Investor Sentiment Purged: A Powerful Predictor in the Cross-Section of Stock Returns

Liya Chu
Singapore Management University

Qianqian Du
University of Stavanger

Jun Tu*
Singapore Management University

First Version: Oct 2015

Current Version: June 2016

*Corresponding author. Send correspondence to Jun Tu, Lee Kong Chian School of Business, Singapore Management University, Singapore 178899; telephone: (+65) 6828 0764. E-mail: tujun@smu.edu.sg. We are grateful to Doron Avramov, Paul Pengjie Gao, Phil Dybvig, Xiaoyan Zhang and seminar participants at Singapore Management University for helpful comments. Jun Tu acknowledges that the study was funded through a research grant from Sim Kee Boon Institute for Financial Economics. The usual disclaimer applies. We also thank Wurgler for kindly providing data on his web pages.

Investor Sentiment Purged: A Powerful Predictor in the Cross-Section of Stock Returns

Abstract

Numerous studies treat the Baker and Wurgler (2006) sentiment index as a behavioral variable. However, these studies could be misleading given that the six proxies used to construct the Baker and Wurgler sentiment index are closely related to overall fundamental business environment. In this paper, we remove fundamental information thoroughly from Baker and Wurgler sentiment index to obtain a new purged sentiment index. Empirically, we find that our purged investor sentiment index has a similar or greater power in predicting the stock returns cross-sectionally compared with the original Baker and Wurgler sentiment index. Our study indicates that the Baker and Wurgler sentiment index captures behavioral driven investor sentiment component. Therefore, it seems fine for many studies to adopt the Baker and Wurgler sentiment index with a behavioral interpretation.

JEL classifications: C53, G12, G14

Keywords: Investor Sentiment, Asset Pricing, Return Predictability, Cash Flow

1. Introduction

Investor sentiment may affect asset prices due to the well-known psychological fact that people with high (low) sentiment tend to make overly optimistic (pessimistic) judgments and choices (e.g., Keynes (1936), Shiller (1981, 2000), Neal and Weatley (1998), Hirshleifer (2001), Antoniou, Doukas and Subrahmanyam (2013)). De Long, Shleifer, Summers, and Waldmann (1990), Shleifer and Vishny (1997), Shleifer (2000), among others, provide theoretical explanations why sentiment can cause asset price to deviate from its fundamental in the presence of limits of arbitrage even when informed traders recognize the opportunity. Empirically, however, it is rather challenging to test the importance of investor sentiment, since it is not directly observable.[1] Baker and Wurgler (2006) construct a novel investor sentiment index that aggregates the information from six proxies, and find that high investor sentiment strongly predicts lower returns in the cross-section, such as stocks that are speculative and hard to arbitrage.[2]

Since the creation of the influential Baker and Wurgler sentiment index (BW index hereafter), numerous papers adopt it for extensive applications, including risk-return trade-off in Yu and Yuan (2011), stock price response to earnings news in Mian and Sankaraguruswamy (2012), asset pricing anomalies in Stambaugh, Yu and Yuan (2012), analysts' forecast error in Hribar and McInnis (2012), institutional/individual investors' demand shocks in Devault, Sias and Starks (2016), the slope of security market line in Antoniou, Doukas and Subrahmanyam (2015), investments of public and private firms in Badertscher, Shanthikumar and Teoh (2016), and hedge fund returns in Chen, Han and Pan (2016), etc. These papers usually treat the BW sentiment index as a behavioral variable and interpret their empirical results as consistent with the idea that investors sentiment, unrelated to systematic risks, drives prices and returns in the market. However, most of the un-

[1] Extant studies use a broad range of variables measuring sentiment, such as survey-based approach (Brown and Cliff (2004), Lemmon and Portniaguina (2006)), search-based sentiment index (Da, Engelberg and Gao (2015)) and mood-related factors (Kaplanski, Levy, Veld and Veld-Merkoulova (2015)). Nevertheless, some of the search/survey-based measures are only available from recent years, and some survey questions may not be answered carefully or truthfully.

[2] There are many other studies employing sentiment measures based on market data. However, most of these studies use single proxy, such as retail investor trades, mutual fund flows, closed-end fund discounts and net equity issues (Kumar and Lee (2006), Ben-Rephael, Kandel and Wohl (2012), Lee, Shleifer, and Thaler (1991), Swaminathan (1996), Baker and Wurgler (2000)).

derlying six proxies used to construct the BW sentiment index also tend to be closely related to the overall business environment. For instance, more IPO numbers can be due to high sentiment but also can be due to the higher demand of investment when the economy is booming.[3] We further summarize fundamental information content from a broad range of more than one hundred economic variables to obtain 14 representative fundamental variables. The regression of the BW sentiment index on these 14 representative fundamental variables has an adjusted R^2 of about 62%, indicating that the BW investor sentiment index is significantly driven by economic fundamental.[4] If the BW sentiment index contains significant amount of fundamental information about rational risk premia or expectations of future cash flows, then numerous studies adopting BW sentiment index as a behavioral variable could be misleading.

In this paper, we take a crucial check on whether BW sentiment index can preserve its cross-sectional predicting power as a sentiment proxy after removing fundamental information as thoroughly as possible. To remove fundamental information from BW index, we regress each one of the six proxies of the BW sentiment index on these 14 representative fundamental variables. Then we exploit the residual of the six sentiment proxies in an efficient manner by using the partial least squares (PLS) method to obtain a new purged index. Empirically, we find that the purged sentiment index (IS-P) can predict the cross-sectional stock returns remarkably well. For instance, in multivariate regressions taking the Fama French factors and Carhart's momentum factor as control variables, IS-P demonstrates significant predictive ability in 13 out of the 16 long-short portfolio returns in Baker and Wurgler (2006) while the original BW sentiment index is statistically significant in predicting 11 out of the 16 long-short portfolio returns.[5] The increase of significant number (from 11 to 13) is due to that IS-P has significant predictive power for the long-short portfolios based on tangibility characteristics (PPE/A portfolio and RD/A portfolio). Specifically, high

[3]Although Baker and Wurgler (2006) has tried taking out a few economic variables, the so-called orthogonal Baker and Wurgler (2006) index $BW^{\perp}$ after taking out a small set of economic variables seems not purging the fundamental information content much given that it has a 0.97 correlation with the original Baker and Wurgler (2006) index.

[4]By contrast, if we regress the BW sentiment index on the set of economic variables in Baker and Wurgler (2006), the adjusted R^2 is approximately 2%.

[5]The orthogonal Baker and Wurgler (2006) index $BW^{\perp}$ is significant in predicting 10 out of the 16 long-short portfolio returns.

IS-P is associated with relatively low future returns of firms with lower tangible assets. This is consistent with the theoretical prediction that firms with less tangible assets are more difficult to value, and those stocks are more likely to be affected by fluctuations in the propensity to speculate. In contrast, the tangibility characteristics fail to exhibit conditional effects based on the original BW sentiment index, which may be due to the possibility that BW sentiment is clouded by too much fundamental information.[6] Moreover, the signs of the coefficients on the various firm characteristics are consistent with the signs documented by Baker and Wurgler (2006). Overall, the results suggest that the predictive ability of IS-P is comparable to or better than the original BW sentiment index containing a large amount of fundamental information.

Furthermore, we find that IS-P is insignificantly correlated with next period macroeconomic activity or business cycle peak-trough dummy, indicating that the purged sentiment index contain little fundamental information. In addition, we find that after removing the 14 representative fundamental variables directly from the BW sentiment index, the residual part of BW sentiment index merely predicts one out of the 16 long-short portfolio returns, and none of the long-leg or short-leg returns. In robustness check, when we consider alternative sets of fundamental variables, our purged sentiment index constructed from PLS consistently present strong predictability for the 16 long-short portfolio returns while the residual part of BW sentiment index has little predictive power. These findings suggest that our purged sentiment index outperforms measures based on other methods in efficiently extracting investor sentiment relevant information, consistent with previous literatures (e.g., Huang, Jiang, Tu and Zhou (2015)).

We provide validation tests for the purged sentiment index. First validation test involves earnings announcement returns. Baker and Wurgler (2006) find that earnings announcement returns are lower after high investor sentiment. Since investors are more likely to suffer errors in valuation for stocks which are speculative and hard to arbitrage, we expect that earnings announcement returns should be inversely related to the purged sentiment index for speculative stocks. The results are consistent with our expectation. In the second validation test, we connect the IS-P with mutual

[6] We also use the orthogonal Baker and Wurgler (2006) index $BW^{\perp}$ and find similar result to that of the original BW sentiment index.

fund flow measured as the net exchanges of equity funds shifting between bond funds and equity funds. The mutual fund flow reflects investor sentiment towards stock market (Ben-Rephael, Kandel and Wohl (2012)). We find that IS-P significantly predicts mutual fund flow while BW sentiment index fails. The third validation test involves mispricing component in Tobin's Q. We find that IS-P captures mispricing information in Tobin's Q and present better predictability for the portfolio returns than mispricing component in Tobin's Q.

In addition, it is of interest to investigate the economic driving force of the predictability of the purged investor sentiment, i.e., whether the predictive power of IS-P stems from time variations in cash flows or discount rates. We find that the purged investor sentiment index significantly forecasts future dividend growth, which is a standard cash flow proxy, but insignificantly forecast future dividend price ratio, which is a common proxy of discount rates. The evidences support that the cash flow channel is the source for predictability. Furthermore, the ability of the purged sentiment index to forecast the cross-section of stock returns is positively associated with its ability to forecast the cross-section of future cash flows as well. Hence, our findings are consistent with Baker and Wurgler (2007) that the lower stock return following high investor sentiment periods seems to represent investors' overly optimistic belief about future cash flows that cannot be justified by subsequent economic fundamentals.

We also compare the predictability of the purged sentiment index with other survey-based sentiment measures, including anxious index, consumer sentiment index, individual investor sentiment index and Gallup survey index. Although we find some explanatory power from the alternative sentiment measures, the predictive power of these alternative measures is much weaker than that of IS-P.

Our study contributes to the previous literature by Baker and Wurgler (2006) and many others who employ Baker and Wurglers sentiment index as investor sentiment proxy. Our study displays that the original Baker and Wurgler (2006) sentiment index may indeed capture behavioral component. However, after taking out fundamental information thoroughly and extract the investor sentiment in an efficient way, BW index can still predict cross-sectional stock returns as implied by sentiment theory. Since forecasting and understanding how stock returns vary over time and across assets is one of the central issues in financial research that has implications in both corporate finance and asset pricing (e.g., Spiegel, 2008 and Cochrane, 2011), our study reassure the importance of investor sentiment as a behavioral force adopted by numerous studies.

Moreover, our study seems complementing extant literature focusing on survey-based non-fundamental factors which could alter individuals' mood or feelings, including weather related issues such as sunshine, clouds and temperature (e.g., Saunders (1993), Hirshleifer and Shumway (2003)), seasonal affective disorder arising from autumn and winter depression (Kamstra, Kramer, and Levi (2003)), and sports results or other abrupt events which could trigger investor sentiment (Edmans, Garcia, and Norli (2007), Kaplanski and Levy (2010a, 2010b)). In contrast to these studies, we extend Baker and Wurgler (2006) by constructing a model-based investor sentiment index purged of fundamental information.

The remainder of the paper is organized as follows. Section II discusses the construction of the purged investor sentiment index. Sections III explains the data and provides summary statistics. We present the predictably of sentiment in Section IV and validation tests of purged sentiment index in Section V. We show further analysis in Section VI , and conclude in Section VII.

II. Construction of Purged Investor Sentiment

A. Concerns about the BW index

Baker and Wurgler (2006) initiate an influential sentiment index, which exerts significant effects on cross-sectional returns. Specifically, BW sentiment index is constructed by taking the first principal component of six investor sentiment proxies, i.e., discount rate of close-end fund (CEFD), average NYSE share turnover (TURN), the number of IPOs (NIPO), average first-day return (RIPO), the equity issuance (EQTI), the log difference of market-to-book ratios between dividend payers and nonpayers (PDND). Since the six raw investor sentiment proxies are highly correlated with business environment, Baker and Wurgler (2006) modify the sentiment index by removing variables related with business cycle from each proxy before principal component analysis (PCA). Specifically, they regress each proxy on growth in the industrial production index, growth in consumer durables, nondurable and services, growth in employment and a dummy variable for NBER recessions and use the residuals from the regressions as cleaner sentiment proxies to construct the orthogonal Baker and Wurgler index $BW^{\perp}$.

We have two main concerns about the BW index. Firstly, the six sentiment proxies used to construct BW index are closely correlated with the overall business environment. For instance, the number of IPOs reflects investor sentiment, but is inevitably determined by the economic condition. In Ritter and Welch (2002) survey of the IPO literature, they conclude that market conditions are the most important factor in the decision to go public. Although the $BW^{\perp}$ index has removed a few variables related with business cycle, it possesses a 0.97 correlation with the original BW index. If BW index does not take out much systematic risks but co-moves largely with fundamental information about rational risk premia or expectations of future cash flows, then adopting the sentiment index as a behavioral variable could be misleading.

Secondly, although econometrically the first principal component is the best combination of all the proxies that maximally represents the total variations of the proxies, the first principal component potentially contains a substantial amount of common approximation errors that are irrelevant for forecasting cross-sectional stock returns influenced by investor sentiment. The higher fraction of the irrelevant common approximation errors, the less important role the unobservable sentiment component will play in PCA. Therefore, PCA may fail to forecast cross-sectional stock returns, such as 16 spread portfolios of Baker and Wurgler (2006), even if sentiment does play an important role in affecting the cross-sectional stock returns. We need a better method to disentangle the information in the proxies that influence the expected cross-sectional stock returns from the common approximation errors.

B. Estimation of purged sentiment index

To more effectively extract non-fundamental information from the six individual sentiment proxies, we adopt partial least squares approach (PLS) to generate a purged investor sentiment index IS-P and apply it on forecasting portfolio returns. In this section, we outline our econometric methodology, which is based on Kelly and Pruitt (2013, 2015), and Huang, Jiang, Tu and Zhou (2015).

B.1 Setup

First we establish the environment wherein we use the PLS method. We define the long-short combined portfolio return *ret* as the mean return of 16 firm characteristics based long-short portfolios documented in Baker and Wurgler (2006). *ret* is composed of two parts – the conditional expectation plus an unpredictable shock,

$$ret_{t+1} = E_t(ret_{t+1}) + e_{t+1}, \tag{1}$$

where ret_{t+1} is the long-short combined portfolio return at time t+1 and e_{t+1} is the unpredicted shock.

We assume that conditioning on information at time t, expected long-short combined portfolio return is explained by unobservable investor sentiment S_t,

$$E_t(ret_{t+1}) = \alpha + \beta S_t. \tag{2}$$

We rearrange equation (1) and obtain,

$$ret_{t+1} = \alpha + \beta S_t + e_{t+1}. \tag{3}$$

We denote $X_t = (x_{1,t}, ..., x_{N,t})'$ as an $N \times 1$ vector of purged individual sentiment proxies at period t. Each purged sentiment proxy is estimated as the regression residual of individual sentiment proxy in Baker and Wurgler (2006) on a wide range of economic fundamental variables. Each purged sentiment proxy has a factor structure,

$$x_{i,t} = \eta_{i,0} + \eta_{i,1} * S_t + \eta_{i,2} * E_t + \varepsilon_{i,t}, \qquad \text{for } i = 1, ..., N. \tag{4}$$

where $\eta_{i,1}$ is the sensitivity of sentiment proxies $x_{i,t}$ to the movements in S_t that matters for forecasting portfolio return, E_t is the common approximation error component of X_t that is irrelevant to returns, and $\varepsilon_{i,t}$ is the idiosyncratic noise.

