), a US employment index (employ), and the National Bureau of Economic Research recession indicator (recess). In addition, four of the five factors from the Fama and French (2015) model are taken as control variables, to capture changes in market risk effects.[11] The variables here are not necessarily carefully selected or perfect, but they do function as a wide selection of controls to account for external effects; the Fama and French (1993) three-factor model variables have at least been shown to be good forecasters of the aggregate market by Panopoulou and Plastira (2014). The variables included are high minus low (HML), small minus big (SMB), conservative minus aggressive (CMA), and robust minus weak (RMW), see Fama and French (2015) for a detailed description of each risk factor.

To avoid biasing the regression results, it is useful to make some of the explanatory variables stationary. To do this in all cases, the first difference is taken, i.e. $\Delta X = X_t - X_{t-1}$. These stationarity alterations not only increase the goodness of fit in this case, but reduce the probability of spurious regressions. To

[10] In contrast to Kelly and Ahmad (2018)'s model, ours is not predictive.

[11] The market returns factor is obviously excluded from the controls, given that this is the dependant variable in our VAR model.

test for stationarity, an augmented Dickey and Fuller (1979) unit root test is performed at the 5% level, and non-stationary variables were made stationary, apart from the BW index.

4.2 Parsimonious model

Because the model specified in (7) has many independent variables, it will almost certainly suffer from over-fitting. Therefore when out-of-sample tests are performed, the model will perform very badly. Instead, for out-of-sample tests, we utilise the more parsimonious model specified as

$$r_t = \beta_0 + \beta_{1,t}\,SMB_t + \beta_{2,t}\,HML_t + \beta_{3,t}\,CMA_t + \beta_{4,t}\,RMW_t + \sum_{t=0}^{3} \beta_{5,t-\tau} I_{t-\tau}, \tag{8}$$

where I is the index to be tested, the second to fifth variables are the Fama and French (2015) factors, and all $\beta_{n,m}$ are the linear coefficients, where n is the coefficient index, and m is the corresponding time of the variable that the coefficient multiplies. This model can then be used with both the CSI and BW index in place of $I_{t-\tau}$.

4.3 Model validation

When performing a regression it is important to ensure that the coefficients are the best linear unbiased estimators, otherwise the results are meaningless. In all regressions performed, several tests were undertaken to ensure that the regression results were in accordance with the Gauss-Markov theorem, and thus robust. One criterion of this is that the error terms have constant variance. To check for this, a White (1980) test is performed. The null hypothesis of this test is that the variance of the error terms is constant. If the p-value of the test statistic is above the chosen 5% significance level, the null hypothesis of homoskedasticity can not be rejected. To make certain that the estimators are unbiased, the residuals must be checked to ensure their mean value is zero. This can be done using a Z-test, with the null hypothesis being that the mean is equal to zero.

Under the Gauss-Markov assumptions, the residuals should be normally distributed. To test for this, a Jarque and Bera (1980) test is performed for each regression, which employs a null hypothesis of normality, i.e. a skewness of zero and a kurtosis of three. The final important robustness check is to confirm that the residuals do not suffer from autocorrelation. In this case, a Durbin and Watson (1950) test is utilised. If the test statistic is sufficiently close to two, then the residuals display no evidence of serial correlation.

5 Empirical results

Several tentative regressions were performed using a selection of variables. We initially regress all twelve sentiment proxies as well as the control variables on the S&P 500 returns from 2006–2018. We remove highly correlated variables and ones which reduce the adjusted goodness of fit, as the aim is to reduce the number of variables while maximising the explained variance of returns.[12] We then choose to include the cross-section of each variable as well as the next three lags.[13] This ensures that the VAR model accounts for short- to medium-term effects.

Throughout this section, the necessary robustness checks are performed to ensure that parameter estimations are valid (see Section 4.3 for a description of these tests). For each VAR model tested, the regression passes each robustness check, unless stated otherwise.

