

# Exploiting investor sentiment for portfolio optimization

N. Banholzer<sup>1</sup> · S. Heiden<sup>2</sup> · D. Schneller<sup>2</sup>

Received: 5 July 2017 / Accepted: 5 April 2018 / Published online: 9 May 2018  
© The Author(s) 2018

**Abstract** The information contained in investor sentiment has up to now hardly been used for portfolio optimization, although theoretical works demonstrate that it should not be neglected and it has already been shown to contain exploitable information on future returns and volatility. Employing the approach of Copula Opinion Pooling, we explore how sentiment information regarding international stock markets can be directly incorporated into the portfolio optimization procedure. We subsequently show that sentiment information can be exploited by a trading strategy that takes into account a medium-term reversal effect of sentiment on returns. This sentiment-based strategy outperforms several benchmark strategies in terms of different performance and downside risk measures. More importantly, the results remain robust to changes in the parameter specification.

**Keywords** Portfolio optimization · Investor sentiment · Copula opinion pooling · Behavioral finance

**JEL Classification** G11 · G15 · G17

---

✉ S. Heiden  
sebastian.heiden@wiwi.uni-augsburg.de  
N. Banholzer  
nbanholzer@ethz.ch

<sup>1</sup> Management Information Systems, ETH Zurich, Weinbergstr. 56/58, 8092 Zurich, Switzerland

<sup>2</sup> Department of Statistics, University of Augsburg, Universitaetstr. 16, 86159 Augsburg, Germany

## 1 Introduction

De Long et al. (1990) model sentiment as an overly optimistic or pessimistic view of the so-called ‘noise traders’ on the financial markets. Through their correlated behavior, noise traders collectively drive asset prices away from their fundamental values. In addition, De Long et al. (1990) show that this over- or under-pricing may not be arbitrated by rational arbitrageurs in the short term, as they could earn substantial negative returns when the noise traders drive up prices even further. However, when the mispricing becomes apparent, asset prices return to their fundamental values and, as a result, present arbitrage opportunities for investors in the medium-to-long term. Consequently, exploiting the mean-reversion effect of sentiment is essentially a matter of market timing. Sentiment has shown predictive power for future asset returns and volatility. In addition, theoretical works like Fu et al. (2015) show that it is an important factor for the mean–variance relation and should be included in portfolio optimization. Up to now, to the best of our knowledge, sentiment has been neglected in empirical studies of portfolio optimization. Our paper’s central research question is, therefore, how sentiment can be usefully incorporated into a portfolio optimization framework. For the first step, this requires finding a suitable measure of sentiment as a latent factor for the valuation of asset prices.

### 1.1 Measuring sentiment

There are different approaches to measuring sentiment: initially, surveys or market-based quantities were the only two routes to gather investor sentiment. Over the last decade, text mining, big-data analysis, social media, and search engine data have given birth to a third source for obtaining measures of investor sentiment.<sup>1</sup>

Concerning surveys, a frequent point of criticism is whether and how the participants of surveys act on financial markets according to their expressed sentiment. In our empirical investigation, we extract sentiment from market-based measures with data availability over long periods.

We thereby extend the empirical data set of Baker and Wurgler (2006) and Baker et al. (2012) (for a description of our data set, see Sect. 3.1). Sibley et al. (2016) question whether the sentiment index of Baker and Wurgler (2006) really measures sentiment or if, on the contrary, its inherent predictive power can be mainly explained by economic fundamentals and risk factors. They conclude that the predictive power of the sentiment index of Baker and Wurgler (2006) is mainly driven by information related to fundamentals and the part unexplained by those fundamentals possesses only very modest predictive power. Neely et al. (2014), on their part, discover that the index of monthly sentiment changes of Baker and Wurgler (2007) can be significantly predicted, especially during periods of recession, by technical indicators, whereas macroeconomic variables do not possess predictive power for future changes in sentiment. Taking these criticisms of the index of Baker and Wurgler (2006) into account, Internet-based measures of

<sup>1</sup> We thank an anonymous referee for valuable input regarding this strand of the literature.

sentiment could be attractive as well for our study: for example, Tetlock (2007) analyzes whether a ‘pessimism factor’ extracted from a column of the *Wall Street Journal* contains predictive power for market prices. He concludes that media pessimism exerts negative price pressure on the daily returns of the Dow Jones Stock Index. This downward pressure is followed by significant price reversals during the following trading days. The FEARS index of Da et al. (2015) is built using sentiment-related terms, based on the search queries of Google users. Da et al. (2015) show its predictive power for return reversals for 1- and 2-day horizons, and point out several advantages of search-query-based sentiment data over questionnaires: high-frequency data sets are easily available and Internet searches should reveal more personal information. Recently, Bucher (2017) finds that the return predictability of the FEARS index concerning a cross-section of stocks with different exposures to FEARS sentiment does not stem from its sentiment loading, but can be explained by return reversals and momentum.

Nevertheless, despite the criticisms raised by Sibley et al. (2016) and the attractiveness of search-based sentiment measures, we favor the index of Baker and Wurgler (2006) for several reasons: first, it is a very well-established quantity in the literature and the results of Stambaugh et al. (2014), who, in a vast simulation study, test whether different capital market anomalies could be spuriously explained by the sentiment index of Baker and Wurgler (2006), strongly support the view that the index of Baker and Wurgler (2006) possesses informational value. Second, although the analysis of Sibley et al. (2016) reveals that ‘pure sentiment’ may not be the main component of the index, the index seems to reflect very well a multitude of different economic variables and ‘soaks up’ their information. Finally, even Sibley et al. (2016) acknowledge that the causality between sentiment and fundamental variables is still unclear and highlight that ‘it is possible that our economic fundamental variables are influenced by sentiment itself’ (Sibley et al. 2016, p. 178). Therefore, we employ the index of Baker and Wurgler (2006), although we acknowledge that its information may not be ‘pure sentiment’ but does contain a rational part. Concerning the attractiveness of search-based measures, it should also be possible to select combinations of keywords to extract sentiment from Google searches also for our index data. Despite this, we do not follow this route for two reasons. On one hand, Google-based search engine data are only available after 2004. As we aim to test a sentiment-based portfolio strategy over as many business cycles as possible, the choice of search engine data would severely limit the time span of our data set.

Finally, Bucher (2017) remarks that the FEARS index is primarily interesting because of its cross-sectional features at daily and weekly horizons. We, therefore, do not suspect that it will be useful for our portfolio strategies addressing much longer horizons.

## 1.2 Sentiment and its information about returns, volatility, and portfolios

While the influence of sentiment on both returns and volatility has been examined in various empirical studies, the potential of sentiment for portfolio optimization has hitherto been largely neglected. We aim to fill this research gap by integrating sentiment directly into a portfolio optimization procedure. We especially expect

sentiment to be a signal for medium-term mean reversion in stock markets. This is supported by empirical evidence in the literature which covers the interrelation between sentiment and returns. Regarding sentiment and its connection to future returns, the early studies, such as Solt and Statman (1988) and Clarke and Statman (1998), rather explore the genesis of the expectations contained in sentiment and come to mixed results. These rather point to sentiment following the trend of the market. Later studies benefit from the possibility of using more extensive data sets and come to different conclusions. Brown and Cliff (2005) and Schmeling (2007) find empirical evidence for sentiment being a risk factor in international asset markets. Brown and Cliff (2005) integrate various classical asset pricing factors and discover, in a predictive regression setup, that sentiment contains valuable information for future returns.

They interpret their findings as an indication of ‘price pressure’ by irrational sentiment traders in financial markets who drive markets up (down) by their excessive optimism (pessimism), before rational arbitrageurs can force the prices to revert back to fundamentally justified levels. Stambaugh and Yuan (2016) underline the importance of sentiment in the stock market, as their four mispricing factors, created by combining multiple stock market anomalies, can be partly predicted by investor sentiment in monthly regressions.

Regarding medium-term horizons, there are strong empirical indications that high (low) private investor sentiment predicts negative (positive) asset returns: Schmeling (2009) examines the relation between sentiment and international stock markets and discovers that sentiment on average has negative predictive power for the aggregate stock market. Especially for horizons of 1–6 months, his analysis finds negative predictive power of the sentiment measure, which fades away for longer horizons (12–24 months). Brown and Cliff (2005) are able to show that this mean-reverting behavior can be interpreted as the current optimism being a predictor of substantially lower subsequent returns on the horizon of 2–3 years. In our analysis, we aim to use the mean reversion of sentiment-driven markets over medium-to-long-term horizons starting from 6 months, to ensure the robustness of the results to the choice of the selected horizon.

While the effect of sentiment on volatility has not been as extensively investigated as its effect on returns, there is nevertheless a growing body of empirical research on this topic. Theoretically, the framework of De Long et al. (1990) already models how the sentiment of irrational traders should increase the volatility of financial assets that are held by ‘noise traders’. Employing the average ‘bullishness’ of private investors on stock message boards as a sentiment proxy, Antweiler and Frank (2004) show that it is strongly linked to the next day’s volatility. Very recently, realistic out-of-sample forecasting approaches to realized volatility and sentiment data have begun to be employed. For example, Schneller et al. (2018) use sentiment data from a survey of (mainly) German and European investors, and find that investor sentiment can be used to profit from a local information advantage when forecasting realized volatility.

While many studies have examined the relation between sentiment, volatility, and returns, the potential of sentiment for portfolio optimization has hitherto been largely neglected. Meanwhile, the results of Yu and Yuan (2011) show that

sentiment alters the mean–variance trade-off. This suggests that the influence of sentiment on the optimal combination of risk and return of different assets in a portfolio setting should be the subject of further empirical investigation. Theoretically, the influence of sentiment on portfolios is also well founded: Fu et al. (2015) extend the classical setting of Markowitz to include sentiment. They conclude, especially by considering further empirical evidence on financial markets, that ‘a rational investor neglecting the effect of aggregate investor sentiment may end up selecting a sub-optimal portfolio’ (Fu et al. 2015, p. 272). The main goal of our empirical study, therefore, is to examine whether sentiment can be used directly for portfolio optimization. To gain insights into the ways in which sentiment has been used up to now, we briefly review the literature on trading strategies employing sentiment measures.

### 1.3 Sentiment-based trading strategies

As sentiment has been shown to contain information on future returns and volatility, various studies have investigated whether this information could be exploitable using trading strategies. In the following, we highlight the results of empirical approaches that use sentiment for trading strategies. Schmeling (2007) analyzes whether the predictive power of survey-based sentiment variables targeting private and institutional investors can be used to develop trading strategies. Schmeling (2007) discovers that even simple and easily implementable trading strategies yield consistently higher Sharpe ratios than the Buy-and-Hold benchmarks, on five international markets. One possible objection to the results of Schmeling (2007) is that his data set is rather small, as it only covers about 260 weekly observations. Therefore, it does not cover multiple business cycles and Schmeling (2007, p. 143) acknowledges that there are clear signs of a structural break in the middle of the data set which lets the discovered negative effect of private investors’ sentiment on future returns disappear in the second half of his sample. For that reason, it seems especially interesting to examine whether sentiment-based trading strategies consistently outperform benchmark strategies using a long data set spanning multiple decades and business cycles. Stambaugh et al. (2012) examine whether mispricing in stock markets can be exploited using long–short strategies for pricing anomalies which they attribute to sentiment. They show that especially sentiment-induced overpricing should be exploitable, because exploiting under-pricing should be more difficult due to short-sale impediments (see Stambaugh et al. 2012, p. 301). Very recently, sentiment data gained from the social media platform Twitter have been used for formulating trading strategies, as well. Sul et al. (2017) use the sentiment of Twitter users concerning different stocks in the SP500. By applying a long–short portfolio approach, they show that their strategy yields significant returns even if transaction costs are considered.<sup>2</sup> The profits are not only significant, but

---

<sup>2</sup> Sul et al. (2017) analyze daily sentiment and purchase the 10% of firms with the most positive sentiment and short sell the 10% of firms with the most negative sentiment. They thereby aim to profit from a short-term overvaluation of high sentiment stocks.

also of considerable size: they earn annual returns of about 15% (10-day holding period) and 11% (20-day holding period).