The key advantage of PLS is that it efficiently estimates S_t by extracting the most relevant common component from the investor sentiment proxies according to its covariance with the forecast target. In other words, PLS separates out the information which matters for the future portfolio return from irrelevant information E_t and $\varepsilon_{i,t}$.

B.2 Estimator

Following Huang, Jiang, Tu and Zhou (2015), PLS can be implemented by two stages of OLS regression. In the first stage, for each sentiment proxy $x_{i,t-1}$, which is the residual component of individual investor sentiment proxy after removing fundamental information, we run a time-series regression on a constant and future long-short combined portfolio return ret_t,

$$x_{i,t-1} = \mu_{i,0} + \mu_i * ret_t + \upsilon_{i,t-1}, \qquad \text{for } t = 1, ..., T. \tag{5}$$

The coefficient μ_i captures the sensitivity of each sentiment proxy $x_{i,t-1}$ to investor sentiment S_{t-1} instrumented by future long-short combined portfolio return ret_t. Since the expected component of long-short combined portfolio return is driven by investor sentiment, sentiment proxies are related to the expected long-short combined portfolio return and is uncorrelated with unpredictable return shocks.

In the second-stage, for each time period t, we run a cross-sectional regression of $x_{i,t}$ on the corresponding estimated coefficient $\hat{\mu}_i$ from the first stage,

$$x_{i,t} = c_t + IS\text{-}P_t * \hat{\mu}_i + \omega_{i,t}, \qquad \text{for } i = 1, ..., N, \tag{6}$$

where the coefficient $IS\text{-}P_t$ is the estimated purged sentiment index.

In summary, the first-stage coefficient estimates map the purged sentiment proxies to the forecast target that is assumed to be driven by the unobservable investor sentiment, while second-stage regression use this map to back out estimates of the unobservable investor sentiment at each point in time.

III. Data and summary statistics

A. BW index and purged sentiment proxies

We obtain BW index, $BW^{\perp}$ index and six sentiment proxies used to construct BW index from Wurgler's website[7]. The six individual sentiment proxies are:

- *Close-end fund discount rate*, CEFD: value-weighted average difference between the net asset values of closed-end stock mutual fund shares and their market prices;
- *Share turnover*, TURN: log of the raw turnover ratio detrended by the past 5-year average, where raw turnover ratio is the ratio of reported share volume to average shares listed from the NYSE Fact Book;
- *Number of IPOs*, NIPO: monthly number of initial public offerings;
- *First-day returns of IPOs*, RIPO: monthly average first-day returns of initial public offerings;
- *Dividend premium*, PDND: log difference of the value-weighted average market-to-book ratios of dividend payers and nonpayers; and
- *Equity share in new issues*, EQTI: gross monthly equity issuance divided by gross monthly equity plus debt issuance.

[7] The data are available on the website: http://people.stern.nyu.edu/jwurgler/.

Due to the availability of sentiment data, we restrict our sample period over July 1965 to November 2014.

To construct purged sentiment index IS-P, we remove economic fundamental factors from the six sentiment proxies. Since there is an extensive list of economic fundamental variables, first of all, we extract some common factors from a broad range of macroeconomic variables. Following Kyle, Ludvigson, and Ng (2015)[8], we use the information of 109 macroeconomic variables that are categorized into seven groups, including: (1) output and income, (2) employment, (3) housing, (4) consumption, orders and inventories, (5) money and credit, (6) exchange rates, (7) inflation. We derive the first principal component of the macroeconomic variables in each group and remove them from the sentiment proxies. A detailed description of all the macroeconomic variables is given in the Appendix A. Secondly, we remove two macroeconomic variables which are documented in asset pricing literature as business cycle indicators but not included in the 109 variables: consumption-to-wealth ratio as in Lettau and Ludvigson (2001) and GDP growth as in Lemmon and Portniaguina (2006). Furthermore, we remove three financial variables drawn from the literature which are frequently used as indicators of business cycle: the yield on three-month Treasury Bills, the default spread which is measured as the difference between the yields to maturity on Moody's Baa-rated and Aaa-rated bonds, and the term spread which is measured as the difference in yields between the ten-year Treasury bond and the three-month Treasury Bill[9]. Finally, we remove two risk factors: dividend yield of the value-weighted CRSP market portfolio as in Campbell and Shiller (1988a, 1988b) and liquidity risk factor measured as percentage of stocks with zero returns as in Lee (2011)[10]. For each sentiment proxy, we remove all the 14 fundamental

[8]Kyle, Ludvigson and Ng (2015) construct the latent common factor as the principal components from 132 macroeconomic variables from FRED-MD database. Besides the seven groups we use in our estimation, they include the 8th group: the variables about overall stock market. We exclude the variables in 8th group from the macroeconomics variables, because we control equity risk separately later. We also exclude bond related variables in the 6th group for consideration of redundancy because we remove 3-month treasury bill rate, term spread and default spread which are all bond related variables.

[9]Studies that use financial variables as business cycle indicators include Campbell (1987), Hodrick (1992), and Chen, Roll and Ross (1986).

[10]We tried to remove different macro-economic variables, and have consistent results. We will explain the details in the robustness check part.

factors documented above in the following regression:

$$x_{i,t} = a + b^{'}(Z_t) + \kappa_{i,t},$$

where $x_{i,t}$ represents each sentiment proxy, Z_t denotes the 14 fundamental variables and $\kappa_{i,t}$ is the regression residual. We define the six purged sentiment proxies as CEFDres, TURNres, NIPOres, RIPOres, PDNDres and EQTIres.

Panel A of Table 1 presents summary statistics of BW index, $BW^{\perp}$ and 14 fundamental factors. For each variable, we report the mean, standard deviation, first-order autocorrelation ($\rho(1)$), correlation with BW index and the data source. Many of the fundamental factors possess a common feature that they are highly persistent, and this pattern is quite similar to BW index and $BW^{\perp}$ index. We find BW index is significantly correlated with many of the fundamental factors, such as labour market employment, housing, consumption, orders and inventory, consumption wealth ratio, GDP growth, three-month Treasury bill, default spread, dividend yield and liquidity factor. Among these variables, three-month treasury bill rate, GDP growth, consumption and liquidity have the highest correlations with BW sentiment index. This implies that a considerable proportion of BW sentiment index is related to systematic risk. By contrast, although Baker and Wurlger (2006) has tried to remove several business cycle variables from original BW index to derive $BW^{\perp}$ index, BW index and $BW^{\perp}$ index are highly correlated (the correlation is 0.97).

In Panel B, we report the regression result of BW index on the 14 fundamental factors. We present the estimated coefficients, OLS t-statistics and Newey-West t-statistics which has been adjusted for 12 lags. We find that adjusted R-squares for BW index is about 62%, indicating that the BW index contains a considerable portion of information related to economic fundamental conditions.

We detail the decomposition regression results for each sentiment proxy in the Appendix. For each decomposition, R-squares range from approximately 30% to over 50% except the first-day returns of IPOs (RIPO) with a relatively low R-square, indicating a considerable portion of the variation in each sentiment proxy can be explained by economic fundamentals. Specifically, for close-end fund discount rate (CEFD) and dividend premium (PDND), a large proportion of R-square is due to the contribution of three-month treasury bill rate, term spread and liquidity risk. Moreover, the contribution of the fundamental variables varies in different sentiment proxies. For example, housing related variables and liquidity risk factor contribute most for share turnover (TURN), while labour market related variables show up with high explanatory power alongside with three-month treasury bill rate and term spread for number of IPOs (NIPO).

Panels A and B in Table 2 provide summary statistics of the six raw sentiment proxies and six purged sentiment proxies. All the sentiment proxies are standardized to have zero mean and unit variance. Each purged sentiment proxy is estimated as the regression residual of the raw sentiment proxy on 14 fundamental factors. Panel A presents the mean, standard deviation, first-order autocorrelation ($\rho(1)$), minimum, maximum of the six raw sentiment proxies, their correlations with BW sentiment index and their correlation matrix. Four out of the six sentiment proxies are positively correlated with BW sentiment index, except close-end fund discount rate CEFD and dividend premium PDND. Panel B presents summary statistics and correlations of the six purged individual sentiment proxies. Since the common macroeconomic variation has been removed from the purged sentiment proxies, it is not surprising that the six purged sentiment proxies show similar pattern but smaller magnitude in terms of persistency and correlation compared with raw sentiment proxies. Therefore, it would be more challenging to efficiently extract the underlying commonality among the purged sentiment proxies.

B. Purged sentiment index

Following the two-steps of estimation procedures of PLS, we obtain the purged investor sentiment index IS-P from the six purged individual sentiment proxies,

$$\begin{aligned} IS\text{-}P_t = & -0.17 * CEFDres_t + 0.20 * TRUNres_{t-12} + 0.38 * NIPOres_t \\ & +0.29 * RIPOres_{t-12} - 0.49 * PDNDres_{t-12} + 0.27 * EQTIres_t \end{aligned} \tag{7}$$

Since some proxies need longer time to reveal the same sentiment (Huang, Jiang, Tu and Zhou, 2015), the purged share turnover, purged average first-day return of IPO, and purged dividend premium are taken as lagged 12 months relative to other three purged proxies.

We also detail the weights of the six raw sentiment proxies when forming BW sentiment index.

$$\begin{aligned} BW_t = & -0.28 * CEFD_t + 0.18 * TRUN_{t-12} + 0.07 * NIPO_t \\ & + 0.10 * RIPO_{t-12} - 0.58 * PDND_{t-12} + 0.10 * EQTI_t \end{aligned} \tag{8}$$

Compared with the weights of the raw proxies in BW sentiment index, all the six purged proxies have the same signs as the corresponding raw proxies in BW index. The weights of residuals in NIPO, RIPO and EQTI when forming our purged sentiment IS-P are much higher while the weights of non-fundamental component in CEFD and PDND are lower. In contrast to the high correlation of 0.97 between Baker and Wurgler's orthogonal sentiment index $BW^{\perp}$ and their original sentiment index BW, the correlation between purged sentiment IS-P and BW original sentiment index is only 0.56.

Figure 1 plots time series of BW index and IS-P, showing IS-P captures almost all the anecdotal accounts of fluctuations as BW index does. Both sentiment indices are low at the beginning of the sample after the 1961 crash of growth stocks, and then reach a spike in the electronic bubble in 1968 and 1969. Sentiment declines subsequently until the middle of 1970s and rebounds from late 1970s to mid-1980s. During the late 1980s, sentiment falls and reaches a peak again in the Internet bubble period from 1999 to 2001. The sentiment indices decrease during subprime debt crisis from 2008 to 2009 and rebound in 2010. Although the two indices are highly correlated, IS-P is more volatile and appears to lead BW index in some cases. Particularly, during the periods after financial crisis, roughly from year 2009 to 2014, our purged sentiment stays slight above the BW sentiment, inferring that purged sentiment is less dragged down by bust fundamental conditions in the crisis.

We plot time series of the six raw individual sentiment proxies and the six purged sentiment proxies in Figure 2. On the one hand, the raw proxies and the purged proxies show comovement during the whole sample period. On the other hand, the two types of sentiment proxies sometimes apparently deviate from each other. The only exception is the residual component in RIPO, which deviates less from raw RIPO because the fundamental variables contribute less in explaining RIPO than in explaining other sentiment proxies, making the two variables RIPO and RIPOres much closer.

C. Portfolio return

Prior research shows investor sentiment affects cross-section stock return, especially for stocks that are difficult to arbitrage or value. Concretely, Baker and Wurgler (2006) document firms that are newer, smaller, more volatile, unprofitable, non-dividend paying, distressed or with extreme growth potential, and firms with analogous characteristics are more sensitive to investor sentiment. Following Baker and Wurgler (2006), we construct spread portfolios (i.e., high, medium and low) according to NYSE breakpoint of firm characteristics, such as size, age, dividend payment, earnings, tangible assets, R&D, sigma, external finance, sales growth, and book-to-market ratio. We define the top three NYSE deciles as high, firms in the bottom three NYSE deciles as low, and remaining middle four NYSE deciles as medium.

Figure 3 shows future returns of the spread portfolios conditional on firm characteristics and sentiment that is estimated as monthly average of BW sentiment index in previous calendar year. We plot the average monthly portfolio returns following positive sentiment periods in solid bars, and portfolio returns following negative sentiment periods in clear bars. The dashed lines are the unconditional average portfolio returns across two regimes of sentiment periods and the solid lines are the differences. Generally, we find that sentiment effect is stronger for firms that are hard to value and arbitrage, consistent with Baker and Wurgler (2006). For instance, Panel A shows the size effect conditional on sentiment. It reveals that size effect only appears in negative sentiment periods. Specifically, following negative sentiment period average monthly return (clear bar) for the bottom size group is approximately two times as large as the return for the top size group, while following positive sentiment period average monthly returns (solid bar) for the bottom, medium and top size deciles are nearly the same. The difference (solid line) illustrates that the future returns on smaller size firms are more sensitive to the fluctuation of sentiment. A similar pattern is apparent when conditioning on firm age (presented in Panel B), earnings (presented in Panel D) and dividends (presented in Panel E[11]). We find that younger firms, unprofitable firms, and non-dividend paying firms are more sensitive to sentiment.

Panel C shows that firms in high volatility risk group earn higher returns than in low volatility risk group following negative sentiment period, while it shows an opposite pattern following positive sentiment period. The solid line summarizes the return difference across two regimes of sentiment, indicating more volatile firms are more influenced by sentiment. In Panel F, the patterns are not so strong, but suggest that the future returns of firms with less tangible assets are more sensitive to sentiment effect. Panel G implies a clear unconditional effect of RD/A portfolios – firms with higher RD/A earn higher returns. The remaining sorting variables, i.e., book-to-market, external finance, and sales growth, show intriguing patterns. First, they all show a monotonic unconditional effect – future returns are generally higher for high BE/ME stocks, low EF/A stocks, and low GS decile stocks. Second, they display a U-shaped pattern in the conditional difference. Specifically, the difference in returns across sentiment regimes are greater for both the bottom and the top deciles than the difference for the medium decile.

Baker and Wurgler (2006) document the conditional effect of sentiment on the spread portfolios which buy the high group and sell the low group (high-low portfolio). For instance, high-low portfolio based on age has higher return when BW sentiment is positive, and has lower return when BW sentiment is negative. However, BW sentiment index has the opposite conditional effect on high-low portfolio based on sigma. For consistency, we construct long-short portfolios which BW sentiment has the same direction of conditional effect on. Specifically, we construct the "high-low" portfolios, which have long legs in the top deciles (less exposed to sentiment) and short legs in bottom deciles (more exposed to sentiment), based on size, age, dividend payment, earnings, fixed assets, book-to-market ratio, and "low-high" portfolios, which have the long legs in the bottom

[11] Baker and Wurgler (2006) suggest that for common investors, the most salient comparisons are those between profitable and unprofitable firms and dividend payers and nonpayers. Therefore, in Panel D and Panel E, we directly compare profitable and unprofitable firms and dividend payers and nonpayers.

deciles and short legs in top deciles in terms of R&D, sigma, external finance, and sales growth. For variables related to growth and distress: external finance, sales growth and book-to-market ratio, the relationships between sentiment and them are not monotonic. Following Baker and Wurgler (2006), we break external finance, sales growth and book-to-market ratio into medium-high and medium-low portfolios. In addition, we construct combined portfolio, which takes equal positions across the 16 firm characteristics based portfolios.

Table 3 summarizes the properties of the 16 characteristics based portfolio as well as the combined portfolio. Panel A shows summary statistics of return variable and all the sorting variables. Panel B presents mean excess return (returns in excess of the monthly Treasury bill rate) and accompanying t-statistics on the long legs and short legs of each portfolio as well as the long-short portfolio. Panel C reports the corresponding values for benchmark-adjusted returns, which are the estimates of a_i from the regression

$$Ret_{i,t} - rf_t = a_i + b * MKT_t + c * SMB_t + d * HML_t + e * WML_t + \varepsilon_t, \tag{9}$$

where $Ret_{i,t} - rf_t$ is the portfolio excess return in month t. Table 4 presents the correlations among the long-short portfolio returns. Not surprising, the spread returns are highly correlated with each other, which is consistent with Baker and Wurgler (2006).