5.1 Comparative regressions vs. baseline model

Three models are tested, the baseline model, the same model with the addition of the BW index, and the baseline model with the addition of the short-term BW index. These regressions show whether the BW index or short term variant have significant explanatory power over the regression period. The regression results for the baseline model are reported in Table 2, results for the baseline model including the BW index in Table 3, and results for the baseline model including the short-term BW index in Table 4.

Table 2: Vector autoregression results ($N = 137$), for the period 2006–2018, of the baseline model. The table shows the first 15 coefficients of variables (upper), sorted by smallest p-value, and the goodness of fit summary (lower).

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
consdur_{t-1}	0.0184	0.0049	3.7672	0.0003
$\Delta SINT_{t-3}$	0.0123	0.0035	3.5344	0.0007
consdur_{t-2}	0.0178	0.0051	3.4901	0.0008
employ_{t-3}	0.0240	0.0070	3.4443	0.0009
indpro_{t}	0.0113	0.0047	2.3919	0.0191
employ_{t}	-0.0184	0.0080	-2.3159	0.0231
indpro_{t-2}	-0.0109	0.0048	-2.2889	0.0247
$TRIN_{t-2}$	0.0086	0.0039	2.2310	0.0285
CMA_{t}	-0.5796	0.2650	-2.1877	0.0316
consserv_{t-3}	-0.0072	0.0036	-2.0018	0.0487
const	0.0050	0.0026	1.9400	0.0559
SMB_{t}	0.3647	0.1915	1.9043	0.0605
cpi_{t-3}	0.0077	0.0045	1.7346	0.0867
AD_{t}	-0.0075	0.0047	-1.6106	0.1112
employ_{t-1}	-0.0122	0.0078	-1.5513	0.1248
$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$	**AIC**	**F-stat**	**Prob. (F-stat)**
0.710	0.506	-540.9	3.491	1.75×10^{-7}

The results shown in Table 2 show that the baseline model fits the data well, with a high R^2 value. Out of the sentiment proxies, $\Delta SINT_{t-3}$ and $TRIN_{t-2}$ have significant coefficients, indicating that they can

[12] Highly correlated according to our threshold dictated by the R^2 of the model.

[13] Lags for the Fama and French (2015) factors were not included in the regression, as this is not common practice and did not increase the adjusted R^2.

explain a portion of the returns variance. As expected, many of the baseline model control variables have significant explanatory power with respect to returns. What is slightly unexpected is that only the CMA_t Fama and French (2015) factor is significant out of the four factors included. Given that Panopoulou and Plastira (2014) show that SMB and HML are good forecasters of returns, we would expect to see corresponding significant coefficients.

Table 3: Vector autoregression results ($N = 137$), for the period 2006–2018, of the baseline model with the BW index included. The table shows the first 15 coefficients of variables (upper), sorted by smallest p-value, and the goodness of fit summary. (lower).

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
employ_{t-3}	0.0261	0.0072	3.6304	0.0005
$\Delta SINT_{t-3}$	0.0117	0.0036	3.2973	0.0015
consdur_{t-1}	0.0156	0.0051	3.0814	0.0029
consdur_{t-2}	0.0153	0.0053	2.9188	0.0046
CMA_t	-0.6858	0.2717	-2.5245	0.0137
indpro_t	0.0116	0.0050	2.3313	0.0224
employ_t	-0.0176	0.0083	-2.1328	0.0362
indpro_{t-2}	-0.0103	0.0049	-2.1092	0.0382
$TRIN_{t-2}$	0.0081	0.0039	2.0712	0.0417
const	0.0052	0.0026	2.0055	0.0485
$NHNL_{t-3}$	-0.0165	0.0085	-1.9471	0.0552
cpi_{t-3}	0.0082	0.0045	1.8124	0.0739
consserv_{t-3}	-0.0064	0.0037	-1.7584	0.0827
SMB_t	0.3358	0.2000	1.6785	0.0973
ΔVOL_{t-3}	0.0072	0.0045	1.6047	0.1127
$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$	**AIC**	**F-stat**	**Prob. (F-stat)**
0.724	0.507	-539.9	3.327	5.18×10^{-7}