Summing up, sentiment has been shown to provide information on future asset returns, which seems to be economically exploitable. While there already have been numerous studies trying to exploit this information, primarily by forecasting either returns or volatility and subsequently using suitable trading strategies or by employing nonparametric long–short portfolio approaches, there are no studies which integrate sentiment into a portfolio optimization framework. We aim to fill this void in the literature by means of a realistic out-of-sample trading strategy. In the following section, we formulate our central research questions.

## 2 Research questions

Regarding the theoretical and empirical literature about the influence of sentiment on future returns and volatility, we discover several research gaps in the literature. Especially, investor sentiment has not yet been used for portfolio optimization, although it has shown to be an important risk factor in financial markets. In addition, the Copula Opinion Pooling (COP) approach of Meucci (2006b) has rarely been used in portfolio applications, although it has highly attractive properties for modeling non-normal markets and integrating investor views. We, therefore, aim to contribute to the sentiment literature using sentiment not for forecasting returns and volatility and forming portfolios based on these forecasts, but rather by integrating investor sentiment directly into the portfolio optimization. In brief, we aim to examine three core research questions in our empirical study:

$Q_1$  : How can investor sentiment be integrated into portfolio optimization using the COP methodology?

To empirically examine this question, we integrate different measures of sentiment into a portfolio optimization procedure. These measures are condensed into a common factor, which represents sentiment regarding international stock markets. We thereby integrate sentiment in a way that incorporates the results of both the theoretical and empirical literature regarding the effects of investor sentiment on stock markets.

$Q_2$  : Does investor sentiment provide profitable information for portfolio optimization?

Quantitatively, we assess this research question by comparing the performance of a portfolio that uses the COP approach and incorporates sentiment with one that neglects sentiment information. We compare the risk-adjusted returns of the sentiment-based strategy to several benchmark portfolios that have been shown to be hard to beat by the classical approach of Markowitz (1952) using a realistic out-of-sample setup. We conduct various robustness checks to ensure that our results are not some sort of ‘statistical artifact’. In addition, we investigate whether the better performance of the ‘sentiment strategy’ can be attributed to the use of sentiment

information or to the COP approach. We, therefore, apply a bootstrap procedure to our sentiment data. The results are summarized in Sect. 4.3.

$Q_3$  : Does a medium- to long-term investment horizon optimally exploit the mean reversion of sentiment-induced mispricing?

Based on the empirical literature (see Sect. 1), the medium- to long-term investment horizon (6 months and beyond) should show the best results. In brief, the literature has shown that investors' optimism (pessimism) significantly predicts negative (positive) returns on different medium-term horizons, as the market mean reverts after a short-term overvaluation (undervaluation). As we use past sentiment concerning different assets to determine the weights in a portfolio, we, therefore, underweight assets where past sentiment has been especially high and overweight assets with low past sentiment.

As it is not a priori known when the reversal effect takes place, we examine different medium-term horizons to add robustness to our findings.

### 3 Data and empirical approach

In this section, we describe our data set and empirical approach. First, we introduce the sentiment measures that are used to construct the sentiment indexes. With these indexes, we then optimize our portfolio using the COP approach of Meucci (2006a). Our portfolio consists of five risky assets. These are the four major stock indexes EURO STOXX 50 (Europe = EU), FTSE100 (United Kingdom = UK), NIKKEI225 (Japan = JP), and SP500 (United States = US), as well as the USD Gold Price per troy ounce (GLD).<sup>3</sup> The latter is added as an alternative asset to the portfolio and should receive higher portfolio weight when the sentiment indexes globally project negative future stock market returns (see Sect. 3.3).

#### 3.1 Sentiment proxies and index computation

Baker and Wurgler (2006) and Baker et al. (2012) use proxies for investor sentiment to construct sentiment indexes that explain future stock returns. However, the monthly proxies in Baker and Wurgler (2006) are restricted to the US market and the proxies for Germany, Japan, the UK, and the US in Baker et al. (2012) are only available on a yearly basis. We draw upon both studies by employing the index of Baker and Wurgler (2006) for the US and computing the indexes for EU, JP, and UK from four of the proxies in Baker and Wurgler (2006) and Baker et al. (2012) on a monthly basis. In addition, a global sentiment index is constructed as in Baker et al. (2012), which is presumed to be inversely related to GLD.

##### Sentiment proxies

The first sentiment proxy in Baker et al. (2012) is the closed-end fund discount (CEFD), i.e., the discount (or premium) if the closed-end fund is trading below (above) its net-asset value. A high discount should arise when investors are sceptical

<sup>3</sup> From now on, we refer to the respective index when speaking of the market in brackets.

about the future performance of the fund. Thus, we expect a negative relation between the closed-end fund discount and sentiment. The discount is calculated by considering all equity funds listed on *Morningstar Direct*. Thereby, each closed-end equity fund is assigned to one market, according to Morningstar's global category. After removing outliers, we compute the net-asset-value-weighted average closed-end fund discount in each market.<sup>4</sup>

The best timing for an initial public offering (IPO) is when investors are bullish and when, therefore, high returns are expected (see Baker and Wurgler 2006). This is the theoretical background for the number of IPOs being a good proxy for investor sentiment. *SDC Platinum* provides the number of IPOs per market and month. As in Baker and Wurgler (2006), we use the sum of IPOs over the last 12 months to smooth the data.

The third measure is log market turnover (TV).<sup>5</sup> High market turnover is often an indicator that a 'bubble' is forming in the market and should, therefore, be positively related to investor sentiment (see Baker and Wurgler 2006).

The last sentiment proxy is the volatility premium (VP). Baker et al. (2012) define this as the log difference of the value-weighted average market to book ratio (PBV) between the stocks with the 30% highest and 30% lowest beta-adjusted idiosyncratic volatility ( $\sigma$ ). Formally, the volatility premium at time  $t$  is as follows:

$$VP_t = \log \left( \sum_{i \in I} c_{it} \cdot PBV_{it} \right) - \log \left( \sum_{j \in J} c_{jt} \cdot PBV_{jt} \right), \quad (1)$$

where  $I$  is the set of high volatility stocks, i.e.,  $I = \{i : \sigma_i \geq \hat{F}_{\sigma_i}^{-1}(0.7)\}$ ,  $J$  is the set of low volatility stocks, i.e.,  $J = \{j : \sigma_j \leq \hat{F}_{\sigma_j}^{-1}(0.3)\}$ , and  $c_{it}$ , respectively,  $c_{jt}$ , is the market capitalization of stock  $i$ , respectively,  $j$ , at period  $t$ .<sup>6</sup>

Baker et al. (2012) point out that the volatility premium is related to the dividend premium considered in Baker and Wurgler (2006) and can be computed monthly for all markets. As Baker et al. (2012) illustrate both on a theoretical and empirical basis, stocks that are hard to value attract noise traders who are willing to buy them even at high prices. Hence, a high volatility premium should indicate a strong sentiment effect on these stocks.

Table 1 provides a brief summary of our data set.

### Sentiment indexes

Analogously to Baker and Wurgler (2006), we construct a monthly sentiment index out of these proxies for EU, JP, and UK. The procedure is based on the intuition that all variables are driven by one latent factor: investor sentiment. To check whether factor analysis is applicable, we conduct Bartlett's sphericity test. It tests the hypothesis  $H_0$  that the correlation matrix is equal to the identity matrix

<sup>4</sup> Outliers are all those discounts,  $CEFD_{OUT}$ , that are in the lower and upper 1% quantile, i.e.,  $CEFD_{OUT} \leq \hat{F}_{CEFD}^{-1}(0.01)$  and  $CEFD_{OUT} \geq \hat{F}_{CEFD}^{-1}(0.99)$ .

<sup>5</sup> Total market turnover is computed [similarly to Baker et al. (2012)] as the ratio of turnover volume of all stocks listed in the country in the current month divided by the market value of the previous month.

<sup>6</sup> Negative price to book values are removed as they lead to negative arguments in the log ratio in some months.



**Table 1** Sentiment variables and sources

Variable	Description	Source
CEFD	Closed-end fund discount, NAV-weighted	Morningstar Direct
IPO	Number of IPOs, sum prior 12 months	SDC Platinum
TV	Total market turnover, log	Datastream
VP	Volatility premium	Datastream

against the alternative hypothesis that the correlation matrix diverges from the identity matrix.  $H_0$  is rejected for each market's proxies, and hence, principal component analysis is applicable.

Baker and Wurgler (2006) define their US sentiment index as the standardized first principal component of their sentiment proxies. The first principal component is a linear combination of the proxies that accounts for as much joint variation of the proxies as possible.

The first principal component explains, respectively, 39% (EU), 40% (JP), and 44% (UK) of the proxies' variance. These proportions are similar to those found in Baker and Wurgler (2006) and Baker et al. (2012).

In addition, we compute a global sentiment index in analogy to Baker et al. (2012). The global sentiment index is computed as the standardized first principal component of the four (standardized) national sentiment indexes. This index should reflect investors' general sentiment about stock markets and, therefore, be inversely related to the alternative asset GLD.

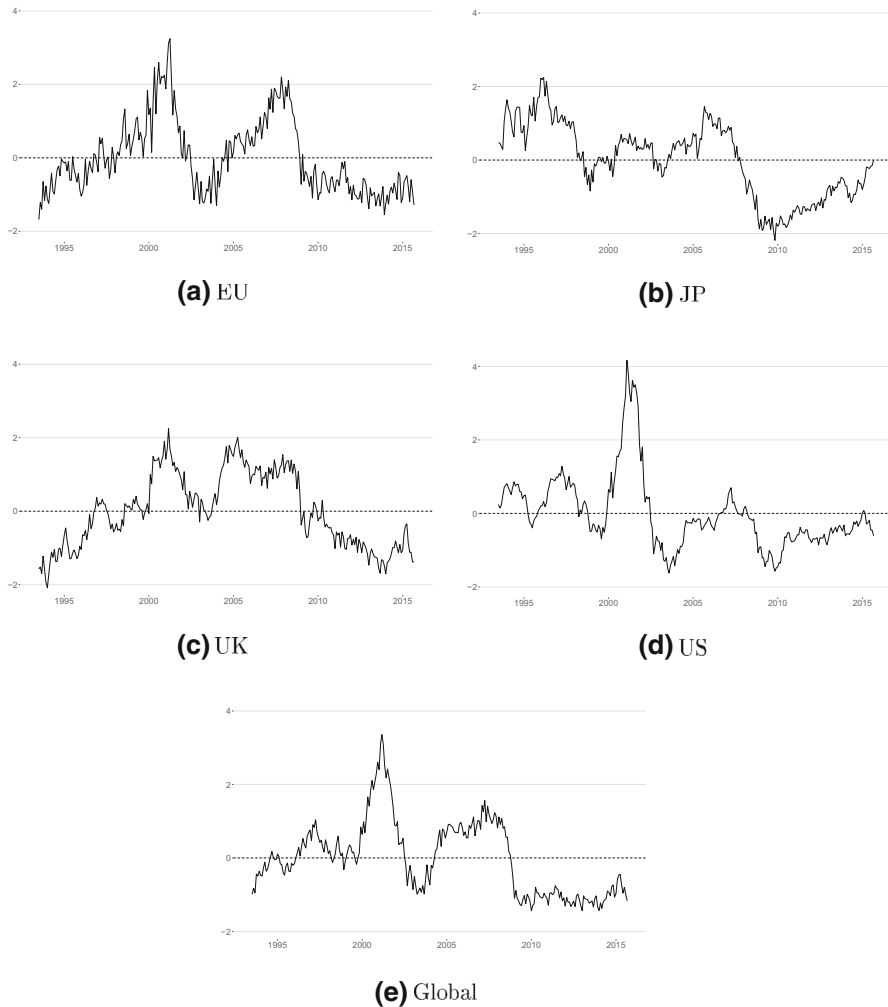
Figure 1 depicts the sentiment index for each market over time.