IV. Predictability of sentiment

In this section, we start with investigating and comparing the predictability of BW index, residual components in BW index (BW$'$ and BW$''$) and purged sentiment index IS-P on cross-section returns.[12] We find the purged sentiment index performs as well as BW index and BW$'$ while it substantially outperforms BW$''$. Furthermore, we analyze the predictability of purged sentiment index on future economic activities and investigate its relation with business cycle.

[12]We regress BW index on a small set of economic variables documented in Baker and Wurgler (2006) and take the regression residual as BW$'$, and in the same way, we remove a large amount of fundamental variables described in Section III and define the regression residual as BW$''$.

A. The predictability of BW index

Table 5 reports the results of using BW index as the predictor for long-short portfolio returns, long-leg returns and short-leg returns of 16 firm characteristics based portfolios.[13] After controlling Fama French three factors and Carhart's momentum factor[14], BW index significantly predicts 11 out of the 16 long-short portfolios. In terms of long or short-leg portfolio returns, it can forecast 4 out of the 16 long-leg returns, and 14 out of the 16 short-leg returns. We present the regression results for the long-short return spreads from column 3 to column 6. We report the estimated coefficients and bootstrapped p-values to correct the bias of autocorrelation (Stambaugh, 1999). The results are consistent with Baker and Wurgler (2006), showing BW can predict most of portfolio returns except for the portfolios based on PPE/A, RD/A, BE/ME, EF/A and GS in which the predictive power disappears after controlling Fama French three factors and Carhart's momentum factor. For portfolios based on Medium-High, Medium-Low strategies of "growth and distress" variables: external finance, sales growth and book-to-market ratio, the regression results illustrate the significant U-shape pattern which is also documented in Baker and Wurgler (2006).

Stambaugh, Yu and Yuan (2012) argue that due to the short-sale constraint, overpricing is more prevalent than underpricing. Specifically, the short legs of the anomalies should be more profitable following high sentiment, and sentiment exhibits no relation with the return of long legs. Although the construction of our long-short portfolios is different from the anomalies in Stambaugh, Yu and Yuan (2012), sentiment should have stronger predictive power for the short legs of portfolio returns since the short legs are set to be more exposed to sentiment. We report the results of predictive regression of BW index for short legs of the portfolios from column 11 to column 14. Without controlling Fama French three factors and Carhart's momentum factor, the BW index significantly and negatively predicts all the short legs, and after controlling four factors, BW index significantly predicts 14 out of the 16 short legs. We report the results of predictive regression for long legs from column 7 to column 10, in which BW index only significantly predicts 4 out of 16 portfolios

[13] We also consider the orthogonal Baker and Wurgler (2006) index $BW^{\perp}$ and find similar results.

[14] When the portfolio is formed based on SMB or HML, SMB or HML is not included as a control variable

controlling Fama French three factors and Carhart's momentum factor. The findings are consistent with Stambaugh, Yu and Yuan (2012)'s prediction that sentiment exhibits asymmetric impacts on the long legs and short legs.

B. The predictability of BW$'$ and BW$''$

Although BW index can predict most of the portfolio returns, we cannot distinguish whether the predictability is driven by investor sentiment or economic fundamental risks. In this section, we try some straightforward methods to remove the fundamental factors from the BW index. Firstly, we directly remove six macroeconomic variables which are documented by Baker and Wurgler (2006) from BW index. We regress BW index on the six macroeconomic variables, i.e., the growth of industrial production, the growth of durable consumption, the growth of nondurable consumption, the growth of service consumption, the growth of employment, and a dummy variable for NBER-dated recessions, and define the regression residual BW$'$ as a new sentiment index. In Panel A of Table 6, we report the regression results of using BW$'$ as investor sentiment proxy to predict cross-section returns. After controlling Fama French three factors and Carhart's momentum factor, BW$'$ significantly predicts 11 out of the 16 long-short portfolio spreads, five out of the 16 long-leg returns, and 14 out of the 16 short-leg returns. The predictability is comparable with BW index, but we have a concern that we fail to remove fundamental information thoroughly.

Next, to alleviate the concern, we remove a comprehensive set of fundamental information, including seven first principle components extracted from 109 macroeconomic variables and another seven business cycle related variables that we have explicitly explained in Section III, from BW sentiment index to obtain the residual component BW$''$. Panel B of Table 6 reports the results of predictive regression based on BW$''$. Compared with BW index, we find that the predictive ability of BW$''$ on long-short portfolios diminishes greatly. In the regressions without control factors, BW$''$ forecasts only 3 out of the 16 long-short return spreads and none of the 16 long/short-leg returns. After controlling the Fama French three factors and Carhart's momentum factor, the number of significance becomes even less: BW$''$ significantly forecast merely one out of the 16 spread returns, and none of the 16 long/short-leg returns. After removing the component related to the economic fundamental, the BW sentiment index almost loses the predictability for the future cross-sectional portfolio return[15]. The diminishing predict power in cross-section returns after removing economic fundamental casts a doubt on whether Baker and Wurgler index is appropriate to serve as an investor sentiment index or whether the PCA method used to construct BW index is appropriate.[16]

C. The predictability of IS-P

Econometrically, the investor sentiment extracted from PCA method may involve a substantial amount of common approximation errors which are irrelevant for forecasting cross-section returns. Therefore, we use an improved econometric way PLS to construct the purged sentiment index IS-P. The purged sentiment index has several desirable features. First, IS-P is constructed from purged sentiment proxies, from which fundamental information has been removed largely. Second, PLS estimation aligns the investor sentiment with the purpose of explaining the future cross-sectional return and only extracts the information relevant for forecasting target.

We study the predictability of IS-P on cross-section stock returns in Panel A of Table 7. We find that the purged sentiment index can predict the cross-sectional stock returns remarkably well. Panel A demonstrates that IS-P significantly predicts 12 out of the 16 long-short portfolio, 13 out of the 16 long-leg returns and all of the 16 short-leg returns. After controlling Fama French three factors and Carhart's momentum factor, IS-P is statistically significant in predicting 13 of the 16 long-short portfolio returns, two of the 16 long-leg returns and 10 of the short-leg returns.

In Panel A of Table 7, the first three rows show that when purged sentiment is higher, returns on small, young and high volatility firms are relatively lower in the next month. In terms of economic magnitudes, for instance, the coefficient for predicting size portfolio indicates that a

[15]As a robust check, we also adopt PCAres, which is the PCA of the six purged sentiment proxies, to predict the portfolio returns. We find that the residual component PCAres almost loses the ability to predict the portfolio returns.

[16]Sibley, Wang, Xing and Zhang (2016) has done similar orthogonalization and also find the predictability of BW becomes much weaker.

one-unit increase in sentiment (which is equivalent to a one standard deviation increase because the indexes are standardized) is associated with a 0.5% higher monthly return on the large minus small portfolio. For profitability and dividend payment, we find that IS-P also has significant predictive power for these portfolios, with higher purged sentiment forecasting relatively lower returns on nonpayers and unprofitable firms. The patterns of long-short and short leg are little affected after controlling for Fama and French factors.

In Baker and Wurgler (2006), the predictability of BW index on long-short PPE/A and RD/A portfolios is insignificant. However, our purged sentiment index significantly predicts the tangibility characteristics based portfolios returns. From row 6 to row7, we show that purged sentiment has significant predictive power for the PPE/A and RD/A portfolios. The higher IS-P, the lower future returns on low PPE/A stocks and high RD/A stocks. The findings are in line with the theoretical prediction that the valuation of a firm with less tangible assets tends to be more subjective, thus its stock is affected more by the fluctuations of investor sentiment.

Baker and Wurgler (2006) demonstrate that "growth and distress" variables do not have simple monotonic relationships with sentiment. We find consistent retulsts from row 8 to 10 showing that purged sentiment does not predict high minus low portfolios formed on BE/ME, EF/A, or GS. However, in the following 6 rows, we present that the predictability of our purged sentiment index on the medium-high and medium-low portfolios of BE/ME, EF/A and GS is strong, matching the U-shaped pattern inferred from Figure 3.

In Panel B of Table 7, we summarize the number of significance in terms of forecasting long-short, long-leg and short-leg portfolio returns employing BW, BW$'$, BW$''$ and IS-P, respectively. In recap, the numbers show that the predictive ability of IS-P is comparable to or even better than the original BW sentiment index. Particularly, IS-P has significant predictive power for the long-short portfolios based on tangibility characteristics (PPE/A portfolio and RD/A portfolio). Furthermore, the signs of the coefficients on the various firm characteristics based on IS-P are consistent with the signs documented by Baker and Wurgler (2006). Hence, our purged sentiment index could be considered as a better measure of a behavioral driven investor sentiment.

D. IS-P and macro-economy

In this section, we examine whether purged sentiment can forecast future macroeconomic conditions. If IS-P contains much information related to fundamental economics, we expect that it would significantly predict the macroeconomic variables. In Table 8, we test the forecasting ability of purged sentiment for several representative macroeconomic indicators which are principle components derived from four categories of macroeconomic variables, including output and income, employment, housing and consumption, orders and inventories (Kyle, Ludvigson and Ng (2015)). The table reports the coefficient estimates, Newey-West t-statistics on the lagged purged sentiment in Panel A (lagged BW sentiment in Panel B) and the R squares of the regressions. As show in Panel B of Table 8, BW sentiment index significantly predicts employment, housing and consumption related macroeconomic activities, and is marginally significant in predicting output and income related macroeconomic variable. By contrast, all of the coefficient estimates in Panel A of Table 8 are not statistically significant, implying that purged sentiment contain little information regarding future macroeconomic conditions.

Figure 4 plots the peaks and troughs of the business cycle as defined by the NBER data along with the contemporaneous purged sentiment. If the purged sentiment index is a proxy for an omitted macroeconomic risk factor, we expect the purged sentiment index tend to be procyclical. However, as shown in the figure, this does not appear to be the case. Specifically, over the 14 reported business cycle peaks and troughs during our sample period, purged sentiment indicator goes into the opposite direction with business cycle indicator for half of the reported peak/trough dates. The evidence further indicates that purged sentiment is less likely to be related to the state of the macroeconomy.

E. Robustness investigation for IS-P

When we construct our purged sentiment proxy, we remove a wide range of fundamental information, including first principal components from seven groups of macroeconomic variables and seven fundamental variables to derive purged sentiment proxies, and use PLS to extract the aligned investor sentiment. As robustness checks, we remove alternative sets of fundamental variables to derive purged sentiment proxies. Firstly, we extract seven common factors from more than one hundred of macroeconomic variables using asymptotic principal component analysis. In this way, we summarize fundamental information from a large number of macroeconomic time series into a small number of estimated common factors.[17]We remove the seven common factors along with the seven fundamental variables the same as we detail in Section III. Secondly, we directly remove a wide range of raw fundamental variables, i.e.,130 macroeconomic variables from FRED-MD and five fundamental related variables, which are consumption-to-wealth ratio, GDP growth, default spread, dividend yield and liquidity risk factor.[18] We present predictability of purged index constructed from PLS on alternative purged sentiment proxies in Panel A of Table 9. For instance, based on the first alternative definition of fundamental variables, the purged sentiment significantly predicts 12 out of 16 long-short return spreads, two out of 16 long-leg portfolios and 12 out of the 16 short-leg portfolios after controlling four factors.

In Panel B, we remove fundamental variables directly from BW sentiment index. When Fama French three factors and Carhart's momentum factor are included as control variables, the residual component of BW sentiment based on alternative 14 fundamental variables can predict only one out of 16 long-short portfolio spreads, and none of the 16 long-leg/short-leg portfolios.

In Panel C, We use PCA way to construct the residual sentiment index from the purged sentiment proxies. Compared with applying PLS, the predictive performance of PCA on purged sentiment proxies diminishes greatly. Specifically, we find that using the 14 fundamental variables in Section III, residual sentiment based on PCA can forecast only two long-short portfolio spread, none of the long-leg portfolio and two out of 16 short-leg portfolios. Using the alternative 14 variables, residual sentiment can forecast four out of 16 long-short portfolio spread, none of the long-leg portfolio and three out of 16 short-leg portfolios.

In summary, it shows that sentiment residual constructed on alternative 14 variables performs similarly to the counterpart constructed on the 14 fundamental variables in Section III. We also

[17]To determine the number of common factors, we use BIC information criterion. (see Schwarz, 1978)

[18]We delete three-month Treasury Bill rate and term spread because of multicollinearity.

reach consistent results using 135 variables as fundamentals.

V. Validation of IS-P

In previous section, we construct the purged sentiment index from which fundamental information has been removed largely and demonstrate its persistent predictability on cross-sectional stock returns. In this section, we further apply three tests to validate our purged sentiment index: earnings annoucment returns, fund flow and non-fundamental component in Tobin's Q.

A. IS-P and earnings announcement returns

Our first validation test involves earnings announcement returns. La Porta, Lakonishok, Shleifer and Vishny (1997) argue that earnings announcement returns reflect investors correction of their errors in earnings expectation. Since investors tend to make errors in the firms that are difficult to value, we expect that earnings announcement returns would be lower for difficult- to-value firms after high sentiment period. For each quarterly earnings announcement, we calculate the three-day cumulative abnormal return around the report date (CAR(-1,1)). The portfolios are constructed the same as in Section III.C. We define firms in the top three NYSE deciles as high, the bottom three NYSE deciles as low, and remaining middle four NYSE deciles as medium, for firm size (ME), age, total risk (Sigma), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS) respectively. Portfolios formed on profitability (E/BE) and dividend (D/BE) are divided into two groups: profitable/unprofitable firms, and dividend payers and non-payers. We average CAR (-1,1) within each characteristic portfolio per month, and run the following regression to examine the relationship between our purged sentiment measure and earnings announcement returns:

$$CAR_{i,t} = a + b * IS\text{-}P_{t-1} + \varepsilon_t, \tag{10}$$

where $CAR_{i,t}$ is the average of CARs around quarterly earnings announcements within each characteristic portfolio in month t, $IS\text{-}P_{t-1}$ is our purged sentiment measure in month t-1.[19]

Table 10 reports the coefficient estimates for each characteristic portfolio using purged sentiment index from column 3 to column 5. In general, the coefficients of purged sentiment are negative, indicating that earnings announcement effects are lower following high sentiment periods. Moreover, the earnings announcement effects are much stronger for uncertain or difficult-to-value stocks which are usually more sensitive to sentiment: heteroskedasticity-robust t-statistics of the coefficients of purged sentiment are larger for small stocks, young stocks, high volatility stocks, non-dividend paying stocks, and also stocks with low PPE and high R&D. Particularly, the stronger earnings announcement effects for stocks with low PPE and high R&D are consistent with our findings in Section IV.C that purged sentiment is more significant for portfolio returns based on firms with less tangible assets.

We also summarize the coefficient estimates based on BW sentiment index from column 6 to column 8 in Table 10 for comparison. Among all cases, the difference between High and Low based on ME, RD/A and D/BE is larger for purged sentiment index IS-P than for BW sentiment index BW. For BE/ME, EF/A and GS, we compare the difference between High and Medium and between Low and Medium: for BE/ME, the difference between High and Medium and between Low and Medium are larger for IS-P than for BW; for GS, the difference are similar for IS-P and BW; for EF/A, only the difference between Low and Medium are smaller for IS-P than for BW whilst the difference between High and Medium are similar for IS-P and BW. Thus, regarding five out of the total 16 cases: ME, RD/A and D/BE and BE/ME (H-M; M-L), the difference are larger for IS-P than for BW. Only for two case - EF/A [M-L] and E/BE - the difference is smaller for IS-P than for BW. Regarding the rest 9 cases, the difference are similar for IS-P and BW. Therefore, IS-P is comparable to or better than BW in predicting CAR, which could be related to sentiment caused time-varying return.