The results shown in Table 3 are perhaps more interesting. When the BW index is included in the model specification, the adjusted R^2 increases and yet the BW index does not have a statistically significant coefficient. This means that the BW index does not sufficiently explain market returns, despite its wide adoption in sentiment analysis. The significant regression coefficients remain very similar to the previous regression, indicating that when the model is perturbed with the BW index there is little change, and so the regression estimators are reasonably stable.

The results shown in Table 4 summarise a regression of the baseline model with the addition of the short-term BW index. As the short-term BW index does not have a significant coefficient, the results suggest that constructing an index using more relevant data is not particularly important in determining whether an index can sufficiently explain market returns. However, there is a marginal increase in the goodness of fit compared to the model with the original BW index. When the short-term BW index is included in the model, the variables with significant coefficients change slightly, such that $NHNL_{t-3}$ and AD_t, which did not previously have significant coefficients, are now significant. This was not a drastic change however, as both $NHNL_{t-3}$ and AD_t had relatively low p-values in the baseline regression. This baseline model is obviously over-fitted due to having so many independent variables, however this is not a major issue as this model will not be used for out-of-sample tests. Instead, the parameters and their significance represent the explanatory power of each variable with respect to market returns.

Table 4: Vector autoregression results ($N = 136$), for the period 2006–2018, of the baseline model with the short-term BW index included. The table shows the first 15 coefficients of variables (upper), sorted by smallest p-value, and the goodness of fit summary (lower).

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
consdur_{t-1}	0.0191	0.0051	3.7415	0.0004
$\Delta SINT_{t-3}$	0.0127	0.0035	3.5848	0.0006
employ_{t-3}	0.0250	0.0073	3.4186	0.0010
consdur_{t-2}	0.0165	0.0053	3.1150	0.0026
CMA_t	-0.6495	0.2837	-2.2893	0.0249
employ_t	-0.0175	0.0081	-2.1571	0.0342
$NHNL_{t-3}$	-0.018	0.0087	-2.0668	0.0422
$TRIN_{t-2}$	0.0081	0.0039	2.0653	0.0424
AD_t	-0.0102	0.0050	-2.0462	0.0442
indpro_t	0.0098	0.0049	2.0017	0.0489
employ_{t-1}	-0.0166	0.0084	-1.9852	0.0508
indpro_{t-2}	-0.0095	0.0049	-1.9525	0.0546
const	0.0047	0.0026	1.8145	0.0736
SMB_t	0.3517	0.1981	1.7757	0.0798
consserv_{t-3}	-0.0063	0.0037	-1.7092	0.0916
$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$	**AIC**	**F-stat**	**Prob. (F-stat)**
0.727	0.509	-535.7	3.335	5.47×10^{-7}

5.2 BW index vs. CSI

Providing evidence that the BW index does not explain returns well is not particularly helpful without providing at least a rudimentary solution. In this case we test, in a direct comparison, whether using more proxies to construct an index is this very solution. To do this we estimate parameters for the baseline model with the addition of both the BW index and CSI. Table 5 reports these results. Ignoring the control variables, the only significant sentiment proxies or indices are $\Delta SINT_{t-3}$, CSI_t, and ΔVOL_{t-1}. Our results therefore show that the CSI is a significant improvement on the BW index, and could be a contender for a new mainstream sentiment indicator.

As one might expect, the other two proxies which have significant coefficients in this regression are ones which only make up a very small component of the CSI, i.e. they have small loadings. As previously mentioned, the relationship between trading volume and sentiment is somewhat controversial; either trading volume does not proxy sentiment (and explains returns well regardless), or it does and the first principal component of the CSI proxies is not the only sentiment component. Because short interest almost certainly contains a component of sentiment, and does load strongly onto the CSI, we believe that the latter argument, i.e. that more than one principal component may represent sentiment, is likely.