The sentiment indexes are in accordance with major stock market episodes. For example, all indexes reach a very high level of investor sentiment prior to the global financial crisis starting in 2007 and experience a large drop as the crisis unfolds. In addition, the recovery from the crisis starting around 2010 is visible. Likewise, the period of the dot.com bubble is captured by sudden shifts in investor sentiment before and after the bubble's collapse between 2000 and 2002, with the exception of the Japanese index. Unsurprisingly, the global crises and recoveries are also observable in the global sentiment index.

The following section describes the COP method used to incorporate investor sentiment into portfolio optimization.

### 3.2 Copula opinion pooling

The following description of the COP approach is based on Meucci (2006a, b) and Gochez et al. (2015). An investor can express views about the returns of  $N$  assets in that investor's portfolio. The views can be absolute (e.g., 'The DAX will have a return of 3%.') or relative ('The DAX will outperform the Nikkei by 2%.'). To specify a view in the COP framework, three components have to be defined: the  $I \times N$  'pick matrix'  $\mathbf{P}$ , with  $I$  indicating the number of views, the 'view distribution', and the confidence in the view.



**Fig. 1** Sentiment indexes by market over time. The sentiment indexes for EU, JP, and UK are computed as the first principal component of sentiment proxies used in Baker and Wurgler (2006) and Baker et al. (2012): CEFD (closed-end fund discount), NIPO (number of IPOs), TV (turnover), and VP (volatility premium). For the US, the index by Baker and Wurgler (2006) is used. In addition, a global sentiment index is computed as the first principal component of the four market indexes. All indexes are standardized

First, the multivariate distribution of the asset returns is fitted by a multivariate distribution. Therefore, any multivariate distribution is applicable. As the COP approach has no closed-form solution, simulations from this multivariate distribution are necessary.

The number of simulations is denoted by  $S$ . The  $N \times S$  matrix  $\mathbf{M}$  contains the  $S$  simulations. Each column represents one ‘market scenario’, meaning one

simulation from the multivariate distribution with the ‘market’ consisting of the  $N$  assets.

In the next step, the so-called ‘market-implied views’ are computed:

$$\mathbf{V} = \mathbf{PM}. \tag{2}$$

$\mathbf{V}$  results in an  $I \times S$  matrix out of the pick matrix and the market simulations. For each market scenario/simulation  $s$ , each element  $v_{is}$  of  $\mathbf{V}$  represents the outperformance that would result from the view  $i$ .

The investor has potentially a differing opinion about the returns that are generated by view  $i$ . Therefore, the investor has to specify a ‘view distribution’ for each of these  $I$  views. The view distribution is allowed to be any distribution. The  $i$ th row  $\hat{\mathbf{V}}_i$  of the  $I \times S$  matrix  $\hat{\mathbf{V}}$  contains simulations from the distribution of the investor’s view  $i$ . This means that the number of simulations from the multivariate distribution of the assets corresponds to the number of simulations from each of the investor’s view distributions.

In the next step, the market-implied views and the investor’s views are combined. This means that the resulting  $I \times S$  matrix  $\tilde{\mathbf{V}}$  contains elements of  $\mathbf{V}$  as well as of  $\hat{\mathbf{V}}$ . For each element  $\tilde{v}_{ij}$ , one has:

$$P(\tilde{v}_{ij} = \hat{v}_{ij}) = \lambda_i \tag{3}$$

and

$$P(\tilde{v}_{ij} = v_{ij}) = 1 - \lambda_i, \tag{4}$$

with  $\lambda_i \in [0; 1]$  indicating the investor’s confidence in view  $i$ . Hence, the elements of  $\mathbf{V}$  and  $\hat{\mathbf{V}}$  are sampled into  $\tilde{\mathbf{V}}$  depending on the confidence  $\lambda_i$  that the investor has in the view  $i$ . As  $\tilde{\mathbf{V}}$  combines realizations of the market-implied views as well as realizations of the investor’s views, the dependence structure of these combined views is quite likely to be different from the dependence structure of the market-implied views  $\mathbf{V}$ .

In the next step of the COP approach, the dependence structure of  $\mathbf{V}$  [Eq. (2)] is retained and transferred to  $\tilde{\mathbf{V}}$ .

Therefore, the  $I \times S$  matrix  $\mathbf{Z}$  is introduced. Each row  $\mathbf{Z}_i$  contains the values of the empirical cdf of  $\mathbf{V}_i$ . Thereby,  $z_{ij}$  is the value of the empirical cdf of  $v_{ij}$  in  $\mathbf{V}_i$ , that is

$$z_{ij} = \hat{F}_{\mathbf{V}_i}(v_{ij}). \tag{5}$$

The matrix  $\mathbf{Z}$  contains the dependence structure, i.e., the copula of  $\mathbf{V}$ . The elements  $d_{ij}$  of the  $I \times S$  matrix  $\mathbf{D}$  are computed as follows:

$$d_{ij} = \hat{F}_{\tilde{\mathbf{V}}_i}^{-1}(z_{ij}). \tag{6}$$

As each row  $\mathbf{D}_i$  is computed according to  $\mathbf{Z}_i$ , the dependence structure, i.e., the copula of  $\mathbf{V}$ , is transferred to the combined views  $\tilde{\mathbf{V}}$ . The resulting matrix  $\mathbf{D}$  contains the investor’s views and the market-implied views ( $\tilde{\mathbf{V}}$ ) and exhibits the dependence structure (copula) of  $\mathbf{V}$ .

If, as in the setup of this study, only one view is used, no copula has to be used but only the values of the empirical cdf of  $\mathbf{V}$ , which is then a  $1 \times S$  vector. If the number  $I$  of views is smaller than the number  $N$  of assets, an  $(N - I) \times N$  matrix  $\mathbf{Q}$  is introduced to ensure that  $((\mathbf{PQ}))$  is an invertible matrix. The  $(N - I) \times S$  matrix

$$\mathbf{R} = \mathbf{QM} \quad (7)$$

consequently completes the market-implied views. These so-called ‘orthogonal views’ do not affect the mathematical calculations regarding the other views. For the choice of  $\mathbf{Q}$ , see Meucci (2006a, b).

In the last step of the COP approach, the simulations of the market that incorporate the investor’s views are computed as follows:

$$\tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{P} \\ \mathbf{Q} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{D} \\ \mathbf{R} \end{pmatrix}. \quad (8)$$

Equation (8) rolls back the computations in Eqs. (2) and (7). Each of the  $S$  columns of the resulting  $N \times S$  matrix  $\tilde{\mathbf{M}}$  contains one simulated market scenario that combines the investor’s views and the implied views based on the investor’s confidences  $\lambda_i$ .<sup>7</sup> These market scenarios are subsequently used to compute the portfolio weights in period  $t$ .

Using the posterior market distribution, we obtain optimal asset weights applying classical mean–variance optimization and computing the Maximum Sharpe ratio portfolio. Although Meucci (2006b, 2006a) recommends calculating the Minimum CVaR portfolio, we apply the mean–variance approach, as it is standard in the literature. Furthermore, we use the Sharpe ratio as the main evaluation criterion and it should, therefore, also be used as the optimization criterion. The results for the Minimum CVaR portfolio are comparable to those of the Maximum Sharpe ratio portfolio and are available on request. Furthermore, we impose a long-only constraint and assume the portfolio to be always fully invested.

Up to now, despite its attractive properties, the COP approach has rarely been used in academic papers. This is even more surprising as the few papers that apply the approach show that its theoretical advantages also carry over into practical applications and yield economically significant results. Stein et al. (2009) analyze whether bond portfolios can benefit from the inclusion of hedge-fund strategies by means of the risk-adjusted performance. As hedge-fund investment styles are often characterized by the use of derivatives with non-linear payoffs, Stein et al. (2009) use the COP method to model the dependence structure of those strategies with the bond market. They use government bond indexes for Germany, the UK, Japan, and the US, as well as hedge-fund returns indexes regarding the following investment

<sup>7</sup> As our sentiment measure, which is extracted from many different financial variables, is supposed to represent the sentiment of the whole ‘crowd of investors’, we, therefore, treat it as highly valuable, discretionary information on future market movements, and aim to fully benefit from its informational value. We, therefore, set  $\lambda_i = 1$ .

styles: convertible arbitrage, fixed-income arbitrage, and event-driven/distressed securities.<sup>8</sup> Their results challenge the conventional view that hedge funds should be included in every well-diversified portfolio. Stein et al. (2009) highlight that the individual views and risk aversion of the investor play a major role in determining the investor's optimal proportion of hedge-fund investments. Simonian (2014) shows how conflicting views of different portfolio managers regarding the same asset can be modeled using the COP approach. To coherently aggregate conflicting views, he proposes a variant of the Shapley value, an approach known from game theory. By this means, the COP method can also be employed if more complex and even conflicting views regarding the same asset have to be integrated. This decision situation, of conflicting opinions, seems to be well known to practitioners in the investment industry, where portfolio decisions are often made by investment committees and different levels of hierarchy come into play. Regarding our own empirical analysis, we aim to benefit from the same attractive properties of the COP approach as Stein et al. (2009), but aim to explore whether investor sentiment can be used to improve portfolio allocations. To this end, we directly extract signals from sentiment data for the COP-based portfolio optimization. Our empirical approach is summarized in the following section.

### 3.3 Empirical approach

In this section, we describe the empirical approach that includes Copula Opinion Pooling (COP) (see Sect. 3.2). The aim of our approach is to examine whether information in sentiment can be used in a COP-based approach<sup>9</sup> and how our 'sentiment strategy' performs against various benchmark strategies.

It should be noted that, in contrast to the formulation of COP in Meucci (2006a, b), we do not form the view based on the assessment of only one investor, but based on the sentiment of a 'crowd of investors'. As this sentiment measure should be more accurate than the market assessment of a single investor, we, therefore, aim to improve the COP approach and make it more applicable using real-world financial data extracted from financial markets.