Overall, the results support the view that investors are more likely to suffer errors in earn-

[19] To eliminate the noise from individual stocks, we require that the number of CARs used to calculate $CAR_{X_{it}=H/M/L,t}$ in month t is larger than a critical value of 15.

ings expectations for stocks which are more sensitive to sentiment and further validate our purged sentiment measure as a proxy for sentiment with behavioral explanation.

B. Purged sentiment and fund flow

In the second validation test, we connect the IS-P with mutual fund flow measured as investor inflows into equity-oriented mutual funds. The mutual fund flow reflects investor sentiment towards stock market (Ben-Rephael, Kandel and Wohl (2012)).[20] In Specifications (1) and (2) of Table 11, we find that IS-P is positively and significantly correlated with contemporaneous mutual fund inflows with a Newey-West t-statistics of 2.89. We also find $IS\text{-}P_t$ positively predicts next period's equity-oriented mutual funds inflows with significant t-statistics of 2.79. In Specifications (3) and (4), we use BW index as the explanatory variable. We find that BW index has no significant relationship with current or next period aggregate fund inflows, which is consistent with prior findings in the literature (e.g., Ben-Rephael, Kandel and Wohl (2012)). The evidences that purged sentiment is consistent with investors' actual behavior indicate that purged sentiment reflects widely shared investor beliefs rather than merely noises.

C. Non-fundamental component in Tobin's Q

The last validation test involves non-fundamental component of Tobin's Q. We follow Rhodes-Kropf, Robinson and Viswanathan (2005) and decompose Tobin's Q into fundamental component and non-fundamental component, which capture firm's growth opportunities and mispricing in Tobin's Q respectively. We use the mispricing part in Tobins Q (mQ) as a proxy for investor sentiment and compare its predictability on portfolio returns.

Panel A of Table 12 shows the regression of the purged sentiment on the contemporaneous mispricing component in Tobin's Q (mQ). The relation between IS-P and mQ is positive and significant, with a Newey-West t-statistics of 4.90, indicating IS-P captures relevant mispricing

[20] We obtain monthly mutual fund inflows data from Investment Company Institute and scale the net dollar inflows in each month by the aggregate capitalization of the U.S. stock market.

information in Tobins Q. In Panel B of Table 12, we investigate the predictability of mQ on long-short, long-leg and short-leg combined portfolio returns respectively. We find that mQ fails to significantly forecast portfolio returns in the next month. Panel C further test the forecasting ability of IS-P on portfolio returns controlling mQ. The results show that IS-P significantly forecast combined portfolio returns while the predictability of mQ is weak, suggesting IS-P better captures investor sentiment in mispricing than mispricing part in Tobins Q.

VI. Further analysis

A. Economic explanation

Stocks are priced by discounting their future cash flow at a discount rate. Campbell and Shiller (1988b) develop a convenient framework to analyze cash flow and discount rate, in which they establish a loglinear approximate identity

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t, \tag{11}$$

where $r_{t+1} \equiv log(P_{t+1} + D_{t+1}) - log(P_t)$, and ρ is a positive log-linearization constant. P denotes price, D denotes dividend, and lowercase letters indicate the log transforms. Following Cochrane (2011), we rearrange the approximate identity and derive

$$r_{t+1} \approx k + \Delta d_{t+1} - \rho dp_{t+1} + dp_t, \tag{12}$$

where Δd_{t+1} is the log dividend growth rate and dp_{t+1} is log dividend price ratio. (12) implies that the predictability of our purged sentiment index on portfolio returns should stem either from cash flow channel or discount rate channel. In this section, we explore through which channel our purged sentiment index influences portfolio return by testing the predictability of our purged sentiment index on future cash flow and discount rate. Dividend growth is widely used as the proxy for cash flow in the previous asset pricing literatures (e.g., Campbell and Shiller, 1988b, Cochrane, 2008, 2011), therefore we use dividend growth of the portfolio as our cash flow proxy. Cochrane (2008, 2011) documents that the variation from discount rate is largely driven by dividend price ratio, hence, we use dividend price ratio of the portfolio as the proxy for discount rate. We focus on the short legs of the portfolios on which our sentiment index has more pronounced effect and use following bivariate predictive regression model:

$$y_{t+1} = \alpha + \beta IS\text{-}P_t + \varphi dp_t + \omega_{t+1}, y = \Delta d_{t+1} \; or \; dp_{t+1} \tag{13}$$

where Δd_{t+1} is the annual log dividend growth rate on the short leg of the portfolio from July of year t to June of year t+1, dp_{t+1} is the log dividend price ratio on the short leg of the portfolio in June of year t+1, and $IS\text{-}P_t$ is the purged sentiment in June of year t. Following the literature, we use annual data to avoid spurious predictability arising from seasonality.

Table 13 reports the estimation results for the predictability regression. In Panel A, Δd_{t+1} and dp_{t+1} are constructed based on the short leg of the combined portfolio and used as the dependent variables respectively. The estimated coefficient of $IS\text{-}P_t$ for Δd_{t+1} is -6.65 with t-statistics of -3.38, while the estimated coefficient of $IS\text{-}P_t$ for dp_{t+1} is close to zero and insignificant. The lagged dividend price ratio dp_t has strong forecasting power for future dividend price ratio dp_{t+1} with a mean reverting coefficient of around 0.9, whereas its forecasting ability for dividend growth rate is much weaker, consistent with Cochrane (2008, 2011) showing dividend price ratio capture time variation in discount rate.

In Panel B, Δd_{t+1} and dp_{t+1} are constructed based on the short legs of individual characteristics portfolios respectively. We find consistent results as in Panel A that $IS\text{-}P_t$ significantly and negatively predicts Δd_{t+1} while $IS\text{-}P_t$ has no significant predictability under conventional 5% significance level for dp_{t+1} in portfolios formed on various firm characteristics, such as age, volatility, dividend, tangibility, etc. The negative coefficients of $IS\text{-}P_t$ for Δd_{t+1} (β) further confirms the negative coefficients of $IS\text{-}P_t$ for short-leg portfolio returns r_{t+1} documented in Section IV.

Since dividend growth rate and dividend price ratio represent cash flow channel and discount rate channel separately, the results indicate that the negative predictability of purged sentiment for short legs of portfolio returns comes from the cash flow channel, i.e, the higher investor sentiment leads to the overoptimistic opinion towards future cash flow, and when the realized cash flow is lower than investors' expectation, the future return decreases.

B. Alternative survey based sentiment measures

In a recent study, Greenwood and Shleifer (2014) find that survey based investor expectations of market returns negatively predict future stock returns. The evidence is consistent with Baker and Wurgler's findings that investor sentiment negatively affect cross-sectional stock returns and favors a behavioral explanation. In this section, we investigate several survey-based sentiment measures and compare their predictability on portfolio returns with our purged sentiment index.

Table 14 reports the relationship between the purged sentiment index and several non-fundamental components of survey based sentiment measures. We obtain anxious index (AI) from Federal Reserve Bank of Philadelphia that measures the probability of a decline in real GDP, consumer sentiment index (ICS) from Michigan University, individual investor sentiment index (AAII) from American Association of Individual Investor survey and rescaled Gallup investor index GA from Galllup survey.[21] We orthogonalize each survey based sentiment index to the 14 macroeconomic variables that we described in Section IV. We take the fitted value as fundamental component and the regression residual as non-fundamental component. After decomposing anxious index (AI), consumer sentiment index (ICS), individual investor sentiment index (AAII), and rescaled Gallup investor index GA, we derive AIres, ICSres, AAIIres and GAres respectively.

Panel A of Table 14 reports the univariate regressions of the relation between purged sentiment and the contemporaneous survey based sentiment measures respectively. We find that non-fundamental component in anxious index AIres is positively and significantly correlated with our

[21]We rescale Gallup investor index GA by projecting the stock return expectation (available between 1999 and 2003) onto the raw Gallup series.

purged sentiment index IS-P, with a Newey-West t-statistics of 2.59, while ICSres is negatively and significantly correlated with our purged sentiment index IS-P. The relation with AAIIres and GAres are not significant.

Then, we investigate the predicative power of each survey based sentiment measure on long-short, long leg and short leg of the combined portfolio return in the next month. We report the results from specifications (1) to (4) in Panels B, C and D respectively. All these predictability are insignificant or weak under the conventional 5% significance level, and the only exception is using GAres, which is the residual from rescaled Gallup survey, to predict the long leg of combined portfolio return. GAres is positively correlated with next month long-leg combined portfolio return with a Newey-West t-statistics of 2.42.

For the convenience of comparison, we also examine the predictability of the purged sentiment index on portfolio returns in the sample period adjusted to the data length of each survey based sentiment measure.[22] We present the results in specifications (5) to (8) in Panels B, C and D respectively. We find that under the 5% significance level, the purged sentiment significantly predicts short-leg returns and long-short returns of the combined portfolio with the only exception that IS-P during 1996 and 2011 (the GA period) significantly forecasts next month short-leg combined portfolio return under 10% significance level. Modifying the sample period does not qualitatively affect the predictability of our purged sentiment index.

Furthermore, we present the regressions of combined portfolio returns on survey based sentiment measures together with purged sentiment in multivariate settings in the last four columns in Panels B, C and D respectively. Specifications (9) to (12) in Panel B of Table 13 show that IS-P significantly predicts next month combined portfolio spread return while estimated coefficients of the survey based measures are insignificant at all. Regarding the long leg, neither IS-P nor the survey based measure can forecast long-leg combined portfolio return except rescaled Gallup survey series. For the short leg, both AIres and ICSres fail to forecast the portfolio returns, while IS-P

[22] Anxious index AI is from October 1968 to November 2014; Consumer sentiment index ICS is from January 1978 to November 2014; AAII survey data is from July 1987 to December 2011; Gallup survey data is from October 1996 to December 2011.

significantly forecasts combined portfolio return with negative sign. Although under 5% significance level, IS-P is not significant for short-leg combined portfolio return in regressions together with AAIIres and GAres, the t-statistics of IS-P is larger in magnitude than the t-statistics of survey based measures. Generally, our purged sentiment index shows higher predicative power for portfolio returns than other survey based sentiment measures.

VII. Conclusion

Since the creation of the influential Baker and Wurgler (2006) sentiment index, numerous studies treat the index as a behavioral variable and interpret their empirical results as consistent with the idea that investors sentiment, unrelated to fundamental risks, drives prices and returns in the market. However, given that the six proxies used to construct the Baker and Wurgler sentiment index are closely related to overall fundamental business environment, these studies could be misleading if the Baker and Wurgler sentiment index is driven by fundamental forces but not non-fundamental behavioural ones.

In this paper, we first remove fundamental information thoroughly from Baker and Wurgler (2006) six sentiment proxies by orthogonalizing each proxy to a broad series of economic fundamental variables. Then we exploit the residual information content of Baker and Wurgler (2006) six sentiment proxies via the partial least squares approach (PLS) to obtain a new purged sentiment index, which is likely driven by non-fundamental behavioural forces. Empirically, we find that our purged investor sentiment index has a similar or greater power in predicting the stock returns cross-sectionally compared with the original Baker and Wurgler sentiment index containing a large amount of fundamental information. Our study indicates that the original Baker and Wurgler (2006) sentiment index could contain behavioral related sentiment component and it seems fine for many studies to adopt the Baker and Wurgler (2006) sentiment index as a behavioral indicator.

A.1 Description of macroeconomic variables in FRED-MD

The column TCODE denotes the following data transformation for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$. (7) $\Delta(x_t/x_{t-1} - 1.0)$. The FRED column gives mnemonics in FRED followed by a short description. The comparable series in Global Insight is given in the column GSI.

Group 1: Output and Income

	id	tcode	fred	description	gsi	gsi:description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	Real personal income ex transfer receipts	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M_116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	19	1	NAPMPI	ISM Manufacturing: Production Index	M_110157212	NAPM prodn
17	20	2	CUMFNS	Capacity Utilization: Manufacturing	M_116461602	Cap util

Group 2: Labor Market

	id	tcode	fred	description	gsi	gsi:description
1	21*	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22*	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M_110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M_110156527	U < 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	M_110156523	U 5-14 wks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	M_110156525	U 15-26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M_110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours : Manufacturing	M_14386098	Avg hrs: mfg
29	49	1	NAPMEI	ISM Manufacturing: Employment Index	M_110157206	NAPM empl
30	127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	M_123109182	AHE: goods
31	128	6	CES2000000008	Avg Hourly Earnings : Construction	M_123109341	AHE: const
32	129	6	CES3000000008	Avg Hourly Earnings : Manufacturing	M_123109552	AHE: mfg

Group 3: Consumption and Orders

	id	tcode	fred	description	gsi	gsi:description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M_110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M_110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M_110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M_110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M_110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M_110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M_110155534	BP: West

Group 4: ORders and Inventories

	id	tcode	fred	description	gsi	gsi:description
1	3	5	DPCERA3M086SBEA	Real personal consumption expenditures	M_123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	M_110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
4	60	1	NAPM	ISM : PMI Composite Index	M_110157208	PMI
5	61	1	NAPMNOI	ISM : New Orders Index	M_110157210	NAPM new ordrs
6	62	1	NAPMSDI	ISM : Supplier Deliveries Index	M_110157205	NAPM vendor del
7	63	1	NAPMII	ISM : Inventories Index	M_110157211	NAPM Invent
8	64	5	ACOGNO	New Orders for Consumer Goods	M_14385863	Orders: cons gds
9	65*	5	AMDMNOx	New Orders for Durable Goods	M_14386110	Orders: dble gds
10	66*	5	ANDENOx	New Orders for Nondefense Capital Goods	M_178554409	Orders: cap gds
11	67*	5	AMDMUOx	Unfilled Orders for Durable Goods	M_14385946	Unf orders: dble
12	68*	5	BUSINVx	Total Business Inventories	M_15192014	M&T invent
13	69*	2	ISRATIOx	Total Business: Inventories to Sales Ratio	M_15191529	M&T invent/sales
14	130*	2	UMCSENTx	Consumer Sentiment Index	hhsntn	Consumer expect

Group 5: Money and Credit

	id	tcode	fred	description	gsi	gsi:description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
4	73	6	AMBSL	St. Louis Adjusted Monetary Base	M_110154995	MB
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M_110155011	Reserves tot
6	75	7	NONBORRES	Reserves Of Depository Institutions	M_110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	78	6	NONREVSL	Total Nonrevolving Credit	M_110154564	Cons credit
10	79*	2	CONSPI	Nonrevolving consumer credit to Personal Income	M_110154569	Inst cred/PI
11	131	6	MZMSL	MZM Money Stock	N.A.	N.A.
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N.A.	N.A.
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N.A.	N.A.
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A.	N.A.