5.3 Short-term BW vs. CSI

While the evidence so far has suggested that including more proxies is the reason for the increase in performance of the CSI relative to the BW index, there is another possibility; the CSI index is formed using information over the same period as the regression. This means that the loadings of the index are based on much more relevant information to the market period in question. It is possible that the power

Table 5: Vector autoregression results ($N = 136$) for the period 2006–2018, including both the BW index and the CSI. The table shows the first 15 coefficients of variables (upper), sorted by smallest p-value, and the goodness of fit summary (lower).

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
employ_{t-3}	0.0265	0.0071	3.7197	0.0004
consdur_{t-1}	0.0182	0.0051	3.5428	0.0007
$\Delta SINT_{t-3}$	0.0126	0.0037	3.3636	0.0012
indpro_t	0.0142	0.0051	2.7740	0.0070
consdur_{t-2}	0.0149	0.0054	2.7478	0.0076
SMB_t	0.5157	0.2171	2.3748	0.0202
CSI_t	0.0495	0.0211	2.3462	0.0217
CMA_t	-0.6299	0.2750	-2.2908	0.0249
indpro_{t-2}	-0.0116	0.0052	-2.2361	0.0284
const	0.0055	0.0026	2.1302	0.0366
ΔVOL_{t-1}	0.0127	0.0062	2.0534	0.0437
employ_t	-0.0166	0.0083	-2.0166	0.0475
ΔVOL_t	-0.0104	0.0053	-1.9743	0.0522
AD_{t-1}	-0.0096	0.0049	-1.9454	0.0556
$\Delta DEBT_{t-2}$	0.0117	0.0062	1.8659	0.0661
$\mathbf{R}^2$	**Adj. $\mathbf{R}^2$**	**AIC**	**F-stat**	**Prob. (F-stat)**
0.766	0.555	-548.2	3.626	1.26×10^{-7}

of the CSI stems not from its wider selection of sentiment proxies, but from the time period in which the loadings are calculated. Conversely, the BW index calculates loadings using the period 1962–2001. Using this longer time period has the benefit of providing a principal component which explains a large amount of the variance in the data. The principal component which forms the basis for the BW index explains about 50% of the variance in the underlying data, whereas the principal component for the CSI only explains around 18%. However, the BW index may contain much less relevant information, unless one views sentiment as some universal unchanging constant. Factor loadings are unlikely to be persistent over long periods, especially given that the entire turnover ratio factor was dropped from the BW index, following the rise of high-frequency trading.

To test if the supposed relevance of the index loadings is the main reason that the CSI outperforms the BW index in explanatory power, we compare the CSI and short-term BW index (as constructed in section 3.2) directly as additions to the baseline VAR model.

The regression results shown in Table 6 indicate that the increased information content of the CSI with respect to market returns is not solely due to the period in which the factor loadings are fit. The CSI coefficient remains significant, while the short-term BW index coefficient is not significant at the 5% level. The model fit does improve however, with a superior adjusted R^2 to that of the previous regression (seen in Section 5.2). This shows that there does seem to be an increase in the explanatory power of an index when it is fit in a more recent time period. The coefficient of the original BW index in the first VAR has a p-value of 0.738 compared to the short-term BW index coefficient in the second VAR which has a p-value of 0.054, further showing that there is a significant benefit in the short-term fitting process. However, given that the short-term BW index does not have a significant coefficient, the time period in which the index is constructed *is not the dominant consideration*. Instead, the main focus should be on

Table 6: Vector autoregression results ($N = 137$) for the period 2006–2018. The BW index is replaced in the regression with the short-term BW index variant. The table shows the first 15 coefficients of variables (upper), sorted by smallest p-value, and the goodness of fit summary (lower).