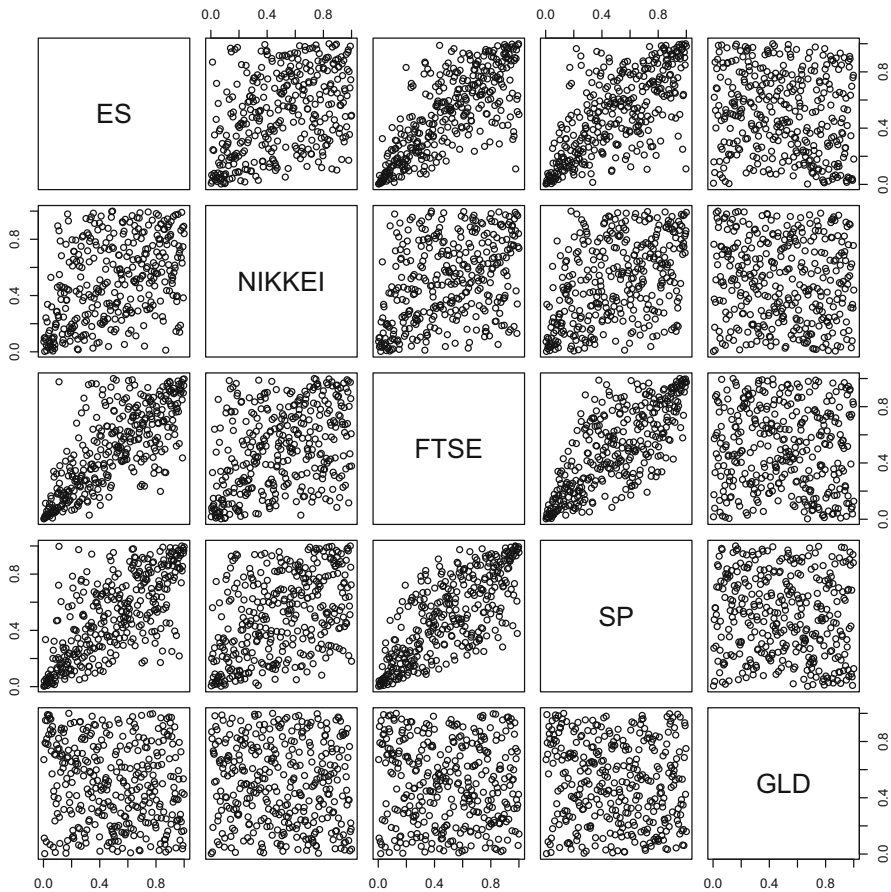
The first step is to specify the initial market distribution. Therefore, the distribution of market returns is fitted by a multivariate distribution. Figure 2 shows the pairwise scatter plots of the monthly returns of the stock market indexes and of the USD Gold price per troy ounce (GLD). GLD is considered as an alternative asset when global sentiment projects bearish stock markets.

As expected, the returns of the stock market indexes are highly correlated, whereas the pattern with GLD looks rather Gaussian.<sup>10</sup> In addition, the stock market indexes exhibit a tail dependence that can be modeled by means of a multivariate  $t$ -distribution. If there is also skewness to some extent, the model could be further improved by a multivariate skew  $t$ -distribution (see Azzalini and Capitanio 2003).

<sup>8</sup> For a very concise description of these investment styles, see Stein et al. (2009).

<sup>9</sup> For this purpose, we use the *BLCOP* package developed by Gochez et al. (2015).

<sup>10</sup> However, the pairwise correlation, as measured by Pearson's correlation coefficient, between each stock market's returns and GLD is negative.



**Fig. 2** Pairwise return scatter plot. The returns are transformed by their empirical cumulative distribution function, so that the dependence structure is not distorted by the marginal distributions and easier to interpret. ES denotes returns of the EuroStoxx50 index, NIKKEI of the Nikkei225, FTSE of the FTSE100, SP of the SP500, and GLD of gold

Therefore, we test the null hypothesis that the shape parameter  $\alpha$  of the skew  $t$ -distribution equals zero. The  $p$  values of the respective likelihood ratio test are depicted in Fig. 3 and suggest that there is significant skewness throughout the period that our data set covers. Consequently, we fit a multivariate skew  $t$ -distribution in every period  $t$  and estimate the parameters from the last 10 years of monthly market data. However, given the start of our sentiment indexes, the initial market distribution in  $t = 1$  is based only on market data from February 1987 to July 1993 ( $t = 1$ ).

Having specified the initial market distribution, we focus on a method to exploit the information contained in our sentiment indexes. We aim to achieve this goal by embedding sentiment into the view distribution. Originally, the view distribution allows managers to express their views about the performance of the assets in the portfolio (Black and Litterman 1992). In our setting, the value of the sentiment

index represents the view about the respective market. It is possible to express single or multiple views as well as absolute and relative views about each market. For every view, the corresponding distribution has to be specified. For example, suppose that the view is uniformly distributed between 1 and 2%, and that the sentiment indexes induce the following pick matrix:

$$P = \begin{pmatrix} EU & UK & JP & US & GLD \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (9)$$

This view suggests that the EURO STOXX 50 will yield a return between 1 and 2% in the next month. As another example, the pick matrix

$$P = \begin{pmatrix} EU & UK & JP & US & GLD \\ 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{pmatrix} \quad (10)$$

together with the same view distribution as above would imply that the EURO STOXX 50 outperforms an equally weighted portfolio of all other markets in the next month. Motivated by empirical findings, as outlined in Sect. 1, we specify a pick matrix that models the negative effect of sentiment on subsequent returns. That is, when the sentiment index for one stock market is positive, the pick element is negative and vice versa.

Inversely, if the global sentiment index is positive, GLD gets a positive pick and vice versa.<sup>11</sup>

The reason that we add GLD to the portfolio in the first place is to have an alternative asset when the sentiment indexes globally project negative stock market returns. We choose GLD, because it is a well-known and frequently traded commodity which often performs inversely to the stock market. In fact, GLD is negatively correlated with the four stock market indexes in our data set. Furthermore, it is a risky asset and can, therefore, be incorporated into mean-variance optimization when computing the optimal risky market portfolio. Alternatively, one could find a rule for determining the weight of the risk-free asset.<sup>12</sup> However, and considering the number of parameters that have to be specified, we find it more elegant to add GLD to the basket and use the sentiment indexes solely to specify the pick.

Other possible specifications of the pick matrix include a positive sentiment effect as well as weighting the absolute sentiment of each index relative to all other indexes. The latter may result in a pick matrix of the form:

<sup>11</sup> Regarding the implementation, it is more feasible to first mirror the global sentiment index, so that the specification of the pick matrix stays the same for all indexes.

<sup>12</sup> In terms of Markowitz's Modern Portfolio Theory, this means finding the point on the capital market line that maximizes the investor's utility. Thus, the weight of the risk-free asset could be determined by relating the global sentiment index to the parameter of risk-aversion in the investor's utility function.

$$P = \begin{pmatrix} EU & UK & JP & US & GLD \\ -\frac{2}{10} & -\frac{1}{5} & -\frac{1}{2} & -\frac{1}{10} & 1 \end{pmatrix}. \quad (11)$$

We consider such a weighted specification in the robustness checks in Sect. 4.3. Thereby, the individual negative (positive) pick is computed by dividing the respective positive (negative) sentiment index value by the sum of all other positive (negative) index values in such a way that both positive and negative picks sum to 1. After specifying the view distribution, we follow the procedure in Meucci (2006a) and simulate the market posterior distribution from which the optimal posterior portfolio weights are derived.

Several parameters have to be specified for the application of the COP method. First, the lag, meaning the time between the observed sentiment and the point in time when changes are made to the portfolio based on this sentiment. As implied by sentiment theory (see Sect. 1), sentiment has a medium-term reversal effect on returns. Therefore, the sentiment in period  $t - \tau$  is used to specify the portfolio weights in period  $t$ .

Second, the view distribution is parameterized. Instead of a uniform distribution, the views could also follow a normal distribution or be fixed at a certain value. Independently of the distribution, the parameters need to be set in a way that allows for changes in the market distribution and the portfolio weights to be perceived and attributed to the sentiment effect. Because finding a possible distribution for the payoff expected by the investor is not obvious, we chose the uniform distribution. As a starting point, we chose a reasonable parameterization of  $U(-0.005, 0.03)$ , which covers slightly negative as well as significantly positive payoffs, and translates into possible yearly returns of the strategy between  $-6\%$  and  $+36\%$ . However, we apply several different distributions as robustness checks (see Sect. 4.3).

We compare our ‘sentiment strategy’ against several benchmarks:

1. *Buy-and-Hold 1/n* This benchmark simply invests 20% in each asset in  $t = 0$  and holds this portfolio.
2. *Dynamic 1/n* In every period  $t$ , the portfolio is rebalanced to give equal weight to each asset, which could, otherwise, gain weight from price changes. It is worthy of note that DeMiguel et al. (2009) find that none of the 14 mean–variance optimal portfolio models considered in their paper is able to consistently outperform this naive diversification strategy out-of-sample.
3. *Dynamic market cap* The portfolio weight of index  $i$  in  $t$ ,  $w_{it}$ , is given by the relative market capitalization of the index in  $t$ ,  $c_{it}$ . Compared to the naive diversification of  $1/n$ , this is a market neutral portfolio.
4. *Initial Market* This benchmark uses the optimized portfolio weights from the initial market distribution. It is very important to compare the sentiment strategy against the initial market strategy so as to assess whether sentiment information improves on the results of a strategy that does not consider the sentiment effect. That is, if our ‘sentiment strategy’ performs better than the



initial market strategy, this improved performance can be attributed to the sentiment effect.

5. *Markowitz* This benchmark selects, in each period, the tangency portfolio based on Markowitz's Modern Portfolio Theory using the same historical market data as in our sentiment strategy and the initial market strategy.

In addition to a graphical visualization and analysis of our results, every strategy is evaluated regarding eight performance measures, which are based on excess returns, i.e., the market return over the respective 3-month government bond yield (for GLD, we use the 3-month US T-Bill). First, the *Annualized Return* and the *Annualized Sharpe ratio* are computed. Furthermore, we consider the *Omega ratio* (Keating and Shadwick 2002) and the *Downside Deviation* (Sortino and Price 1994). The higher the Omega ratio, the more gains there are relative to losses, when the loss threshold (also often referred to as the benchmark return) is set to zero.

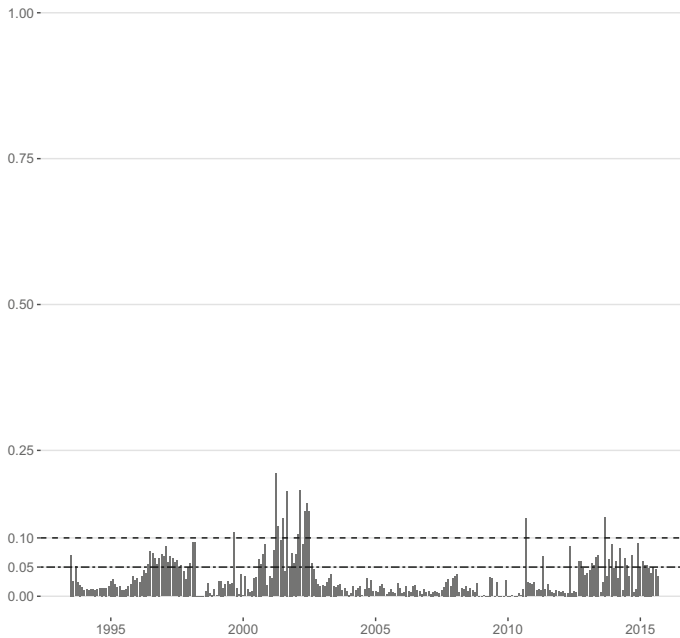
The Downside Deviation measures risk after eliminating positive returns when setting the minimum acceptable return to zero. Additional downside risk measures include the 95% *Value-at-Risk* (VaR) and *Conditional Value-at-Risk* (CVaR), which are computed using the quantiles from the empirical cumulative distribution. Finally, the *Certainty Equivalent Return* (CER) (with risk-aversion parameter  $\lambda = 1$ ) and the monthly monetary *portfolio turnover* complement the measures for comparing our results and are computed as in DeMiguel et al. (2009).

The following section presents the empirical results.