Group 6: Interest rate and Exchange Rates

	id	tcode	fred	description	gsi	gsi:description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	85*	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	CPF3M	Comm paper
3	86	2	TB3MS	3-Month Treasury Bill:	M_110155165	3 mo T-bill
4	87	2	TB6MS	6-Month Treasury Bill:	M_110155166	6 mo T-bill
5	88	2	GS1	1-Year Treasury Rate	M_110155168	1 yr T-bond
6	89	2	GS5	5-Year Treasury Rate	M_110155174	5 yr T-bond
7	90	2	GS10	10-Year Treasury Rate	M_110155169	10 yr T-bond
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
10	93*	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS		CP-FF spread
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
18	101	5	TWEXMMTH	Trade Weighted U.S. Dollar Index: Major Currencies		Ex rate: avg
19	102*	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M_110154768	Ex rate: Switz
20	103*	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M_110154755	Ex rate: Japan
21	104*	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	M_110154772	Ex rate: UK
22	105*	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M_110154744	EX rate: Canada

Group 7: Prices

	id	tcode	fred	description	gsi	gsi:description
1	106	6	PPIFGS	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	PPIFCG	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	PPIITM	PPI: Intermediate Materials	M_110157527	PPI: int matls
4	109	6	PPICRM	PPI: Crude Materials	M_110157500	PPI: crude matls
5	110*	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	M_110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and metal products:	M_110157335	PPI: nonferrous
7	112	1	NAPMPRI	ISM Manufacturing: Prices Index	M_110157204	NAPM com price
8	113	6	CPIAUCSL	CPI : All Items	M_110157323	CPI-U: all
9	114	6	CPIAPPSL	CPI : Apparel	M_110157299	CPI-U: apparel
10	115	6	CPITRNSL	CPI : Transportation	M_110157302	CPI-U: transp
11	116	6	CPIMEDSL	CPI : Medical Care	M_110157304	CPI-U: medical
12	117	6	CUSR0000SAC	CPI : Commodities	M_110157314	CPI-U: comm.
13	118	6	CUUR0000SAD	CPI : Durables	M_110157315	CPI-U: dbles
14	119	6	CUSR0000SAS	CPI : Services	M_110157325	CPI-U: services
15	120	6	CPIULFSL	CPI : All Items Less Food	M_110157328	CPI-U: ex food
16	121	6	CUUR0000SA0L2	CPI : All items less shelter	M_110157329	CPI-U: ex shelter
17	122	6	CUSR0000SA0L5	CPI : All items less medical care	M_110157330	CPI-U: ex med
18	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
19	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	gmdcd	PCE defl: dlbes
20	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	gmdcn	PCE defl: nondble
21	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Group 8: Stock Market

	id	tcode	fred	description	gsi	gsi:description
1	80*	5	S&P 500	S&P's Common Stock Price Index: Composite	M_110155044	S&P 500
2	81*	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M_110155047	S&P: indust
3	82*	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield		S&P div yield
4	83*	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio		S&P PE ratio

A.2 Regression of individual sentiment proxies on fundamental variables

Table A.2

This table presents the decomposition results of six individual sentiment proxies in the following regression: $x_{i,t} = a + b^{'}(Z_t) + \kappa_t$, where $x_{i,t}$ represents one of the six individual sentiment proxies, which are close-end fund discount rate (CEFD), share turnover (TURN), number of IPOs (NIPO), first-day returns of IPOs (RIPO), dividend premium (PDND) and equity share in new issues (EQTI), Z_t is the 14 monthly fundamental variables described in Section III and κ_t is the regression residual. Turnover, the average monthly first-day return, and the dividend premium are lagged one year relative to the other three measures. We report the regression coefficients estimates, ols t-statistics, Newey-West t-statistics with 12 lags and the R-squares in the decomposition.

	CEFD			TURN			NIPO			RIPO			PDND			EQTI		
	coef.	OLS t-stat	NW t-stat	coef.	OLS t-stat	NW t-stat	coef.	OLS t-stat	NW t-stat	coef.	OLS t-stat	NW t-stat	coef.	OLS t-stat	NW t-stat	coef.	OLS t-stat	NW t-stat
fundamental variables																		
Intercept	8.92	6.81	2.85	-0.15	-4.89	-2.02	20.89	6.70	2.91	19.07	6.88	2.93	4.17	2.72	0.99	-0.04	-3.01	-1.46
Output and income	0.28	0.91	0.83	0.00	1.00	0.95	-0.09	-0.24	-0.23	0.03	0.09	0.08	0.21	1.10	1.03	0.00	-0.07	-0.06
Labour market	0.21	0.50	0.28	0.00	0.57	0.28	2.80	5.76	3.69	1.50	3.40	2.81	-1.85	-7.55	-4.74	0.00	1.74	1.38
Housing	-1.90	-6.62	-2.77	0.03	8.45	4.07	0.32	0.87	0.35	-0.46	-1.36	-1.06	1.01	5.40	2.34	-0.01	-3.27	-1.86
Consumption, orders and inventories	0.74	1.98	1.13	-0.01	-1.75	-1.00	-0.92	-1.31	-0.87	-2.75	-4.38	-3.22	2.14	6.16	3.58	0.01	2.77	1.73
Money and credit	0.12	0.56	0.82	0.01	2.06	2.22	-0.34	-0.70	-1.18	-0.54	-1.22	-2.37	0.41	1.68	2.38	0.00	0.37	0.18
Exchange rates	0.17	0.75	0.78	-0.01	-0.97	-0.95	-0.17	-0.30	-0.33	-0.26	-0.49	-0.53	0.03	0.12	0.12	0.00	0.53	0.39
Inflation	-0.05	-0.20	-0.47	0.00	0.69	1.17	0.00	0.02	0.03	0.27	1.07	1.66	-0.10	-0.71	-1.37	0.00	-1.70	-1.78
Consumption wealth ratio	0.16	0.57	0.19	2.64	5.07	2.32	79.22	1.54	0.60	15.29	0.33	0.18	-128.99	-4.97	-1.99	-1.77	-8.29	-3.87
GDP growth	1.24	3.82	2.77	0.01	1.87	1.43	-0.94	-3.20	-1.88	1.44	5.47	3.67	0.13	0.87	0.70	0.00	1.82	1.52
3m Treasury Bill	-5.05	-9.59	-3.24	0.04	6.61	3.03	6.66	10.84	4.49	1.87	3.42	1.51	-5.97	-19.73	-7.43	0.02	8.21	5.08
Default spread	1.86	5.39	2.10	0.14	5.05	2.03	-3.84	-1.36	-0.56	4.26	1.67	0.90	2.53	1.79	0.72	0.06	4.99	3.01
Term spread	-3.72	-10.52	-3.89	0.00	-0.28	-0.14	8.58	8.40	3.74	-1.26	-1.38	-1.09	-5.25	-10.39	-4.83	0.02	5.14	3.24
Dividend yield	-0.14	-0.28	-0.09	0.02	1.28	0.56	-7.56	-4.73	-2.36	-6.08	-4.24	-1.25	1.83	2.31	0.71	0.00	0.54	0.36
Zero return ratio	6.46	13.56	4.82	-0.01	-7.68	-4.16	-0.35	-3.02	-1.18	-0.21	-2.04	-0.91	0.86	14.91	6.10	0.00	-1.51	-0.74
R-square			46.10%			36.96%			30.56%			17.33%			54.65%			44.01%
adj R-square			44.79%			35.43%			28.88%			15.33%			53.55%			42.65%

A.3 Description of characteristics-based portfolios

- Market equity (ME): Constructed as price times shares outstanding from CRSP in the June prior to t. Size is the log of market equity.
- Age: Measured as the number of years between the firms first appearance on CRSP and t.
- Total risk (sigma): Measured as the annual standard deviation in monthly returns from CRSP for the 12 months ending in the June prior to t.
- Earnings to book equity (E/BE): A profitability measure defined as earnings divided by the book value of equity. Earnings (E) is defined as income before extraordinary items (Item 18) plus income statement deferred taxes (Item 50) minus preferred dividends (Item 19). Book equity (BE) is defined as shareholders equity (Item 60) plus balance sheet deferred taxes (Item 35).
- Dividends to book equity (D/BE): Measured as dividends per share times shares outstanding divided by the book value of equity. Dividends (D) are equal to dividends per share at the ex date (Item 26) times shares outstanding (Item 25). Book equity (BE) is defined as shareholders equity (Item 60) plus balance sheet deferred taxes (Item 35).
- Plant, property, and equipment to total asset (PPE/A): A measure of asset tangibility constructed as the ratio of property, plant and equipment (Item 7) to total assets.
- Research and development expense to total assets (RD/A): Another measure of asset tangibility constructed as the ratio of research and development expense (Item 46) to total assets.
- Book equity to market equity (BE/ME): A proxy for either growth opportunities or distress constructed as the log of the ratio of book equity to market equity.
- External finance to assets (EF/A): Measured as external finance divided by assets. External finance (EF) is equal to the change in assets (Item 6) less the change in retained earnings

(Item 36). When the change in retained earnings is not available we use net income (Item 172) less common dividends (Item 21) instead.

- Growth in sales (GS): A measure of growth opportunities defined as the change in net sales divided sales of the previous year. Sales growth decile is formed using NYSE breakpoints for sales growth. Sales growth is the percentage change in net sales (Item 12).

Reference

Ang, A., and G. Bekaert. 2007. Return predictability: Is it there? *Review of Financial Studies* 20, 651–707.

Antoniou, C., J. A. Doukas, and A. Subrahmanyam. 2013. Cognitive dissonance, sentiment, and momentum. *Journal of Financial and Quantitative Analysis* 48, 245–275.

Antoniou, C., J. A. Doukas, and A. Subrahmanyam. 2015. Investor Sentiment, Beta, and the Cost of Equity Capital. *Management Science* 1–21.

Arif, S., and Charles M.C. Lee. 2014. Aggregate investment and investor sentiment. *Review of Financial Studies* 27(11), 3241–3279.

Badertscher, B., D. Shanthikumar, and S. Teoh. 2016. Private firm investment and public peer misvaluation. Working paper.

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

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

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

Ben-Rephael, A., Kandel, S. and Wohl, A. 2012. Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics* 104, 363–382.

Brown, G. W., and Cliff, M. T. 2005. Investor sentiment and asset valuation. *Journal of Business* 78(2), 405–40.

Campbell, J. Y. 1987. Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.

Campbell, J. Y., and R. J. Shiller. 1988a. Stock prices, earnings and expected dividends. *Journal of Finance* 43, 661–676.

Campbell, J. Y., and R. J. Shiller. 1988b. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.

Chen, N., Roll, R. and Ross, S. A. 1986. Economic forces and the stock market. *Journal of Business* 59(3), 383–403.

Chen, Y., Han, B. and Pan, J. 2016. Sentiment risk, sentiment timing, and hedge fund returns. Working paper.

Cochrane, J. H. 2008. The dog that did not bark: A defense of return predictability. *Review of Financial Studies* 21, 1533–1575.

Cochrane, J. H. 2011. Presidential address: Discount rates. *Journal of Finance* 66, 1047–1108.

Da, Z., Engelberg, J. and Gao, P. 2015. The sum of ALL FEARS Investor sentiment and asset prices. *Review of Financial Studies* 28, 1–32.

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

Devault, L., Sias, R. and Starks, L. 2016. Who are the sentiment traders? Evidence from the cross-section of stock returns and demand. Working paper.

Edmans, A., Garía, D. and Norli, O. 2007. Sports sentiment and stock returns. *Journal of Finance* 62, 1967–1998.

Fama, E. F., and K. R. French. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–25.

Gao, Z., Ren, H. and Zhang, B. 2016. Googling investor sentiment around the world. Working paper.

Garía, D. 2013. Sentiment during recessions. *Journal of Finance* 68, 1267–1300.

Greenwood, R. and Shleifer, A. 2014. Expectations of returns and expected returns. *Review of Financial Studies* 27(3), 714–746.

Goetzmann, W. N., Kumar, A., Kim, D. and Wang, Q. 2014. Weather-induced mood, institutional investors, and stock returns. *Review of Financial Studies* 28(1), 73–111.

Goyal, A., and I. Welch. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455–508.

Goyenko, R. Y., Holden, C. W. and Trzcinka, C. A. 2009. Do liquidity measures measure liquidity?

Journal of Financial Economics 92, 153–181.

Han, B. 2008. Investor sentiment and option prices. *Review of Financial Studies* 21(1), 387–414.

Hirshleifer, D. 2001. Investor psychology and asset pricing. *Journal of Finance* 56(4), 1533–1597.

Hirshleifer, D. and Shumway, T. 2003. Good day sunshine: Stock returns and the weather. *Journal of Finance* 58, 1009–1032.

Hodrick, R. J. 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5(3), 357–386.

Hribar, P. and McInnis, J. 2007. Investor sentiment and analysts' earnings forecast errors. *Management Science* 58(2), 293–307.

Huang, D., Jiang, F., Tu, J. and Zhou, G. 2015. Investor sentiment aligned: A powerful predictor of stock returns *Review of Financial Studies* 28, 791–837.

Jiang, F., Lee, J., Martin, X. and Zhou, G. 2015. Manager sentiment and stock returns. Working paper.

Kahneman, D., and A. Tversky. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124–1131.

Kamstra, M. J., Kramer, L. A. and Levi, M. D. 2003. Winter blues: A SAD stock market cycle. *American Economic Review* 93, 324–343.

Kaplanski, G., Levy, H 2010a. Exploitable predictable irrationality: The FIFA World Cup effect on the U.S. stock market. *Journal of Financial and Quantitative Analysis* 45(2), 535–553.

Kaplanski, G., Levy, H 2010b. Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics* 95(2), 174–201.

Kaplanski, G., Levy, H., Veld, C. and Veld-Merkoulova, Y. 2015. Do happy people make optimistic investors? *Journal of Financial and Quantitative Analysis* 50, 145–168.

Kelly, B., and S. Pruitt. 2015. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics* 186(2), 294–316.

Kelly, B., and S. Pruitt. 2013. Market expectations in the cross-section of present values. *Journal of Finance*, forthcoming.

Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.

Kumar, A. and Lee, C. 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61(5), 2451–2486.

Kyle, J., Ludvigson, S. C., S. Ng. 2015. Measuring uncertainty. *American Economic Review* 105(3), 1177–1216.

Lee, C., Shleifer, A. and Thaler, R. H. 1991. Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46, 75–109.

Lee, K. 2011. The world price of liquidity risk. *Journal of Financial Economics* 99, 136–161.

Lemmon, M., and E. Portniaquina. 2006. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies* 19, 1499–1529.

Lettau, M., and S. Ludvigson. 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109, 1238–1287.

Ludvigson, S. C., S. Ng. 2007. The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economcis* 83, 171–222.

McCracken, M. W., S. Ng. 2015. FRED-MD: A monthly database for macroeconomic research. Working paper, Federal Reserve Bank of St. Louis.

Mian, G. M. and Sankaraguruswamy, S. 2012. Investor sentiment and stock market response to earnings News. *The Accounting Review* 87(4), 1357–1384.

Neal, R., and S. Wheatley. 1998. Do measures of investor sentiment predict stock returns? *Journal of Financial and Quantitative Analysis* 34, 523–547.

Papapostolou, N. C., Pouliasis, P. K., Nomikos, N. K. and Kyriakou, I. 2015. Real assets investor sentiment and the predictability of stock returns. Working paper, City University London.

Rhodes-Kropf, M., Robinson, R.M.D. and Viswanathan, S. 2005. Valuation waves and merger activity: the empirical evidence. *Journal of Financial Economics* 77, 561–603.

Ritter, J. R., and I. Welch. 2002. A Review of IPO Activity, Pricing, and Allocations. *Journal of Finance* 57, 1795–1828.

Saunders, E. M. 1993. Stock prices and wall street weather. *American Economic Review* 83, 1337–1345.

Schwarz, G. W. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2), 461–464.

Shiller, R. J. 1981. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review* 71, 421–436.

Shiller, R. J. 2000. Irrational Exuberance *Princeton UP, Princeton*.

Shleifer, A. and R. W. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52, 35–55.

Shleifer, A. 2000. Inefficient markets. *Oxford UP, Oxford*

Sibley, S., Y. Wang, Y. Xing, and X. Zhang. 2016. The information content of the sentiment index. *Journal of Banking and Finance* 62, 164–179.

Sohn, B. S. 2013. What does investor sentiment reflect: Animal spirits or risks? Working paper.

Spiegel, M. 2008. Forecasting the equity premium: Where we stand today. *Review of Financial Studies* 21, 1453–1454.