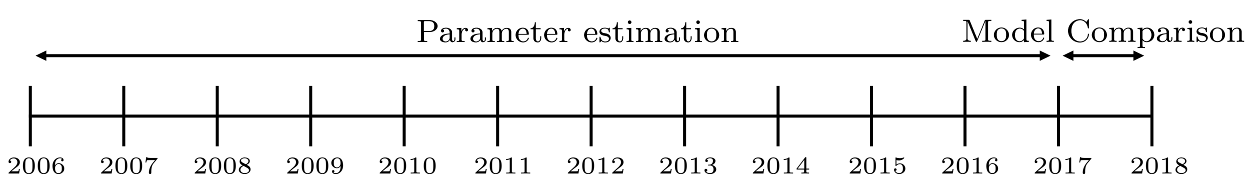
Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
$\Delta SINT_{t-3}$	0.0136	0.0037	3.6799	0.0005
consdur_{t-1}	0.0181	0.0050	3.5855	0.0006
employ_{t-3}	0.0221	0.0071	3.1162	0.0026
employ_{t}	-0.0229	0.0079	-2.8957	0.0050
indpro_{t}	0.0148	0.0051	2.8850	0.0052
CSI_t	0.0651	0.0228	2.8491	0.0057
consdur_{t-2}	0.0147	0.0052	2.8012	0.0066
ΔVOL_t	-0.0136	0.0053	-2.5581	0.0127
AD_{t-1}	-0.0117	0.0049	-2.4117	0.0185
CMA_t	-0.5773	0.2753	-2.0973	0.0395
indpro_{t-2}	-0.0104	0.0051	-2.0599	0.0431
SMB_t	0.4020	0.1992	2.0183	0.0473
$TRIN_{t-2}$	0.0085	0.0042	1.9981	0.0495
$\Delta DEBT_{t-1}$	-0.0124	0.0063	-1.9677	0.053
ΔVOL_{t-3}	0.0094	0.0049	1.9311	0.0575
$\mathbf{R}^2$	**Adj. $\mathbf{R}^2$**	**AIC**	**F-stat**	**Prob. (F-stat)**
0.777	0.550	-546.9	3.429	5.06×10^{-7}

which proxies are selected to be in the index; careful consideration should be taken to select proxies which do reflect sentiment, although this can be very challenging.

5.4 Out-of-sample analysis

Despite the in-sample tests conducted being fairly conclusive, out-of-sample tests must be conducted to ensure that the extra information content of the CSI is retained outside of the period in which regression parameters are estimated. For this test there is limited data available, and so only a one year out-of-sample comparison is made, as an extra check on the validity of the results of the previous tests. Because the original baseline model suffers from over-fitting, the more parsimonious model specified in (8) is used.

Figure 2: Timeline of the one year out-of-sample model comparison. The parameters are estimated using the data from 2006–2017, and the models are compared using the data from 2017.



To perform the out-of-sample analysis, the timeline shown in Figure 2 is utilised; the coefficients are estimated using data up to the final year, and the two regression models are then compared using data from the final year. Once the parameters have been estimated, the out-of-sample returns according to the model can be analysed. For this, the mean square error is calculated as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r_i})^2, \tag{9}$$

where n is the out-of-sample size, r is the true S&P 500 return, and $\hat{r}$ is the model return. A smaller MSE indicates a superior model.

The parameter estimation results are given in Table 7 for the CSI model, and Table 8 for the BW index model. The in-sample fit for the CSI index model is better than that of the BW index based model, as expected. Furthermore, none of the BW coefficients are statistically significant at the 5% level, while the CSI_{t-3} coefficient is. The out-of-sample MSE of the CSI model is 5.52×10^{-4}, while the out-of-sample MSE of the BW index model is 6.18×10^{-4}. This indicates that even in limited out-of-sample tests, the CSI still contains more explanatory power with respect to returns than the BW index.