## 4 Results

The data set is split into a training period and a test period. The training period ranges from July 1993 to July 2004. The returns data prior to July 1993 are used for the initial fit of the initial market distribution. The training set is used to identify the best specification of the pick matrix, the time lag of the sentiment variables, and specifications for the other parameters. The resulting setup is subsequently applied in the test period (August 2004 to September 2015) to verify the results of the training period. In the test period, therefore, the method is applied using the parameters that were optimal in the training period: these parameters are not re-fitted. In Sect. 4.3, several robustness checks are presented.

In the training as well as in the test period, the empirical setup is structured as follows. At period  $t$ , we only use market returns data prior to period  $t$  to specify the initial market distribution. The pick matrix is generated using the information contained in the sentiment measures at period  $t - \tau$ , with  $\tau$  being the lag between the sentiment measures and the current period  $t$ . Using the COP method, we compute the posterior market simulations and use them for the computation of the optimal weights for the portfolio at time  $t$ . After holding the resulting portfolio for 1 month, the described procedure is repeated monthly. The portfolio weights for the benchmark strategies are also re-calculated monthly (except for the Buy and Hold).



**Fig. 3** The  $p$ -values of the likelihood ratio test on the shape parameter from the multivariate skew  $t$ -distribution over time. For each time  $t$ , the  $p$  values are obtained by fitting a multivariate skew  $t$ -distribution with (first model) and without a symmetry constraint (second model) to the  $t - 120$  historical market returns. A likelihood ratio test compares the models' goodness of fit. The dashed lines represent the 5 and 10% significance levels. If the bars are below the dashed lines, the shape parameter  $\alpha$  is significant and the use of a multivariate skew  $t$ -distribution justified

#### 4.1 Training period results

In this subsection, we briefly discuss the procedure for finding the best parameter combination in the training period, which is a multi-step optimization procedure. The first step in this procedure is to determine the best specification of the pick matrix. As described in Sect. 3.3, a pick matrix that models the negative effect of sentiment should exploit the reversal effect of investor sentiment on returns. This means that if the sentiment index  $S_m$  of market  $m$  at period  $t$  is positive (negative), the corresponding value in the pick matrix is negative (positive). The results in the training period identify this pick matrix as better than the positive pick matrix in terms of several performance measures. Having identified the general orientation, the next step is to determine if the weighted or unweighted version of the pick matrix should be applied. The comparison shows that both versions are very similar regarding their performance. Therefore, we chose the simpler, unweighted version of the pick matrix. As pointed out in Sect. 1, sentiment does not have an immediate effect on returns. Therefore, in the next step, the best lag  $\tau$  has to be determined. We find that the sentiment strategy outperforms the benchmark strategies on medium-term lags, which we define as between 6 and 24 months. This implies that for the

computation of the portfolio weights at period  $t$ , sentiment indexes with a medium-term lag should be used. As pointed out in Sect. 3.3, we want to impose as a few assumptions as possible regarding the prospective performance of the view in advance. Thus, the view distribution is selected to be uniform with parameterization  $U(-0.005; 0.03)$ . It should be noted that the mean of this distribution is positive. We chose a uniform view distribution with positive mean, since the outperformance that is generated by the view is expected to be positive. Therefore, this view distribution is not entirely uninformative, but uninformative regarding the probability of each value in the range  $[-0.005; 0.03]$ . Nevertheless, alternative parameterizations of the view distribution are part of the robustness checks in Sect. 4.3. In addition, the initial market distribution is estimated using 10 years of historical data.

Hereafter, by way of example, the results of the sentiment strategy with lag  $\tau = 12$  are shown. The cumulated returns of all strategies are shown in Fig. 4. The sentiment strategy outperforms all of the benchmark strategies in the training period. The fact that the sentiment strategy also outperforms the initial market strategy, which does not use sentiment information, is an initial indication that investor sentiment carries relevant information for portfolio optimization. This is examined in greater detail in Sect. 4.3. Regarding only returns, it should be noted that the outperformance of the sentiment strategy mostly stems from the last third of the training period, but the strategy already starts performing comparatively well in the second third. In the first two-thirds, however, the sentiment strategy is much less volatile than the other strategies and performs very well during the dotcom bubble, during which all the benchmarks show a clear downward trend.

However, the higher performance of the sentiment strategy could merely be generated by a higher risk of the strategy, as the cumulative returns shown in Fig. 4 are not a risk-adjusted performance measure. Therefore, we compute several performance measures, as condensed in Table 2.



**Fig. 4** Cumulative returns over time (training period). Cumulative returns of the sentiment strategy and the five benchmark strategies in the training period. The sentiment strategy uses lag  $\tau = 12$ , a uniform view distribution with  $U(-0.005; 0.03)$ , and the ‘negative pick matrix’

**Table 2** Performance measures for the training period

Strategy	AR	SR	DD	OR	VaR	CVaR	TO	CER
Sentiment	0.0360	0.3285	0.0205	1.3235	0.0456	0.0597	0.3481	0.0029
Initial market	0.0095	0.0686	0.0292	1.1079	0.0756	0.0924	0.2105	0.0008
Dynamic $1/n$	− 0.0198	− 0.1616	0.0263	0.9301	0.0582	0.0776	0.0362	− 0.0017
Buy-and-Hold $1/n$	− 0.0080	− 0.0637	0.0269	0.9995	0.0608	0.0806	0.0443	− 0.0007
Dyn. Market Cap	0.0102	0.0720	0.0293	1.1056	0.0678	0.0893	0.0454	0.0008
Markowitz	0.0151	0.1089	0.0291	1.1414	0.0758	0.0925	0.0908	0.0013

Performance measures for the sentiment strategy with the ‘negative pick matrix’ for lag  $\tau = 12$  and a uniform view distribution, as well as for the five alternative strategies for the training period. For an explanation of the performance measures, see Sect. 3.3. VaR and CVaR are displayed as absolute values. AR annualized return, SR Sharpe ratio, DD downside deviation, OR Omega ratio, VaR value at risk, CVaR conditional value at risk, TO turnover, CER certainty equivalent return

In terms of the Sharpe ratio, the CER, as well as the annualized returns, the sentiment strategy outperforms all of the benchmark strategies. The sentiment strategy has the highest Omega ratio, the lowest VaR and CVaR (both displayed in absolute values), and the lowest downside deviation (DD). As the sentiment strategy of course is an active strategy, it exhibits a higher turnover than the rather passive benchmark strategies.

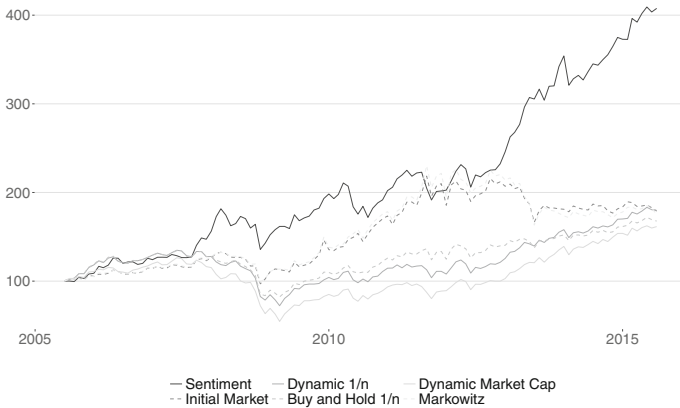
To conclude the results of the training period, the sentiment strategy outperforms the benchmark strategies in terms of seven of the eight performance measures. After the specification of the best parameter combination in the training period, the sentiment strategy is applied in the test period, since an out-of-sample trading exercise should be a ‘real test’ of the performance of the strategy.

## 4.2 Test period results

The test period ranges from August 2004 to September 2015 and, therefore, spans downturn and expansion periods of the market. As the theoretically well-founded ‘negative pick matrix’ is also identified as the best specification regarding the performance measures in the training period in conjunction with a medium-term lag  $\tau$  and a uniform view distribution, we again show the results for a lag of  $\tau = 12$  months. Therefore, we first examine the cumulative returns in Fig. 5.

Cumulatively, the sentiment strategy outperforms the benchmark strategies. The outperformance starts in 2008 and persists throughout the rest of the test period (with a slight drop in 2011 and 2012). Again, we compute several performance measures for the strategies and compare them in Table 3. Thereby, we again focus on the risk/return characteristics of the strategies.

The sentiment strategy outperforms all benchmarks in terms of annualized return, Sharpe ratio, Omega ratio, and CER, and performs slightly worse in terms of the other performance measures. Again, the sentiment strategy exhibits a higher turnover than the benchmarks. The resulting annualized returns and Sharpe ratio are more than twice those of the best benchmark.



**Fig. 5** Cumulative returns over time (test period). Cumulative returns of the sentiment strategy and the five benchmark strategies in the test period. The sentiment strategy uses lag  $\tau = 12$ , a uniform view distribution with  $U(-0.005; 0.03)$ , and the ‘negative pick matrix’

**Table 3** Performance measures for the test period

Strategy	AR	SR	DD	OR	VaR	CVaR	TO	CER
Sentiment	0.1353	0.8884	0.0300	2.0229	0.0702	0.1039	0.3664	0.0106
Initial market	0.0467	0.3030	0.0312	1.3631	0.0634	0.1139	0.2515	0.0038
Dynamic 1/n	0.0480	0.3269	0.0316	1.3715	0.0707	0.1036	0.0393	0.0039
Buy-and-Hold 1/n	0.0428	0.3193	0.0287	1.3668	0.0507	0.0969	0.0505	0.0035
Dyn. market cap	0.0328	0.1912	0.0377	1.2484	0.0968	0.1282	0.0470	0.0027
Markowitz	0.0438	0.2764	0.0322	1.3410	0.0623	0.1193	0.1011	0.0036

Performance measures for the sentiment strategy with the ‘negative pick matrix’ for lag  $\tau = 12$  and a uniform view distribution, as well as for the five alternative strategies for the test period. For an explanation of the performance measures, see Sect. 3.3. For abbreviations of the performance measures, see Table 2. VaR and CVaR are displayed as absolute values. Table 4 shows the performance measures for the test period when considering transaction costs

To summarize, the sentiment strategy outperforms the five benchmark strategies in terms of most performance measures.<sup>13</sup> To demonstrate that the outperformance of the sentiment strategy is not generated by pure chance or due to a certain combination of parameters, we subject our research setup to several robustness checks in the next subsection.

### 4.3 Robustness checks

The starting point for each of the robustness checks is the sentiment strategy with the parameter specification pointed out in Sect. 4.1.

<sup>13</sup> The only drawback is that the sentiment strategy exhibits considerably higher portfolio turnover. However, considering transaction costs changes the results only slightly (see Table 4).