Stambaugh, R. F. 1999. Predictive regressions. *Journal of Financial Economics* 54, 375–421.

Stambaugh, R. F., J. Yu, and Y. Yuan. 2012. The short of it: Investor sentiment and anomalies. *Journal of Financial Economics* 104, 288–302.

Swaminathan, B. 1996. Time-varying expected small firm returns and closed-end fund discounts. *Review of Financial Studies* 9, 845–887.

Wold, H. 1966. Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiaah (eds.), *Multivariate Analysis*, 391–420. New York: Academic Press.

Wold, H. 1975. Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Cappecchi (eds.), *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*. New York: Academic Press.

Yu, J., and Y. Yuan. 2011. Investor sentiment and the mean-variance relation. *Journal of Financial Economics* 100, 367–381.

Figure 1. Time series plot of BW index and purged investor sentiment index.

This figure plots Baker and Wurgler's investor sentiment index and purged sentiment index from July 1965 to November 2014. The solid black line depicts purged investor sentiment index, while the blue dashed line plots BW index. Both series are normalized to have zero mean and a standard deviation of one.

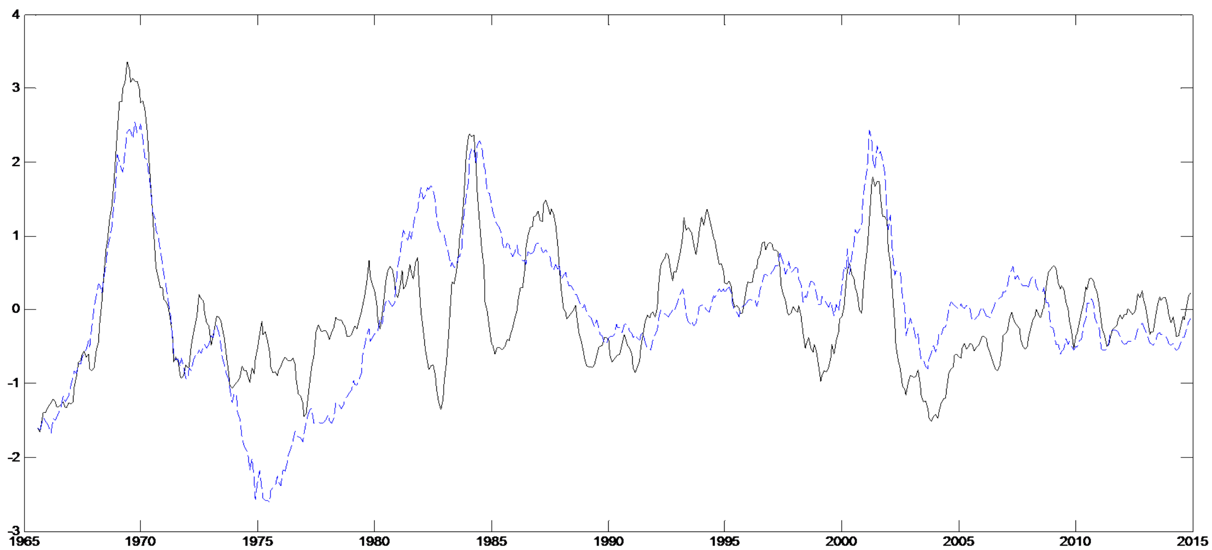

Figure 2. Individual investor sentiment proxies.

The figures plot raw investor sentiment proxies and purged investor sentiment proxies. In each panel, the solid black line depicts purged investor sentiment proxies, while the blue dashed line plots raw investor sentiment proxies. The first panel shows the value-weighted average difference between the net asset values of closed-end stock mutual fund shares and market prices. The second panel shows detrended log turnover. Turnover is the ratio of reported share volume to average shares listed from the NYSE Fact Book. We detrend using the past five-year average. The third panel shows the monthly number of initial public offerings. The fourth panel shows the average monthly first-day returns of initial public offerings. The fifth panel shows the log ratio of the value-weighted average market-to-book ratios of dividend payers and nonpayers. The sixth panel shows gross monthly equity issuance divided by gross monthly equity plus debt issuance. Sample period is over July 1965 to November 2014.

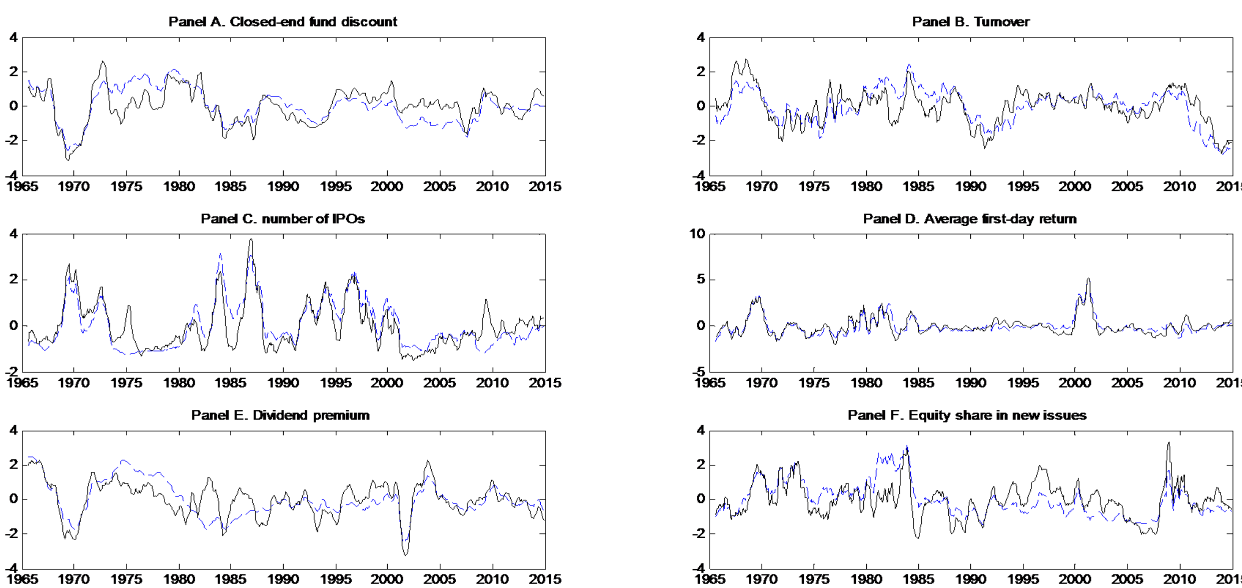

Figure 3. Two-way sorts: Future returns by sentiment index and firm characteristics.

For each month, we form 3 portfolios (i.e., high, medium and low) according to the NYSE breakpoints of firm size (ME), age, total risk, fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth (GS). We also calculate portfolio returns for profitable and unprofitable firms, as well as dividend payers and nonpayers. The solid bars are returns following positive sentiment periods, and the clear bars are returns following negative sentiment periods. The dashed line is the average across both periods and the solid line is the difference. Positive sentiment periods are periods when monthly average of Baker and Wurgler's sentiment during the prior year is positive. Sample period is over Augest 1965 to December 2014.

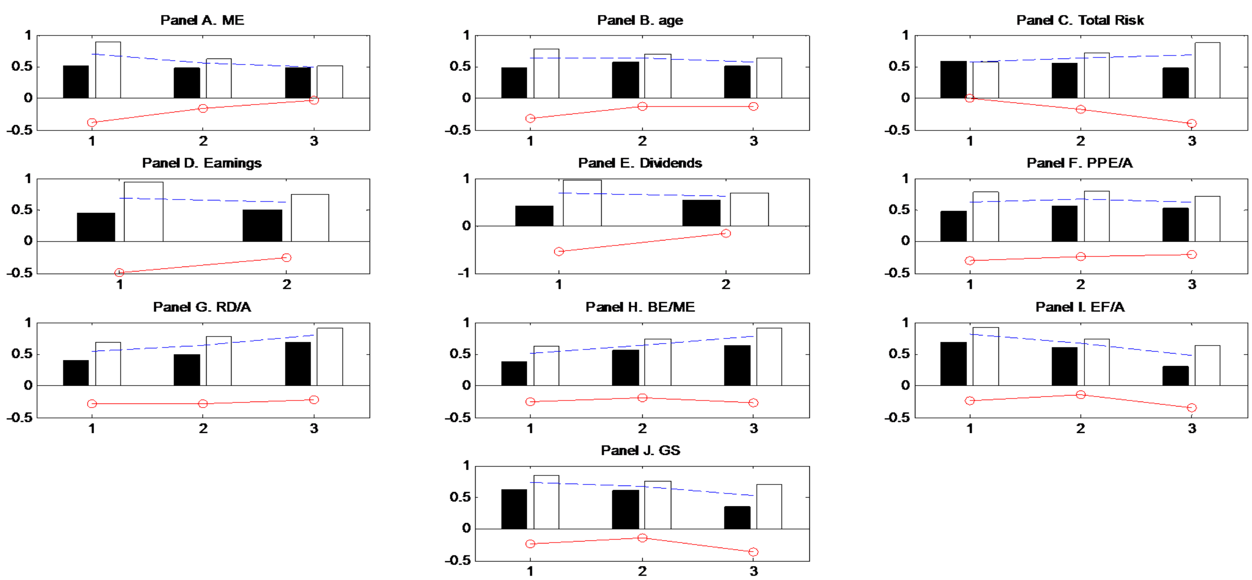

Figure 4. NBER indicator of peaks and troughs and purged sentiment indictor

This figure plots NBER indicator of peaks and troughs (the solid bars) and contemporaneous purged sentiment indictor (the clear bars). NBER indicator equals one when the economy is at peak and equals zero when it is at trough. Sentiment indicator equals one when sentiment is above the medium level and equals zero when it is below the medium level.

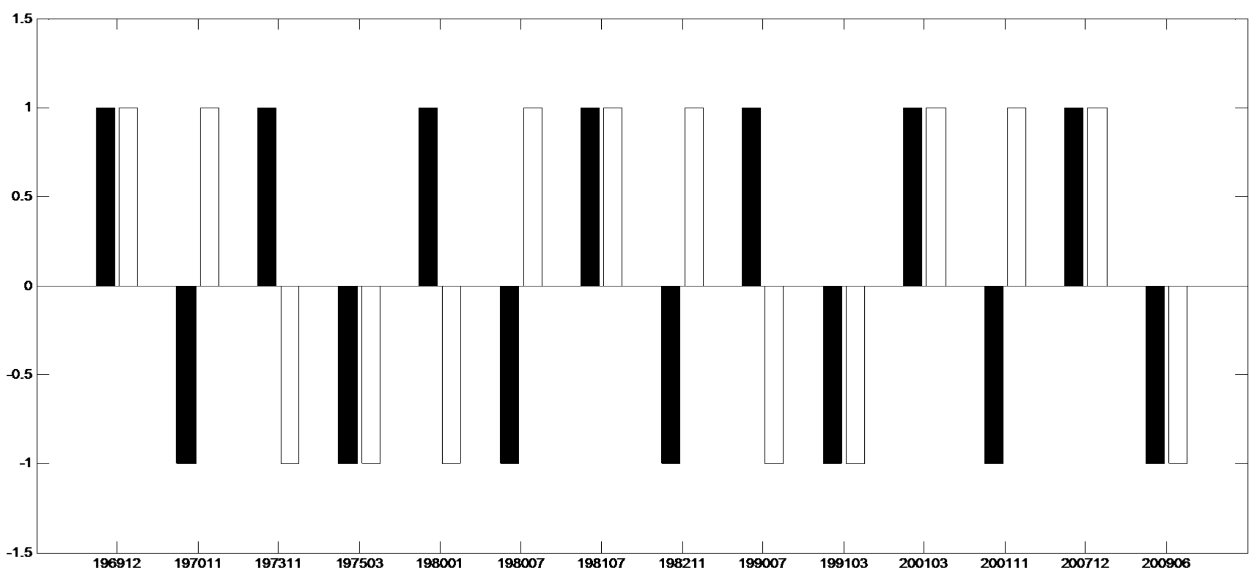

Table 1 Summary statistics of fundamental variables and decomposition

This table presents summary statistics for BW index and fundamental variables, as well as decomposition results. Panel A provides summary statistics for Baker and Wurgler's investor sentiment index (BW) and 14 fundamental variables. For each variable, we report the means, standard deviations, first-order autocorrelations ($\rho(1)$) and their correlations with investor sentiment index (BW index). The 14 fundamental variables are: the first principle components from seven categories of macroeconomic variables (i.e., (1) output and income, (2) employment, (3) housing, (4) consumption, orders and inventories, (5) money and credit, (6) exchange rates, (7) inflation.), consumption-to-wealth ratio, GDP growth, three-month Treasury Bill rate, default spread, term spread, dividend yield and zero return ratio. Our sample period is from July 1965 to November 2014. All variables are measured at monthly frequency. Panel B presents the decomposition results of BW index in the following regression: $BW_t = a + b^{'}(Z_t) + \kappa_t$, where Z_t is the 14 monthly fundamental variables in Panel A and κ_t is the regression residual. We report the regression coefficients estimates, ols t-statistics, Newey-West t-statistics adjusted for 12 lags and the R-square in the decomposition.

Panel A. Summary statistics of BW index and fundamentals

	mean	std	$\rho(1)$	corr with BW	p-value	Source
BW	0.00	1.00	0.992	1.00	0.00	Wurgler's website
$BW^{\perp}$	0.00	1.00	0.987	0.97	0.00	Wurgler's website
Output and income	-0.05	2.95	0.357	-0.07	0.08	FRED-MD
Employment	-0.06	3.19	0.800	-0.11	0.01	FRED-MD
Housing	-0.01	2.88	0.983	0.10	0.02	FRED-MD
Consumption, orders and inventories	-0.06	1.98	0.788	-0.18	0.00	FRED-MD
Money and credit	0.00	1.71	-0.208	0.01	0.81	FRED-MD
Exchange rates	0.00	1.47	0.320	0.06	0.14	FRED-MD
Inflation	0.00	2.90	-0.212	-0.01	0.79	FRED-MD
Consumption wealth ratio	0.00	0.02	0.968	0.16	0.00	Ludvigson's website
GDP growth	6.72	4.13	0.825	-0.21	0.00	U.S. Dept. of Commerce: Bureau of Economic Analysis
3m Treasury Bill	5.06	3.19	0.990	0.23	0.00	Board of Governors of the Federal Reserve System
Default spread	1.06	0.46	0.965	0.10	0.02	Board of Governors of the Federal Reserve System
Term spread	1.60	1.29	0.957	-0.01	0.85	Board of Governors of the Federal Reserve System
Dividend yield	3.04	1.18	0.992	-0.09	0.03	CRSP
Zero return ratio	23.64	15.43	0.995	-0.18	0.00	CRSP

Panel B. Decomposition of BW index				
	coef.	OLS t-stat	NW t-stat	
fundamental variables				
Intercept	-0.52	-5.46	-2.21	
Output and income	-0.03	-2.32	-2.44	
Employment	0.04	2.60	1.30	
Housing	0.04	3.94	1.55	
Consumption, orders and inventories	-0.02	-1.12	-0.59	
Money and credit	0.00	0.06	0.08	
Exchange rates	0.00	0.06	0.06	
Inflation	0.00	0.48	1.03	
Consumption wealth ratio	4.62	2.97	1.12	
GDP growth	-0.06	-6.51	-3.61	
3m Treasury Bill	0.46	24.73	9.21	
Default spread	-0.17	-1.96	-0.75	
Term spread	0.46	15.04	7.29	
Dividend yield	-0.15	-3.16	-1.13	
Zero return ratio	-0.07	-18.80	-5.30	
R^2				62.51%
Adj. R^2				61.60%