Table 7: Vector autoregression results (N=125) of baseline model with CSI over the period 2006–2017, for estimating parameters for out-of-sample tests. Note that only the three period lag is significant out of all the CSI index lags.

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
CMA_t	-0.9841	0.2593	-3.7961	0.0002
RMW_t	-0.6716	0.2383	-2.8186	0.0057
SMB_t	0.4083	0.1707	2.392	0.0184
HML_t	0.4184	0.1909	2.192	0.0304
CSI_{t-3}	0.0151	0.0073	2.0816	0.0396
const	0.0055	0.0033	1.6556	0.1005
CSI_{t-3}	-0.0085	0.0064	-1.3377	0.1836
CSI_{t-2}	-0.0073	0.0071	-1.0301	0.3051
CSI_t	0.0025	0.0059	0.4297	0.6682
$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$	**AIC**	**F-stat**	**Prob. (F-stat)**
0.327	0.281	-464.7	7.050	1.47×10^{-7}

Table 8: Vector autoregression results (N=125) of baseline model with BW index over the period 2006–2017, for estimating parameters for out-of-sample tests. It is important to note that this regression did have non-normal residuals, but passed all other robustness checks.

Variable	**Coef.**	**Std. Err.**	t	$P > \|t\|$
CMA_t	-0.9632	0.2645	-3.6409	0.0004
HML_t	0.5352	0.1845	2.9011	0.0044
RMW_t	-0.6702	0.2426	-2.7625	0.0067
SMB_t	0.399	0.1691	2.3591	0.0200
const	0.0056	0.0034	1.6623	0.0992
BW_{t-3}	-0.0113	0.0103	-1.0994	0.2739
BW_t	0.0071	0.0103	0.6913	0.4908
BW_{t-1}	0.0047	0.014	0.3366	0.7370
BW_{t-2}	-0.0039	0.0139	-0.282	0.7784
$\mathbf{R^2}$	**Adj.** $\mathbf{R^2}$	**AIC**	**F-stat**	**Prob. (F-stat)**
0.307	0.259	-546.9	6.417	6.89×10^{-7}

6 Conclusion

This report has pursued an insight into the performance of mainstream sentiment indices, when using the explanatory power on aggregate market returns as a performance metric. The two main questions posed are: 1. Given its dominant position in sentiment research, does the Baker and Wurgler (2006) index require improvement? and 2. How can the BW index be improved? The answer to the first question is fairly simple; the BW index has served as the staple mainstream sentiment index in research since its inception in the seminal works of the 2000s. The consistent use of this index, without question of its performance or true meaning, has led to a compounding effect and now researchers have somewhat taken this index for granted. Given that under one of its own performance metrics the BW index underperforms, there is clear scope for improvement. Furthermore, there is no real understanding of what this index truly represents; the BW index is an abstract index labelled as sentiment.[14]

The answer to the second question is slightly more complex. We have shown that one of the ways to improve this type of sentiment index is to alter the selection of sentiment proxies used. To a lesser extent, constructing the index over a more recent period also leads to improvement. This naturally leads to the question of whether our improvements lead to an index that better represents sentiment. The answer, for now, is unclear; a 'ground truth' for sentiment is a controversial subject and so a proper way to benchmark indirect sentiment indices must be found. There is a lot of research on analysing investor sentiment using textual analysis (e.g. Wuthrich et al., 1998; Antweiler and Murray, 2004), yet many models formed in this type of research remain empirically untested in a financial context. One clear channel for future research is a direct comparison between these textual-based models and indirect sentiment indices.

Using the aforementioned improvements, we have constructed a new sentiment index, namely the comprehensive sentiment index, which serves as an example of the potential of sentiment indices when given a small number of changes. The improvements made to the BW index in this report are fairly trivial, and yet still produce an index which has significantly greater explanatory power. With some more careful variable selection, and other non-trivial techniques, it is not unreasonable to expect that sentiment indices with far more explanatory power could be generated.