**Table 4** Performance measures for the test period when considering transaction costs

Strategy	AR	SR	DD	OR	VaR	CVaR	TO	CER
Sentiment	0.1343	0.8819	0.0300	2.0146	0.0703	0.1040	0.3664	0.0106
Initial market	0.0460	0.2989	0.0312	1.3587	0.0634	0.1139	0.2515	0.0038
Dynamic $1/n$	0.0479	0.3263	0.0316	1.3708	0.0708	0.1036	0.0393	0.0039
Buy-and-Hold $1/n$	0.0426	0.3184	0.0287	1.3658	0.0507	0.0969	0.0505	0.0035
Dyn. market cap	0.0327	0.1905	0.0377	1.2477	0.0968	0.1283	0.0470	0.0027
Markowitz	0.0436	0.2749	0.0322	1.3393	0.0623	0.1193	0.1011	0.0036

Performance measures for the sentiment strategy with the ‘negative pick matrix’ for lag  $\tau = 12$  and a uniform view distribution, as well as for the five alternative strategies for the test period when considering transaction costs of 2 bp per transaction of the turnover. For an explanation of the performance measures, see Sect. 3.3. For abbreviations of the performance measures, see Table 2. VaR and CVaR are displayed as absolute values

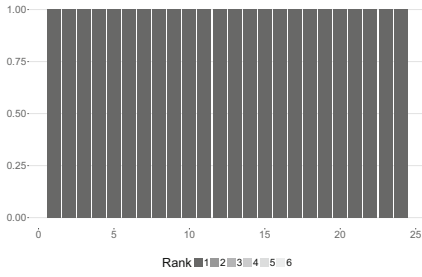
**Fig. 6** Ratios of performance measures in the test period. Ratios of the six performance measures of the sentiment strategy and the corresponding best strategy for lags  $\tau = 1, \dots, 24$  in the test period. The ratio is computed as  $PM_{\text{sent}}/PM_{\text{best}}$ , with  $PM_{\text{sent}}$  the value of the performance measure for the sentiment strategy and  $PM_{\text{best}}$  the value for the best performing strategy. If the sentiment strategy is the best performing strategy, the ratio takes the value 1. For negative Sharpe ratios of the sentiment strategy, the ratio takes the value 0. The colors of the bars indicate the rank of the sentiment strategy among all strategies in terms of that performance measure

The alternative parameterizations of the sentiment strategy serve different purposes in the training than in the test period. In the training period, they are necessary to identify the best parameterization in terms of several performance measures, whereas, in the test period, the results of the alternative parameterizations are used as robustness checks. To present an overview of the results, in this section, the results of the alternative parameterizations are not only shown for the test but also for the training period.

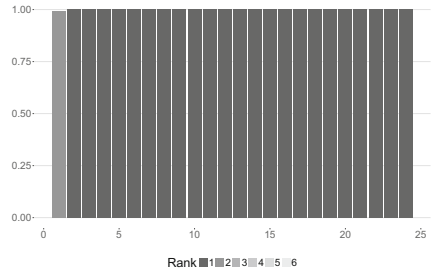
First, the robustness of the lag  $\tau$  is investigated. Recall that the sentiment literature suggests that sentiment does not have an immediate but rather a medium-term effect on stock markets (see Sect. 1). In Sects. 4.1 and 4.2, the results for the lag  $\tau = 12$  are presented. We vary the lag from 1 to 24 months to check if the results depend on a specific lag. The results are summarized in Fig. 6.

For the sake of better comparability, a normalization ratio is computed:  $PM_{\text{sent}}/PM_{\text{best}}$ , with  $PM_{\text{sent}}$  the value of the performance measure of the sentiment strategy and  $PM_{\text{best}}$  its value for the best performing strategy. The ratio takes the value 1 if the sentiment strategy is the best performing strategy for the respective performance measure and lag  $\tau$ . For medium-term lags, starting at about 6 months, the sentiment strategy always outperforms the best benchmark strategy in terms of annualized return, Sharpe ratio, Omega ratio, and CER. For shorter lags, the ratio is usually unity. However, it should be noted that the absolute differences between the Sharpe ratios of the sentiment strategy and the second best strategy are small for lags smaller than 6 months.<sup>14</sup> Furthermore, in the training period, ratios are clearly

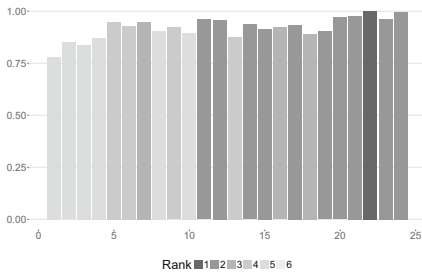
<sup>14</sup> To conserve space, the results are not reported.



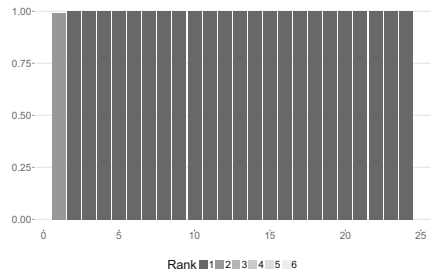
(a) Ratio of returns



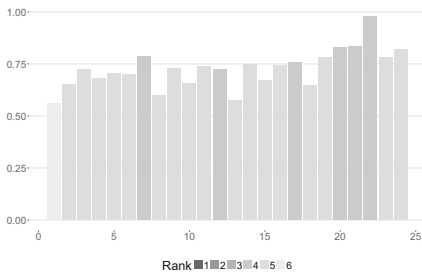
(b) Ratio of Sharpe ratios



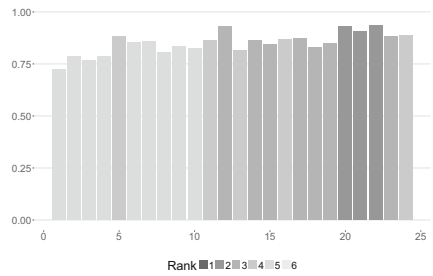
(c) Ratio of Downside Deviations



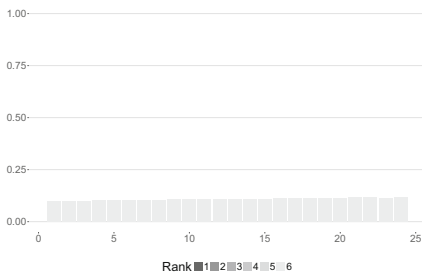
(d) Ratio of Omegas



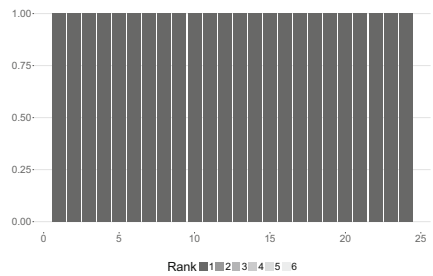
(e) Ratio of VaRs



(f) Ratio of CVaRs



(g) Ratio of Turnovers



(h) Ratio of CERs

smaller than 1 for short lags (see, e.g., subfigure (a) of Fig. 10). Therefore, a clear pattern of the negative sentiment of sentiment is observable only in the medium term.

Regarding the DD, CVar, and VaR, the sentiment strategy performs only slightly worse than the benchmarks as the ratio takes values close to 1 for medium lags. This is in line with the previous literature on the relation between sentiment and returns, as our results confirm the previous findings that the information contained in investor sentiment cannot be exploited in the short term but can when using medium-term lags (see Brown and Cliff 2005).

Another parameter that requires a robustness check is the number of observations used to estimate the initial market distribution. In the previous sections, 10 years of returns data are used. On one hand, we extend this period to 11 years and, on the other hand, shorten it to 9 years. The results for the ratio of the Sharpe ratios of the sentiment strategy and the best benchmark strategy for different lags  $\tau = 1, \dots, 24$  are shown in Fig. 7.

It can be seen that both the extension and the reduction of the number of observations lead to comparable performances of the sentiment strategy.

Next, we check the robustness of the results concerning the choice of the view distribution. We, therefore, vary the parameters of the view distribution along a grid.<sup>15</sup> For the uniform distribution, the lower limit is varied in the interval  $[-0.04; 0]$ , whereas the upper limit is varied in  $[0.0025; 0.0425]$ , each in steps of 0.0025. For the normal distribution, we let  $\mu$  take values in  $[-0.02; 0.02]$  and  $\sigma$  in  $[0.0025; 0.0425]$ , again in steps of 0.0025. We show the results as 3D plots (Fig. 8) and as heat maps (Fig. 9).

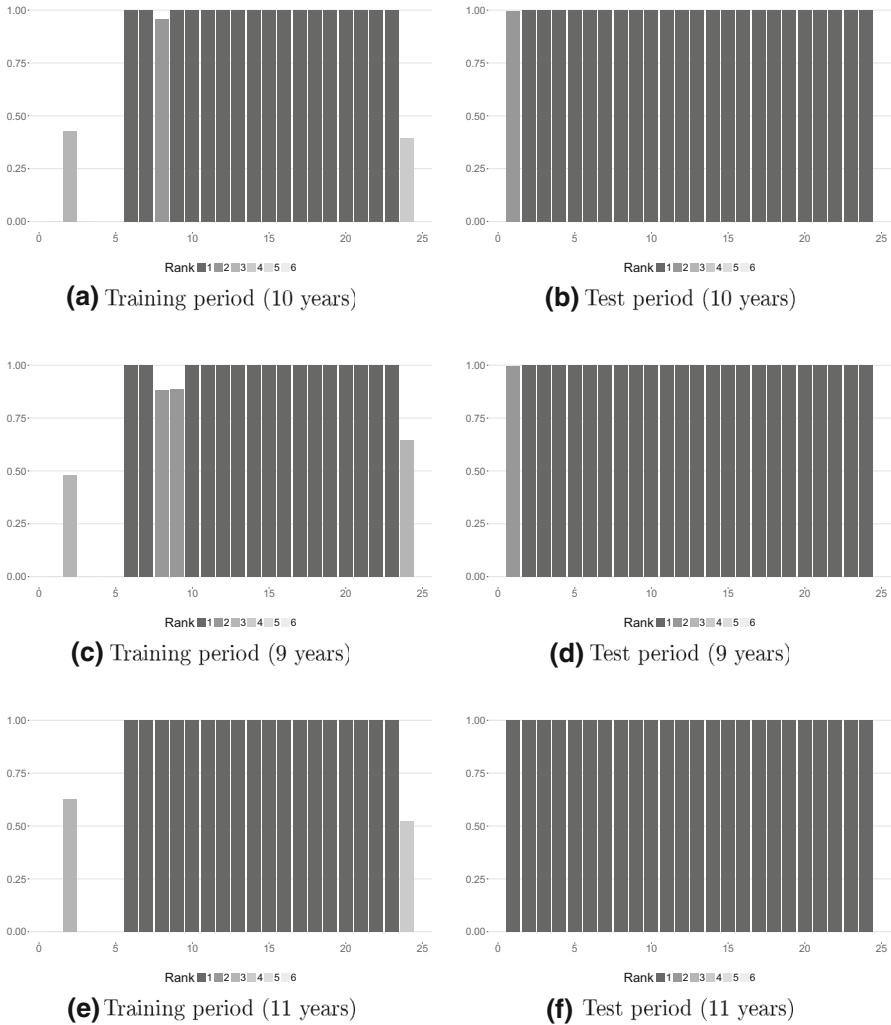
Starting with the 3D plots of the test period in Fig. 8b, d, for the normal distribution, the sentiment strategy exhibits positive Sharpe ratios for positive values of  $\mu$ . Thereby, the value of  $\sigma$  only plays a minor role for the performance of the sentiment strategy. For the uniform distribution, the performance of the sentiment strategy follows a similar pattern. The sentiment strategy generates positive Sharpe ratios for parameter combinations that result in a positive mean for the view distribution. This is in line with our theory, as we expect the return being generated by the specified view to be positive.

In the training period, the sentiment strategy also generates slightly positive Sharpe ratios for some parameter combinations with a slightly negative mean. For parameter combinations with a positive mean, the results are similar to those of the test period. As the 3D plots regard only the value of the Sharpe ratio, we also show heat maps containing the rank of the sentiment strategy among all strategies.