Table 2 Summary statistics of sentiment proxies

This table reports summary statistics for raw investor sentiment proxies and purged sentiment proxies. Panel A presents the means, standard deviations, first-order autocorrelations ($\rho(1)$), minimums, maximums of the six raw sentiment proxies, their correlations with BW sentiment index and their correlation matrix. The first sentiment proxy (CEFD) is the value-weighted average difference between the net asset values of closed-end stock mutual fund shares and market prices. The second sentiment proxy (TURN) is detrended natural log turnover. Turnover is the ratio of reported share volume to average shares listed from the NYSE Fact Book. We detrend using the past five-year average. The third sentiment proxy (NIPO) is the monthly number of initial public offerings. The fourth sentiment proxy (RIPO) is the average monthly first-day returns of initial public offerings. The fifth sentiment proxy (PDND) is the log difference of the value-weighted average market-to-book ratios of dividend payers and nonpayers. The sixth sentiment proxy (EQTI) is gross monthly equity issuance divided by gross monthly equity plus debt issuance. Turnover, the average monthly first-day IPO return, and the dividend premium are lagged one year relative to the other three measures. In Panel B, we present summary statistics and correlations of the six purged individual sentiment proxies. Purged sentiment proxies are the residuals from regressing corresponding sentiment proxies on 14 fundamental variables defined the same as in Table 1. Superscripts a, b, and c denote statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A. Summary statistics and correlations of raw sentiment proxies

	Mean	Std	$\rho(1)$	min	max	correlation with BW	correlations CEFD	TURN	NIPO	RIPO	PDND	EQTI
CEFD	0.00	1.00	0.965	-2.77	2.29	-0.67^a	1.00					
TURN	0.00	1.00	0.814	-2.91	3.08	0.47^a	-0.12^a	1.00				
NIPO	0.00	1.00	0.862	-1.13	4.07	0.49^a	-0.29^a	0.19^a	1.00			
RIPO	0.00	1.00	0.653	-2.37	5.25	0.46^a	-0.17^a	0.27^a	0.17^a	1.00		
PDND	0.00	1.00	0.976	-3.15	2.50	-0.90^a	0.58^a	-0.32^a	-0.43^a	-0.39^a	1.00	
EQTI	0.00	1.00	0.744	-1.46	4.22	0.13^a	0.22^a	0.21^a	0.26^a	0.11^a	-0.04	1.00

Panel B. Summary statistics and correlations of purged sentiment proxies

	Mean	Std	$\rho(1)$	min	max	correlation with BW	correlations CEFDres	TURNres	NIPOres	RIPOres	PDNDres	EQTIres
CEFDres	0.00	1.00	0.898	-3.11	2.80	-0.40^a	1.00					
TURNres	0.00	1.00	0.709	-2.88	2.98	0.19^a	-0.01	1.00				
NIPOres	0.00	1.00	0.756	-1.96	4.83	0.23^a	-0.24^a	0.00	1.00			
RIPOres	0.00	1.00	0.565	-2.04	5.33	0.30^a	-0.07	0.22^a	0.13^a	1.00		
PDNDres	0.00	1.00	0.841	-3.68	2.78	-0.53^a	0.30^a	-0.14^a	-0.14^a	-0.40^a	1.00	
EQTIres	0.00	1.00	0.512	-2.24	5.30	0.09^b	0.09^b	0.03	0.37^a	0.09^b	0.05	1.00

Table 3 The properties of the characteristics-based portfolios

This table reports the properties of characteristics-based portfolios. The sample period spans from August 1965 to December 2014. Panel A summarizes six groups of variables: the returns variables, the size, age, and risk characteristics, profitability variables, dividend variables, tangibility measures and variables used as proxies for growth opportunities and distress. Returns are measured monthly. For the four groups of variables regarding profitability, dividend policy, tangibility and growth opportunities and distress, accounting data from the fiscal year ending in t-1 are matched to monthly returns from July of year t through June of year t+1. All variables are winsorized at 99.5 and 0.5%. Panel B reports the means and t-statistics of excess returns for the 16 characteristics based portfolios: firm size (ME), age, total risk (sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). We also construct a combination strategy of the 16 portfolios. Panel C reports the means and t-statistics of benchmark-adjusted returns for the characteristics based portfolios. Benchmark-adjusted average returns are estimates of a_i in the following regression: $R_{i,t} = a_i + bMKT_t + cSMB_t + dHML_t + eWML_t + u_{i,t}$ where $R_{i,t}$ is excess return in month t for one of the characteristics based portfolios. All t-statistics are based on the heteroskedasticity consistent standard errors of White (1980).

Panel A. Summary statistics

	Full Sample				
	N	Mean	SD	Min	Max
			Returns		
$R_t(\%)$	2185371	1.30	17.55	-98.13	2400.00
			Size, Age and Risk		
ME_{t-1}(\$M)	2185371	1157	4491	1	45205
Age_t(Years)	2185371	13.93	14.23	0.17	73.83
$sigma_{t-1}(\%)$	2184127	13.49	8.83	2.52	60.70
			Profitability		
$E+/BE_{t-1}(\%)$	2185371	10.22	10.26	0.00	68.97
			Dividend Policy		
$D/BE_{t-1}(\%)$	2184513	2.13	3.45	0.00	26.15
			Tangibility		
$PPE/A_{t-1}(\%)$	1969977	53.22	37.96	0.00	193.78
$RD/A_{t-1}(\%)$	2185359	3.27	8.22	0.00	62.60
			Growth Opportunities and Distress		
$BE/ME_{t-1}(\%)$	2185371	0.87	0.73	0.03	5.17
$EF/A_{t-1}(\%)$	2157062	10.04	23.06	-69.27	123.86
GS_{t-1}(decile)	2126840	5.74	3.12	1.00	10.00

Panel B. Excess returns: Long-short

	ME	Age	Sigma	E/BE	D/BE	PPE/A	RD/A	BE/ME	EF/A	GS	BE/ME M-L	EF/A M-H	GS M-H	BE/ME M-H	EF/A M-L	GS M-L	Combine
Long leg (mean)	0.59	0.75	0.75	0.85	0.82	0.84	0.68	1.14	1.21	1.06	0.89	0.93	0.95	0.89	0.93	0.95	0.89
Short leg (mean)	1.01	0.86	0.95	0.98	0.98	0.85	1.18	0.61	0.54	0.65	0.61	0.54	0.65	1.14	1.21	1.06	0.86
Long minus short (mean)	-0.42	-0.11	-0.20	-0.13	-0.16	-0.01	-0.50	0.53	0.67	0.41	0.28	0.39	0.30	-0.26	-0.28	-0.12	0.03
Long leg (t-statistics)	2.88	3.86	4.99	3.89	4.16	4.14	2.88	5.11	5.08	4.23	3.91	4.41	4.59	3.91	4.41	4.59	4.28
Short leg (t-statistics)	3.84	3.27	3.07	2.86	3.23	2.99	3.69	2.22	2.01	2.48	2.22	2.01	2.48	5.11	5.08	4.23	3.18
Long minus short (t-statistics)	-2.52	-0.82	-0.94	-0.71	-0.97	-0.05	-3.11	4.51	9.63	5.25	3.01	4.84	3.49	-4.94	-5.22	-1.39	0.28

Panel C. Benchmark-adjusted returns: Long-short

	ME	Age	Sigma	E/BE	D/BE	PPE/A	RD/A	BE/ME	EF/A	GS	BE/ME M-L	EF/A M-H	GS M-H	BE/ME M-H	EF/A M-L	GS M-L	Combine
Long leg (mean)	0.16	0.12	0.24	0.19	0.14	0.21	0.03	0.38	0.50	0.38	0.22	0.29	0.30	0.22	0.29	0.30	0.25
Short leg (mean)	0.29	0.24	0.24	0.24	0.25	0.24	0.65	0.12	-0.05	0.05	0.12	-0.05	0.05	0.38	0.50	0.38	0.23
Long minus short (mean)	-0.13	-0.11	0.00	-0.06	-0.11	-0.04	-0.62	0.26	0.56	0.34	0.10	0.34	0.26	-0.16	-0.21	-0.08	0.02
Long leg (t-statistics)	3.43	2.52	4.40	3.76	2.74	2.46	0.37	4.73	6.29	3.83	3.54	5.58	5.51	3.54	5.58	5.51	4.89
Short leg (t-statistics)	2.90	2.59	1.98	1.35	2.17	2.40	4.71	1.30	-0.59	0.61	1.30	-0.59	0.61	4.73	6.29	3.83	2.42
Long minus short (t-statistics)	-1.14	-1.16	0.02	-0.35	-0.90	-0.39	-5.24	3.62	10.13	4.96	1.61	5.74	4.86	-3.52	-4.68	-1.03	0.36

Table 4 Correlations of Long-short portfolio returns

This table reports correlations among characteristics-based portfolios. The long−short portfolios are formed based on firm characteristics: firm size (ME), age, total risk (sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. We also construct a combination strategy of the 16 portfolios. Superscripts a, b, and c denote statistical significance at the 1%, 5%, and 10% level, respectively.

		Size, Age and Risk			Profitability, Dividends		Tangibility		Growth Opportunities and Distress			Growth Opportunities			Distress			
		ME	Age	Sigma	E/BE	D/BE	PPE/A	RD/A	BE/ME	EF/A	GS	BE/ME M-L	EF/A M-H	GS M-H	BE/ME M-H	EF/A M-L	GS M-L	Combine
ME	H-L	1.00																
Age	H-L	0.81^a	1.00															
Sigma	L-H	0.69^a	0.87^a	1.00														
E/BE	$>0-<0$	0.73^a	0.82^a	0.84^a	1.00													
D/BE	$>0-=0$	0.75^a	0.91^a	0.94^a	0.89^a	1.00												
PPE/A	H-L	0.61^a	0.79^a	0.77^a	0.65^a	0.78^a	1.00											
RD/A	L-H	0.45^a	0.73^a	0.70^a	0.61^a	0.72^a	0.74^a	1.00										
BE/ME	H-L	0.05	0.49^a	0.55^a	0.36^a	0.53^a	0.51^a	0.64^a	1.00									
EF/A	L-H	0.08^c	0.45^a	0.50^a	0.30^a	0.48^a	0.51^a	0.53^a	0.71^a	1.00								
GS	L-H	−0.21^a	0.00	0.08^b	−0.22^a	0.04	0.22^a	0.15^a	0.42^a	0.66^a	1.00							
BE/ME	M-L	0.24^a	0.65^a	0.68^a	0.57^a	0.70^a	0.61^a	0.72^a	0.91^a	0.66^a	0.28^a	1.00						
EF/A	M-H	0.53^a	0.80^a	0.87^a	0.74^a	0.86^a	0.75^a	0.67^a	0.66^a	0.76^a	0.32^a	0.76^a	1.00					
GS	M-H	0.46^a	0.76^a	0.85^a	0.65^a	0.83^a	0.72^a	0.69^a	0.70^a	0.75^a	0.47^a	0.76^a	0.90^a	1.00				
BE/ME	M-H	0.32^a	0.02	−0.05	0.18^a	0.03	−0.08^c	−0.19^a	−0.67^a	−0.44^a	−0.47^a	−0.30^a	−0.15^a	−0.24^a	1.00			
EF/A	M-L	0.71^a	0.64^a	0.67^a	0.72^a	0.68^a	0.47^a	0.33^a	0.08^c	−0.15^a	−0.39^a	0.29^a	0.53^a	0.39^a	0.34^a	1.00		
GS	M-L	0.66^a	0.76^a	0.78^a	0.86^a	0.79^a	0.52^a	0.56^a	0.30^a	0.13^a	−0.46^a	0.50^a	0.61^a	0.57^a	0.20^a	0.75^a	1.00	
Combine		0.72^a	0.93^a	0.95^a	0.87^a	0.97^a	0.85^a	0.81^a	0.63^a	0.58^a	0.13^a	0.78^a	0.91^a	0.88^a	−0.05	0.63^a	0.77^a	1.00

Table 5 The predictability of BW index

This table presents the results of using Baker and Wurgeler's investor sentiment index BW to predict spread, long and short portfolio returns. The sample spans from August, 1965 to December, 2014. The portfolios are formed based on firm characteristics: firm size (ME), age, total risk (Sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. We provide results for the following two regressions respectively:

$$R_{i,t} = a + bBW_{t-1} + u_t$$

$$R_{i,t} = a + bBW_{t-1} + cMKT_t + dSMB_t + eHML_t + fWML_t + u_t$$

Variable $R_{i,t}$ is the time t monthly return on the spread, long or short portfolio. SMB (HML) is not included as a control variable when SMB (HML) is the dependent variable. Both coefficient estimates and one-sided empirical p-values are reported.

		long-short				long leg				short leg			
		no control		FF(t)		no control		FF(t)		no control		FF(t)	
		BW coef	p value	BW coef	p value	BW coef	p value	BW coef	p value	BW coef	p value	BW coef	p value
ME	High-Low	0.562	0.001	0.539	0.001	-0.248	0.143	0.044	0.098	-0.810	0.002	-0.495	0.001
Age	High-Low	0.450	0.001	0.187	0.013	-0.324	0.062	-0.013	0.405	-0.774	0.003	-0.200	0.004
Sigma	Low-High	0.942	0.000	0.502	0.000	-0.085	0.303	0.149	0.004	-1.027	0.001	-0.353	0.000
E/BE	$>0-<0$	0.758	0.000	0.480	0.000	-0.493	0.016	-0.057	0.145	-1.251	0.000	-0.537	0.000
D/BE	$>0-=0$	0.768	0.000	0.447	0.000	-0.334	0.056	0.009	0.405	-1.102	0.000	-0.439	0.000
PPE/A	High-Low	0.363	0.010	0.084	0.220	-0.469	0.014	-0.122	0.058	-0.832	0.003	-0.205	0.008
RD/A	Low-High	0.319	0.040	0.058	0.369	-0.596	0.008	-0.146	0.035	-0.915	0.003	-0.205	0.067
BE/ME	High-Low	0.174	0.111	0.060	0.375	-0.572	0.008	-0.104	0.112	-0.746	0.005	-0.164	0.037
EF/A	Low-High	0.162	0.022	0.074	0.121	-0.623	0.006	-0.136	0.024	-0.784	0.003	-0.210	0.002
GS	Low-High	0.070	0.266	0.016	0.486	-0.695	0.004	-0.182	0.015	-0.766	0.003	-0.199	0.002
BE/ME	Med-Low	0.267	0.003	0.167	0.040	-0.480	0.024	0.002	0.439	-0.746	0.005	-0.164	0.037
EF/A	Med-High	0.374	0.000	0.223	0.000	-0.410	0.035	0.013	0.320	-0.784	0.003	-0.210	0.002
GS	Med-High	0.395	0.000	0.233	0.000	-0.370	0.048	0.034	0.172	-0.766	0.003	-0.199	0.002
BE/ME	Med-High	0.093	0.028	0.107	0.013	-0.480	0.024	0.002	0.439	-0.572	0.008	-0.104	0.112
EF/A	Med-Low	0.213	0.000	0.149	0.000	-0.410	0.035	0.013	0.320	-0.623	0.006	-0.136	0.024
GS	Med-Low	0.325	0.000	0.216	0.001	-0.370	0.048	0.034	0.172	-0.695	0.004	-0.182	0.015
Combination		0.390	0.000	0.201	0.000	-0.435	0.025	-0.034	0.257	-0.825	0.002	-0.235	0.001