It is hoped that this report encourages researchers and practitioners alike to think deeply about the true underlying mechanism of whichever sentiment index they employ; studying the factor loadings of the index may prove helpful for this. We provide some initial evidence that more than one principal component may explain sentiment, rather than just the first. One could argue that if there is indeed more than one component related to sentiment, then it would be unlikely that these components would be orthogonal. Following this, a non-orthogonal decomposition (e.g. independent component analysis) could be performed, and the components studied. Once again, the difficult question is how one determines the extent to which these components represent sentiment.

Finally, we observe that proxies lagged more than one period appear to hold significant explanatory power on market returns, such that sentiment may operate in a wider time period than initially thought

[14] The BW index may well represent sentiment to the full extent, but there is no real way of knowing without an external benchmark.

by Baker and Wurgler (2006). We therefore expect that sentiment index performance could be improved with a choice of more lags during construction. The parsimony of the resulting index can be maintained by only keeping the most correlative proxy lag to an initial principal component in the index.

References

D. Aggarwal. Defining and measuring market sentiments: a review of the literature. *Qualitative research in financial markets*, 2019.

W. Antweiler and F. Z. Murray. Is all that talk just noise? the information content of internet stock message boards. *Journal of finance*, 59:1259–1294, 2004.

C. Asness, A. Frazzini, N. J. Gormsen, and L. H. Pedersen. Betting against correlation: Testing theories of the low-risk effect. *Journal of financial economics*, 135:629–652, 2020.

J. Bae, C. Kim, and C. R. Nelson. Why are stock returns and volatility negatively correlated? *Journal of empirical finance*, 14(1):41 – 58, 2007.

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

M. Baker and J. Wurgler. The equity share in new issues and aggregate stock returns. *Journal of finance*, 55:2219–2257, 2000.

M. Baker and J. Wurgler. A catering theory of dividends. *Journal of finance*, 59:1125–1165, 2004.

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of finance*, 61:1645–1680, 2006.

M. Baker and J. Wurgler. Investor sentiment in the stock market. *Journal of economic perspectives*, 21: 129–152, 2007.

A. Bandopadhyaya and A. Jones. Measures of investor sentiment: A comparative analysis put-call ratio vs. volatility index. *Journal of Business and economics research*, 6:27–34, 02 2011.

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

D. Berger and H. J. Turtle. Cross-sectional performance and investor sentiment in a multiple risk factor model. *Journal of banking and finance*, 36:1107–1121, 2012.

R. S. Billingsley and D. M. Chance. Put—call ratios and market timing effectiveness. *The Journal of Portfolio Management*, 15(1):25–28, 1988.

F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 meetings of the business and economic statistics section*, pages 177–191, 1976.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81 (3):637–654, 1973.

J. C. Brau and S. E. Fawcett. Initial public offerings: An analysis of theory and practice. *The Journal of Finance*, 61(1):399–436, 2006.

E. F. Brigham and J. F. Houston. *Fundamentals of financial management*. Cengage Learning, 2012.

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

J. Campbell, S. Grossman, and J. Wang. Trading volume and serial correlation in stock returns. *The quarterly journal of economics*, 108:905–39, 02 1993.

E. C. Chang, T. C. Lin, Y. Luo, and J. Ren. Ex-day returns of stock distributions: An anchoring explanation. *Management science*, 65:1076–1095, 2019.

R. Chen, J. Yu, C. Jin, and W. Bao. Internet finance investor sentiment and return comovement. *Pacific-basin finance journal*, 56:151–161, 2019.

C. L. Concetto and F. Ravazzolo. Optimism in financial markets: Stock market returns and investor sentiments. *Journal of risk and financial management*, 12(2):85, 2019.

D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression I. *Biometrika*, 37 (3-4):409–428, 12 1950.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3 – 56, 1993.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116:1–22, 2015.

B. Gebka and M. E. Wohar. Causality between trading volume and returns: Evidence from quantile regressions. *International review of economics & finance*, 27:144 – 159, 2013.