In addition to Fig. 8, the results in Fig. 9 show that the sentiment strategy does not only generate positive Sharpe ratios for a positive mean of the underlying view distribution but also exhibits the highest Sharpe ratio among all strategies. This is indicated by the color white for the respective parameter combination. The results are consistent with the results of Fig. 8. In the training period, the sentiment strategy also is the best performing strategy for some parameter combinations with a slightly negative mean. A reason for this may be that, from each view distribution,

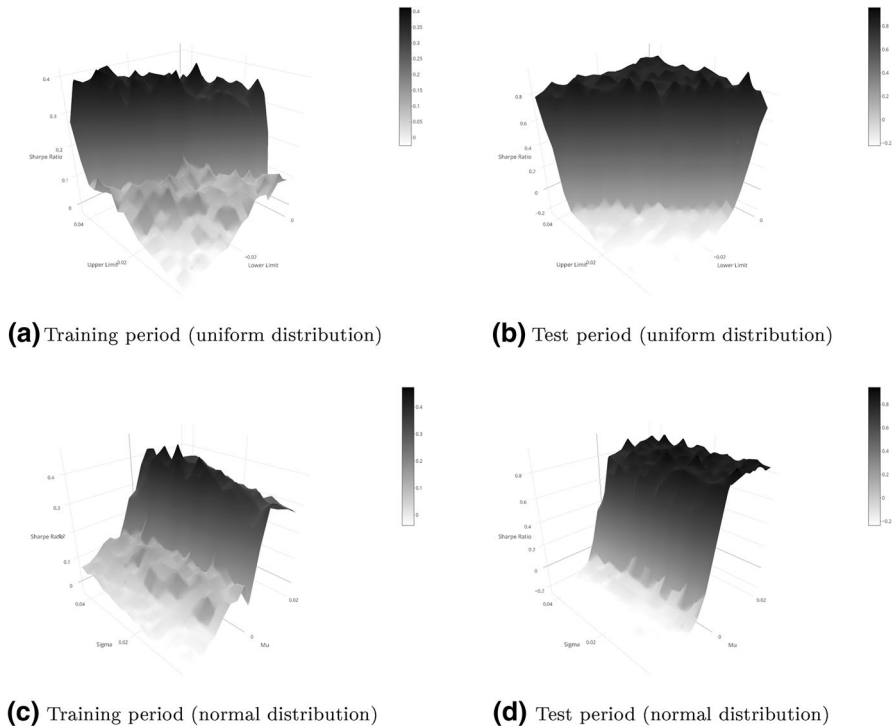
<sup>15</sup> We thank an anonymous referee for suggesting this procedure.





**Fig. 7** Ratios of Sharpe ratios for different numbers of observations for the estimation of the initial market distribution. Ratios of the Sharpe ratios of the sentiment strategy and the best strategy for lags  $\tau = 1, \dots, 24$  for different numbers of observations for the estimation of the initial market distribution in the training (left column) and the test (right column) period. The ratio is computed as  $SR_{sent}/SR_{best}$ , with  $SR_{sent}$  the value of the performance measure for the sentiment strategy and  $SR_{best}$  the value for the best performing strategy. If the sentiment strategy is the best performing strategy, the ratio takes the value 1. For negative Sharpe ratios of the sentiment strategy, the ratio takes the value 0. The colors of the bars indicate the rank of the sentiment strategy among all strategies

simulations are generated. This means that, even if the mean of the underlying view distribution is positive, the simulations can contain a large number of positive values. The most prominent result of Figs. 8 and 9 is that, for parameter combinations with positive mean, the sentiment strategy not only exhibits positive Sharpe ratios but also is the best performing strategy among all strategies. This

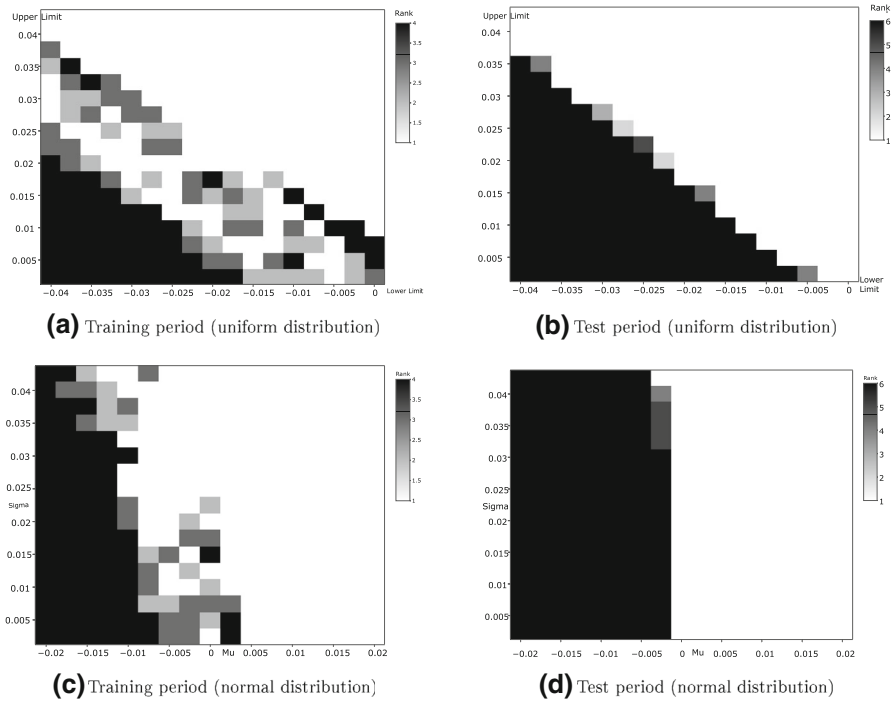


**Fig. 8** 3D plots for varying combinations of the parameters for the view distribution. 3D plots for different combinations of parameters for the view distribution in the training (left column) and the test (right column) period. For the uniform distribution, the lower and upper limits are plotted on the horizontal axes, whereas, for the normal distribution,  $\mu$  and  $\sigma$  are plotted on the horizontal axes. For both distributions, the Sharpe ratio of the corresponding sentiment strategy is plotted on the vertical axis. To interactively rotate the plots, see the following online graphics: - <https://plot.ly/~pl0tn7c/18/> (uniform distr., training period) - <https://plot.ly/~pl0tn7c/21/> (uniform distr., test period) - <https://plot.ly/~pl0tn7c/12/> (normal distr., training period) - <https://plot.ly/~pl0tn7c/15/> (normal distr., test period)

means that the choice of a view distribution with a positive mean is not only reasonable, because the corresponding pick is expected to generate a positive return by incorporating the information contained in sentiment, but also empirically generates a higher Sharpe ratio. The results show that the superior performance of the sentiment strategy does not depend on a specific parameterization of the view distribution and remains robust for alternative specifications.

The next parameter that is subject to a robustness check is the pick matrix. Apart from the intuitive ‘negative pick matrix’, we apply a ‘weighted negative pick matrix’ and a ‘positive pick matrix’. The former is an extension of the ‘negative pick matrix’ in which the entries of the pick matrix are weighted by the magnitude of the sentiment measures. In the latter, a positive sentiment index leads to a positive value in the pick matrix for the respective market. For the results of the application of these pick matrices, see Fig. 10.

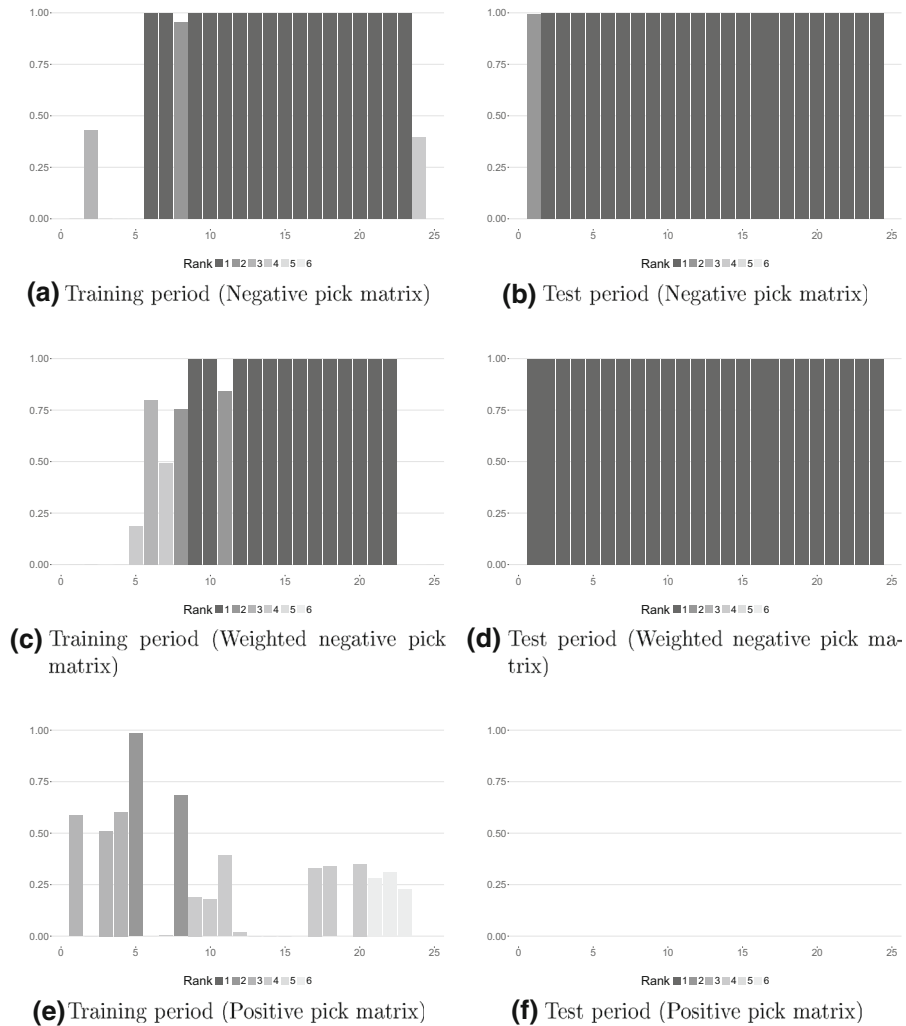
In the training period as well as in the test period, the ‘weighted negative pick matrix’ performs about equally well as the ‘negative pick matrix’. The application



**Fig. 9** Heat maps for varying combinations of parameters for the view distribution. Heat maps for different combinations of parameters of the view distribution in the training (left column) and the test (right column) period. For the uniform distribution (a and b), the lower limit of the distribution is plotted on the x-axis and the upper limit on the y-axis. For the normal distribution (c and d),  $\mu$  is plotted on the x-axis and  $\sigma$  on the y-axis. Regarding the uniform distribution, the lower limit is varied in the interval  $[-0.04; 0]$ , whereas the upper limit is varied in  $[0; 0.04]$ , each in steps of 0.005. For the normal distribution, we let  $\mu$  take values in  $[-0.02; 0.02]$  and  $\sigma$  in  $[0; 0.04]$ , again in steps of 0.005. The color indicates the rank of the sentiment strategy’s Sharpe ratio among all strategies for the respective parameter combination. Note that the color white indicates that the sentiment is the best strategy. In the training period, black indicates that the sentiment strategy is the fourth best performing strategy (as for all parameter combinations, the worst rank of sentiment strategy among all strategy is fourth), whereas, in the test period, black indicates that the sentiment strategy is the worst performing strategy

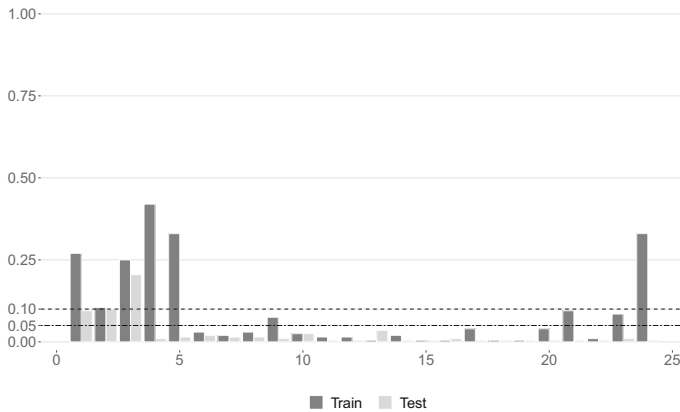
of the ‘positive pick matrix’ leads to the sentiment strategy mostly being the worst or second worst performing strategy. This means that the sentiment measures can only be exploited for portfolio optimization if positive (negative) sentiment leads to a negative (positive) value in the pick matrix. This approach is consistent with the previous findings in the sentiment literature and its results seem very sound.

The last robustness check concerns the question if the superior performance of the sentiment strategy can be attributed to the information contained in the sentiment measures or to the COP method. Therefore, a bootstrap procedure is applied to the sentiment data  $S$ . The intuition of this procedure is to apply the COP method using randomly permuted sentiment information. If the resulting time series still leads to outperformance, this could be attributed to the COP method but not to sentiment. Each row  $S_t$  contains the sentiment measures for the four markets at period  $t$ . The  $k$ th bootstrap sentiment time series  $S^{Bk}$  is generated by drawing rows



**Fig. 10** Ratios of Sharpe ratios for different specifications of the pick matrix. Ratios of the Sharpe ratios of the sentiment strategy and the best strategy for lags  $\tau = 1, \dots, 24$  for different specifications of the pick matrix in the training (left column) and the test (right column) period. The ratio is computed as  $SR_{sent}/SR_{best}$ , with  $SR_{sent}$  the value of the respective performance measure for the sentiment strategy and  $SR_{best}$  the value for the best performing strategy. If the sentiment strategy is the best performing strategy, the ratio takes the value 1. For negative Sharpe ratios of the sentiment strategy, the ratio takes the value 0. The colors of the bars indicate the rank of the sentiment strategy among all strategies

from  $\mathbf{S}$  with replacement, where  $\mathbf{S}^{B_k}$  and  $\mathbf{S}$  correspond in their number of observations. This bootstrap is repeated  $K = 200$  times. For each of the 200 bootstrap sentiment time series, the sentiment strategy is applied in the training and the test period separately. For each lag  $\tau$ , we store the resulting Sharpe ratios and calculate the percentage of Sharpe ratios that are higher than the Sharpe ratio of the



**Fig. 11** Empirical  $p$  values of the sentiment bootstrap. Empirical  $p$  values of the bootstrap procedure for the training (black) and the test (gray) period and lags  $\tau = 1, \dots, 24$ . The empirical  $p$  values indicate the number of bootstrapped time series that generate a higher Sharpe ratio than the sentiment time series. The sentiment time series as well as the 200 bootstrapped time series are incorporated in the COP framework separately

sentiment strategy that uses **S**. We refer to these percentages as the ‘empirical  $p$  values’. The results are summarized in Fig. 11.

It can be seen that, for medium-term lags, the empirical  $p$  values are below 10% or even 5% (with one exception in the training period). This means that applying the COP method to bootstrap data that carries little if any significant information regarding future portfolio allocation leads to only a very small number of higher Sharpe ratios. This number lies below the 5%, respectively, 10% that can be expected by chance. We can, therefore, conclude that sentiment contains a significant information for portfolio optimization and that the COP method is a sensible framework for the incorporation of sentiment into portfolio optimization.

Overall, the performance of the sentiment strategy is robust to changes in several parameters of the applied portfolio optimization procedure. Figures 8 and 9 show that the superior performance of the sentiment strategy does not depend on the specification of the view distribution, but, as Fig. 10 shows, on the choice of a pick matrix that is theoretically well founded. The following section concludes the paper.

## 5 Conclusion

The influence of investor sentiment on returns and volatility is well documented and investigated in several theoretical and empirical studies (see Sect. 1). We build on these studies by exploiting the information carried by investor sentiment for portfolio optimization. The sentiment measures are computed via Principle Component Analysis as in Baker and Wurgler (2006). However, we extend the data set of Baker and Wurgler (2006) and Baker et al. (2012) and compute the sentiment measures for four international markets on a monthly basis. The sentiment measures are integrated into a portfolio optimization procedure by the

application of the Copula Opinion Pooling method introduced by Meucci (2006b). As, therefore, several parameters have to be specified, we use the period from July 1993 to July 2004 for the optimization of the parameters in terms of eight performance and downside risk measures. The consequent parameter specification is tested on the period from August 2004 to September 2015. Regarding sentiment and portfolios, several studies have used investor sentiment to sort assets into different portfolios with respect to their exposure to sentiment. We extend these approaches by directly integrating investor sentiment into the portfolio optimization.

Regarding the research question stated in Sect. 2, we come to the following conclusions.

$C_1$  : The choice of the pick matrix that integrates the view of a ‘crowd of investors’ into the portfolio optimization is theoretically well founded. The negative pick matrix, which models the reversal effect of sentiment-induced mispricing, exhibits the best performance of all considered sentiment strategies (i.e., versions of the pick matrix). Furthermore, the ‘weighted negative pick matrix’ shows a similar performance. Therefore, we conclude that a pick matrix modeling the negative effect of sentiment on returns should be used, as the application of a ‘positive pick matrix’ leads to a much worse performance.

$C_2$  : The sentiment-based strategy outperforms the benchmark strategies in terms of most of the performance and downside risk measures in the training as well as the test period. As the sentiment-based strategy outperforms the initial market strategy that does not include sentiment information, this provides further empirical evidence that sentiment contains useful information for portfolio optimization. This is strengthened by the fact that the empirical  $p$  value used to compare the outcome of our sentiment strategy to a strategy that applies the framework to randomly bootstrapped sentiment time series is below 5% or 10% for most medium-term horizons.

$C_3$  : Using a medium-term lag of the sentiment measures to model the medium-term reversal effect of investor sentiment, the outperformance of the sentiment strategy shows stable results.

Regarding our three research questions, several robustness checks demonstrate that, for different parameter specifications, our results remain qualitatively the same (see Sect. 4.3).

Regarding the ‘sentiment performance premium’ of our strategies, it could be interesting to see whether this could be attributed to some known risk and asset pricing factors, e.g., the DGTW characteristics of Daniel et al. (1997), Fama–French factors (Fama and French (2015)), liquidity factors (e.g., that of Pástor and Stambaugh (2003)), or volatility innovations (Ang et al. (2006)).<sup>16</sup> Another area for future research is the incorporation into the proposed framework of sentiment measures other than the ones used in this study. Furthermore, not only stock indexes, but also single stocks, which are more subject to the influence of sentiment, could be used to form portfolios. Thereby, it could be tested whether the

---

<sup>16</sup> We thank an anonymous referee for pointing out this limitation and highlighting this area for future research.

performance of the strategy could be improved by segmenting the market using hard-to-value and difficult-to-arbitrage portfolios. We leave this investigation for further research, as our aim was foremost to test the sentiment effect on a market level. In addition, it might be fruitful to use the sentiment measures at hand in combination with alternative methods that can even incorporate sentiment as an optimization criterion instead of using it as an input variable.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Ang, A., R.J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *The Journal of Finance* 61 (1): 259–299.
- Antweiler, W., and M.Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance* 59 (3): 1259–1294.
- Azzalini, A., and A. Capitanio. 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (2): 367–389.
- Baker, M., and J. Wurgler. 2006. Investor sentiment and the cross-section of stock returns. *The Journal of Finance* 61 (4): 1645–1680.
- Baker, M., and J. Wurgler. 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives* 21 (2): 129–151.
- Baker, M., J. Wurgler, and Y. Yuan. 2012. Global, local, and contagious investor sentiment. *Journal of Financial Economics* 104 (2): 272–287.
- Black, F., and R. Litterman. 1992. Global portfolio optimization. *Financial Analysts Journal* 48 (5): 28–43.
- Brown, G.W., and M.T. Cliff. 2005. Investor sentiment and asset valuation. *The Journal of Business* 78 (2): 405–440.
- Bucher, M. (2017). Investor attention and sentiment: Risk or anomaly? Working paper.
- Clarke, R.G., and M. Statman. 1998. Bullish or bearish? *Financial Analysts Journal* 54: 63–72.
- Da, Z., J. Engelberg, and P. Gao. 2015. The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies* 28 (1): 1–32.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance* 52 (3): 1035–1058.
- De Long, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98: 703–738.
- DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies* 22 (5): 1915–1953.
- Fama, E.F., and K.R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116 (1): 1–22.
- Fu, C., G. Jacoby, and Y. Wang. 2015. Investor sentiment and portfolio selection. *Finance Research Letters* 15: 266–273.
- Gochez, F., Chandler-Mant, R., Jin, S. & Xie, J. (2015). *BLCOP: Black-Litterman and Copula Opinion Pooling Frameworks*. URL <https://CRAN.R-project.org/package=BLCOP>. R package version 0.3.1.
- Keating, C., and W.F. Shadwick. 2002. A universal performance measure. *Journal of Performance Measurement* 6 (3): 59–84.
- Markowitz, H. 1952. Portfolio selection. *The Journal of Finance* 7 (1): 77–91.
- Meucci, A. 2006a. Beyond Black–Litterman in practice: A five-step recipe to input view on non-normal markets. *Risk Magazine* 19: 114–119.
- Meucci, A. 2006b. Beyond Black–Litterman: Views on non-normal markets. *Risk Magazine* 19: 87–92.

- Neely, C.J., D.E. Rapach, J. Tu, and G. Zhou. 2014. Forecasting the equity risk premium: The role of technical indicators. *Management Science* 60 (7): 1772–1791.
- Pástor, L., and R.F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111 (3): 642–685.
- Schmeling, M. 2007. Institutional and individual sentiment: Smart money and noise trader risk? *International Journal of Forecasting* 23 (1): 127–145.
- Schmeling, M. 2009. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance* 16 (3): 394–408.
- Schneller, D., S. Heiden, M. Heiden, and A. Hamid. 2018. Home is where you know your volatility - Local investor sentiment and stock market volatility. *German Economic Review* 19: 209–236.
- Sibley, S.E., Y. Wang, Y. Xing, and X. Zhang. 2016. The information content of the sentiment index. *Journal of Banking & Finance* 62: 164–179.
- Simonian, J. 2014. Copula-opinion pooling with complex opinions. *Quantitative Finance* 14 (6): 941–946. <https://doi.org/10.1080/14697688.2013.812234>.
- Solt, M.E., and M. Statman. 1988. How useful is the sentiment index? *Financial Analysts Journal* 44: 45–55.
- Sortino, F.A., and L.N. Price. 1994. Performance measurement in a downside risk framework. *The Journal of Investing* 3 (3): 59–64.
- Stambaugh, R.F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104 (2): 288–302.
- Stambaugh, R.F., J. Yu, and Y. Yuan. 2014. The long of it: Odds that investor sentiment spuriously predicts anomaly returns. *Journal of Financial Economics* 114 (3): 613–619.
- Stambaugh, R.F., and Y. Yuan. 2016. Mispricing factors. *The Review of Financial Studies* 30 (4): 1270–1315.
- Stein, M., R. Füss, and W. Drobetz. 2009. Fixed-income portfolio allocation including hedge fund strategies: A copula opinion pooling approach. *The Journal of Fixed Income* 18 (4): 78–91.
- Sul, H.K., A.R. Dennis, and L.I. Yuan. 2017. Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences* 48 (3): 454–488.
- Tetlock, P.C. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62 (3): 1139–1168.
- Yu, J., and Y. Yuan. 2011. Investor sentiment and the mean-variance relation. *Journal of Financial Economics* 100 (2): 367–381.