M. Glaser, P. Schmitz, and M. Weber. Individual investor sentiment and stock returns-what do we learn from warrant traders? *SSRN electronic journal*, 2009.

T. Hahn, M. O'Neill, and J. Swisher. What does book-to-market proxy: Risk or investor sentiment? *Academy of accounting and financial studies journal*, 14:85–100, 01 2010.

A. Han and C. Y. Chung. Are individual investors less informed than institutional investors? unique evidence from investor trading behaviours around bad mergers in korean financial market. *Applied economics letters*, 20(12):1145–1149, 2013.

D. Huang, F. Jiang, J. Tu, and G. Zhou. Investor sentiment aligned: A powerful predictor of stock returns. *The review of financial studies*, 28:791–837, 2014.

S. Huddart, M. Lang, and M. H. Yetman. Volume and price patterns around a stock's 52-week highs and lows: Theory and evidence. *Management science*, 55(1):16–31, 2009.

Carlos M. Jarque and Anil K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255 – 259, 1980.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Journal of the econometric society*, 47:263–291, 1979.

S. Karpavičius and F. Yu. Dividend premium: Are dividend-paying stocks worth more? *International review of financial analysis*, 56:112 – 126, 2018.

S. Kelly and K. Ahmad. Estimating the impact of domain-specific news sentiment on financial assets. *Knowledge-based systems*, 150:116–126, 2018.

A. Kumar and C. M. C. Lee. Retail investor sentiment and return comovements. *Journal of Finance*, 61: 2451–2486, 2006.

C. M. C. Lee, A. Shleifer, and R. H. Thaler. Investor sentiment and the closed-end fund puzzle. *The journal of finance*, 46(1):75–109, 1991.

T. P. Lemke, G. T. Lins, and A. T. Smith. *Regulation of investment companies*, volume 1 of *Regulation of investment companies*. Lexis Nexis Matthew Bender, 1995.

B. Mizrach and S. Weerts. Highs and lows: a behavioural and technical analysis. *Applied financial economics*, 19(10):767–777, 2009.

T. Mullen and N. Collier. Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP-04, 9th conference on empirical methods in natural language processing*, 2004.

R. Neal and S. M. Wheatley. Do measures of investor sentiment predict returns? *The journal of financial and quantitative analysis*, 33(4):523–547, 1998.

E. Panopoulou and S. Plastira. Fama french factors and us stock return predictability. *Journal of asset management*, 15(2):110–128, 2014.

J. R. Ritter. The long-run performance of initial public offerings. *Journal of finance*, 46:3–27, 1991.

H. Shefrin. *A behavioral approach to asset pricing*. Elsevier, 2 edition, 2008.

H. Shefrin. Investors' judgments, asset pricing factors and sentiment. *European financial management*, 21 (2):205–227, 2015.

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

M. Simutin. Cash holdings and mutual fund performance. *Review of Finance*, 18(4):1425–1464, 08 2013.

R. F. Stambaugh, J. Yu, and Y. Yuan. The short of it: Investor sentiment and anomalies. *Journal of financial economics*, 104(2):288 – 302, 2012.

C. Wang. Investor sentiment, market timing, and futures returns. *Applied financial economics*, 13:871–878, 02 2003.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang. Daily stock market forecast from textual web data. In *SMC'98 conference proceedings. 1998 IEEE international conference on systems, man, and cybernetics (Cat. No.98CH36218)*, volume 3, pages 2720–2725, 1998.

C. Yang and R. Zhang. Dynamic sentiment asset pricing model. *Economic modelling*, 37:362 – 367, 2014.

J. Yu and Y. Yuan. Investor sentiment and the mean–variance relation. *Journal of financial economics*, 100(2):367 – 381, 2011.

G. Zhou. Measuring investor sentiment. *Annual review of financial economics*, 10:239–259, 2018.