Table 6 The predictability of BW′ and BW″

This table presents the results of using residual component in BW index (BW′ or BW″) to predict spread portfolio returns, which is defined as the residual from the decomposing regression

$$BW_t = a + b^{'} ECON_t + e_t$$

where BW_t is Baker and Wurgler's investor sentiment index. In Panel A $ECON_t$ represents fundamental variables in Baker and Wurgler (2006), including growth in industrial production, growth in durable consumption, nondurable consumption and service consumption, growth in employment and NBER recession dummy. The derived residual component in BW index is defined as BW′. In Panel B $ECON_t$ represents 14 fundamental variables, including the first principle components from seven categories of macroeconomic variables ((1) output and income, (2) employment, (3) housing, (4) consumption, orders and inventories, (5) money and credit, (6) exchange rates, (7) inflation), consumption-to-wealth ratio, GDP growth, three-month Treasury Bill rate, default spread, term spread, dividend yield and zero return ratio. The derived residual component of BW index is defined as BW″. The portfolios are formed based on firm characteristics: firm size (ME), age, total risk (Sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. We provide results for the following two regressions respectively:

$$R_{i,t} = a + bS_{t-1} + u_t$$

$$R_{i,t} = a + bS_{t-1} + cMKT_t + dSMB_t + eHML_t + fWML_t + u_t$$

Variable $R_{i,t}$ is the time t monthly return on the spread, long or short portfolio, S_{t-1} represents residual component in investor sentiment BW′ (BW″) at time t-1 in Panel B (C). The sample periods include monthly returns from August, 1965 to December, 2010. SMB (HML) is not included as a control variable when SMB (HML) is the dependent variable. Both coefficient estimates and one-sided p-values are reported.

		long-short				long leg				short leg			
		no control		FF(t)		no control		FF(t)		no control		FF(t)	
Panel A.		BW′		BW′		BW′		BW′		BW′		BW′	
		coef	p value	coef	p value	coef	p value	coef	p value	coef	p value	coef	p value
ME	High-Low	0.581	0.001	0.560	0.001	-0.235	0.114	0.031	0.128	-0.817	0.001	-0.529	0.001
Age	High-Low	0.479	0.000	0.159	0.017	-0.313	0.040	-0.027	0.293	-0.792	0.001	-0.186	0.005
Sigma	Low-High	1.011	0.000	0.510	0.000	-0.055	0.271	0.159	0.005	-1.066	0.000	-0.352	0.000
E/BE	$> 0 - < 0$	0.793	0.000	0.461	0.001	-0.486	0.011	-0.049	0.148	-1.279	0.000	-0.510	0.000
D/BE	$> 0 - = 0$	0.802	0.000	0.422	0.000	-0.318	0.042	0.008	0.455	-1.120	0.000	-0.414	0.000
PPE/A	High-Low	0.396	0.007	0.062	0.268	-0.465	0.007	-0.133	0.026	-0.862	0.001	-0.195	0.009
RD/A	Low-High	0.388	0.020	0.070	0.324	-0.592	0.005	-0.143	0.030	-0.980	0.002	-0.213	0.051
BE/ME	High-Low	0.207	0.069	0.091	0.271	-0.560	0.005	-0.065	0.156	-0.767	0.002	-0.157	0.027
EF/A	Low-High	0.175	0.021	0.074	0.146	-0.627	0.004	-0.127	0.021	-0.803	0.001	-0.201	0.004
GS	Low-High	0.081	0.260	0.024	0.499	-0.702	0.002	-0.169	0.015	-0.783	0.001	-0.193	0.002
BE/ME	Med-Low	0.276	0.005	0.168	0.040	-0.491	0.013	0.011	0.523	-0.767	0.002	-0.157	0.027
EF/A	Med-High	0.391	0.000	0.213	0.000	-0.412	0.020	0.012	0.455	-0.803	0.001	-0.201	0.004
GS	Med-High	0.419	0.000	0.231	0.000	-0.365	0.029	0.037	0.273	-0.783	0.001	-0.193	0.002
BE/ME	Med-High	0.069	0.080	0.077	0.047	-0.491	0.013	0.011	0.523	-0.560	0.005	-0.065	0.156
EF/A	Med-Low	0.215	0.000	0.139	0.000	-0.412	0.020	0.012	0.455	-0.627	0.004	-0.127	0.021
GS	Med-Low	0.338	0.000	0.207	0.001	-0.365	0.029	0.037	0.273	-0.702	0.002	-0.169	0.015
Combination		0.414	0.000	0.190	0.000	-0.431	0.016	-0.033	0.195	-0.844	0.001	-0.224	0.002

Panel B.		BW$''$		BW$''$		BW$''$		BW$''$		BW$''$		BW$''$	
		coef	p value	coef	p value	coef	p value	coef	p value	coef	p value	coef	p value
ME	High-Low	0.209	0.045	0.192	0.147	-0.062	0.718	0.021	0.101	-0.271	0.118	-0.172	0.226
Age	High-Low	0.233	0.101	0.099	0.217	-0.046	0.268	0.044	0.304	-0.279	0.206	-0.055	0.300
Sigma	Low-High	0.397	0.063	0.196	0.091	0.015	0.522	0.080	0.109	-0.382	0.172	-0.116	0.181
E/BE	$> 0 - < 0$	0.192	0.157	0.066	0.325	-0.126	0.225	0.030	0.382	-0.318	0.161	-0.036	0.372
D/BE	$> 0 - = 0$	0.349	0.045	0.191	0.079	-0.040	0.262	0.067	0.249	-0.389	0.136	-0.124	0.144
PPE/A	High-Low	0.166	0.159	0.024	0.432	-0.128	0.324	-0.019	0.457	-0.294	0.226	-0.043	0.388
RD/A	Low-High	0.207	0.210	0.053	0.487	-0.147	0.351	0.008	0.393	-0.354	0.297	-0.045	0.563
BE/ME	High-Low	0.077	0.440	0.035	0.558	-0.165	0.233	0.021	0.481	-0.242	0.331	-0.014	0.600
EF/A	Low-High	0.117	0.082	0.070	0.125	-0.153	0.315	0.024	0.333	-0.270	0.224	-0.046	0.334
GS	Low-High	0.092	0.087	0.058	0.149	-0.188	0.302	0.000	0.415	-0.280	0.176	-0.058	0.209
BE/ME	Med-Low	0.115	0.243	0.074	0.355	-0.127	0.348	0.061	0.193	-0.242	0.331	-0.014	0.600
EF/A	Med-High	0.154	0.058	0.081	0.097	-0.116	0.356	0.035	0.205	-0.270	0.224	-0.046	0.334
GS	Med-High	0.204	0.014	0.123	0.013	-0.076	0.434	0.065	0.081	-0.280	0.176	-0.058	0.209
BE/ME	Med-High	0.038	0.193	0.040	0.153	-0.127	0.348	0.061	0.193	-0.165	0.233	0.021	0.481
EF/A	Med-Low	0.036	0.269	0.011	0.421	-0.116	0.356	0.035	0.205	-0.153	0.315	0.024	0.333
GS	Med-Low	0.112	0.121	0.065	0.222	-0.076	0.434	0.065	0.081	-0.188	0.302	0.000	0.415
Combination		0.169	0.056	0.074	0.121	-0.105	0.356	0.032	0.225	-0.273	0.213	-0.042	0.357

Table 7 The predictability of IS-P

Panel A of Table 7 presents the results of using purged investor sentiment IS-P to predict spread, long and short portfolio returns. The sample periods include monthly returns from August, 1965 to December, 2014. The portfolios are formed based on firm characteristics: firm size (ME), age, total risk (Sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). High is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. We provide results for the following two regressions respectively:

$$R_{i,t} = a + bIS\text{-}P_{t-1} + u_t$$

$$R_{i,t} = a + bIS\text{-}P_{t-1} + cMKT_t + dSMB_t + eHML_t + fWML_t + u_t$$

Variable $R_{i,t}$ is the time t monthly return on the spread, long or short portfolio. SMB (HML) is not included as a control variable when SMB (HML) is the dependent variable. Both coefficient estimates and one-sided empirical p-values are reported. Panel B summarizes number of significance in predicting long-short, long-leg and short-leg portfolio returns using sentiment measures in Table 5, 6 and 7.

		long-short				long leg				short leg			
		no control		FF(t)		no control		FF(t)		no control		FF(t)	
Panel A.		IS-P		IS-P		IS-P		IS-P		IS-P		IS-P	
		coef	p value	coef	p value	coef	p value	coef	p value	coef	p value	coef	p value
ME	High-Low	0.543	0.000	0.514	0.001	-0.359	0.281	-0.014	0.688	-0.902	0.000	-0.527	0.001
Age	High-Low	0.439	0.001	0.203	0.009	-0.443	0.057	0.071	0.083	-0.882	0.001	-0.133	0.051
Sigma	Low-High	0.790	0.000	0.318	0.006	-0.275	0.067	0.056	0.161	-1.064	0.001	-0.263	0.008
E/BE	$>0-<0$	0.397	0.006	0.349	0.003	-0.660	0.012	0.002	0.353	-1.058	0.000	-0.347	0.005
D/BE	$>0-=0$	0.575	0.001	0.239	0.009	-0.509	0.021	0.043	0.161	-1.084	0.000	-0.197	0.030
PPE/A	High-Low	0.500	0.002	0.140	0.047	-0.478	0.021	-0.041	0.411	-0.978	0.002	-0.181	0.018
RD/A	Low-High	0.417	0.049	0.399	0.003	-0.639	0.009	0.001	0.563	-1.056	0.009	-0.398	0.002
BE/ME	High-Low	0.200	0.182	0.163	0.150	-0.648	0.002	-0.031	0.351	-0.848	0.003	-0.194	0.024
EF/A	Low-High	0.163	0.052	0.065	0.607	-0.710	0.002	-0.108	0.029	-0.873	0.002	-0.173	0.055
GS	Low-High	0.181	0.101	-0.068	0.033	-0.709	0.001	-0.184	0.008	-0.890	0.003	-0.117	0.125
BE/ME	Med-Low	0.198	0.039	0.191	0.036	-0.649	0.012	-0.003	0.504	-0.848	0.003	-0.194	0.024
EF/A	Med-High	0.280	0.002	0.149	0.020	-0.592	0.012	-0.024	0.411	-0.873	0.002	-0.173	0.055
GS	Med-High	0.319	0.001	0.114	0.049	-0.570	0.015	-0.003	0.582	-0.890	0.003	-0.117	0.125
BE/ME	Med-High	-0.001	0.752	0.028	0.236	-0.649	0.012	-0.003	0.504	-0.648	0.002	-0.031	0.351
EF/A	Med-Low	0.118	0.007	0.084	0.002	-0.592	0.012	-0.024	0.411	-0.710	0.002	-0.108	0.029
GS	Med-Low	0.138	0.023	0.182	0.002	-0.570	0.015	-0.003	0.582	-0.709	0.001	-0.184	0.008
Combination		0.329	0.001	0.166	0.001	-0.566	0.011	-0.020	0.391	-0.894	0.001	-0.187	0.012

Panel B.			BW	BW′	BW″	IS-P
long-short	no FF	bootstrap p	14	13	3	12
	FF(t)	bootstrap p	11	11	1	13
long leg	no FF	bootstrap p	12	14	0	13
	FF(t)	bootstrap p	4	5	0	2
short leg	no FF	bootstrap p	16	16	0	16
	FF(t)	bootstrap p	14	14	0	10

Table 8 Future macroeconomic variables and purged sentiment index

This table reports the results of the regression of future macroeconomic variables on purged sentiment index (IS-P) in Panel A and BW sentiment index (BW) in Panel B. The dependent variables selected are principle components from four categories of macroeconomic variables respectively: (1) output and income, (2) employment, (3) housing, (4) consumption, orders and inventories. We report the regression slopes, Newey-West t -statistics, as well as R-squares. The sample period is from August, 1965 to December, 2014.

	dependent variable			
explanatory variable	(1)	(2)	(3)	(4)
Panel A. IS-P				
Intercept	-0.05	-0.07	-0.01	-0.06
	[-0.43]	[-0.54]	[-0.09]	[-0.75]
IS-P	-0.02	0.15	-0.02	-0.12
	[-0.15]	[1.17]	[-0.29]	[-1.48]
Number of observations	593	593	593	593
$R^2(\%)$	0.00	0.23	0.01	0.38
Panel B: BW				
Intercept	-0.06	-0.08	0.00	-0.07
	[-0.50]	[-0.66]	[-0.01]	[-0.93]
BW	-0.26	-0.41	0.29	-0.39
	[-1.73]	[-2.38]	[3.28]	[-3.73]
Number of observations	593	593	593	593
$R^2(\%)$	0.77	1.66	0.99	3.83

Table 9 Robustness checks

This table presents results for robustness Checks. Panel A reports the number of significant t-statistics for purged sentiment when predicting spread, long and short portfolio returns, using partial least squares with different orthogonalization variables for individual sentiment proxy decomposition. Panel B report number of significant t-statistics for BW sentiment residual, where BW sentiment residual is derived from decomposing BW sentiment index directly using different orthogonalization variables. Panel C report number of significant t-statistics for alternative purged sentiment, using principle component analysis with different orthogonalization variables for individual sentiment proxy decomposition.

			long-short 16 spread portfolios	long leg 16 spread portfolios	short leg 16 spread portfolios
Panel A.			IS-P	IS-P	IS-P
no FF	bootstrap p	14 variables	12	13	16
		alternative 14 variables	12	14	16
		135 variables	10	15	16
FF(t)	bootstrap p	14 variables	13	2	10
		alternative 14 variables	12	2	12
		135 variables	7	2	12
Panel B.			BW''	BW''	BW''
no FF	bootstrap p	14 variables	3	0	0
		alternative 14 variables	6	0	0
		135 variables	0	0	0
FF(t)	bootstrap p	14 variables	1	0	0
		alternative 14 variables	1	0	0
		135 variables	1	0	0
Panel C.			PCAres	PCAres	PCAres
FF(t)	bootstrap p	14 variables	2	0	2
		alternative 14 variables	4	0	3
		135 variables	3	2	0

Table 10 Earnings announcement returns and purged sentiment index

This table reports the results of the regression of average monthly earnings announcement returns on lagged purged sentiment index:

$$CAR_{X_{it}=H/M/L,t} = a + b * IS\text{-}P_{t-1} + \varepsilon_t$$

where $CAR_{X_{it}=H/M/L,t}$ is the average of CARs around quarterly earnings announcements of each characteristic portfolio in month t, $IS\text{-}P_{t-1}$ is our purged sentiment measure in month t-1. The portfolios are formed based on firm characteristics: firm size (ME), age, total risk (Sigma), profitability (E/BE), dividends (D/BE), fixed assets (PPE/A), research and development (RD/A), book-to-market ratio (BE/ME), external finance over assets (EF/A), and sales growth decile (GS). Portfolios formed on earnings are divided into two groups: unprofitable firms and profiable firms. Portfolios formed on dividends are divided into two groups: non dividend paying firms and dividend paying firms. For other firm characteristics, high is defined as a firm in the top three NYSE deciles, low is defined as a firm in the bottom three NYSE deciles, and medium is defined as a firm in the middle four NYSE deciles. We report the regression slopes and heteroskedasticity-robust t-statistics for each characteristic portfolio based on purged sentiment index IS-P and BW sentiment index respectively. The sample period is from January, 1973 to December, 2014.

		IS-P			BW		
		Decile			Decile		
		L	M	H	L	M	H
ME	coef	-0.11	-0.06	0.00	-0.11	-0.04	-0.05
	t-stat	-2.67	-1.82	-0.09	-2.71	-1.09	-1.61
Age	coef	-0.07	-0.09	-0.06	-0.08	-0.06	-0.06
	t-stat	-1.87	-2.50	-1.43	-2.17	-1.65	-1.76
Sigma	coef	0.01	-0.07	-0.13	0.02	-0.08	-0.11
	t-stat	0.47	-2.24	-2.62	0.62	-2.29	-2.42
PPE/A	coef	-0.11	-0.06	-0.05	-0.11	-0.07	-0.04
	t-stat	-2.60	-1.69	-1.28	-2.52	-1.86	-1.15
RD/A	coef	-0.01	-0.17	-0.14	-0.12	-0.18	-0.12
	t-stat	-0.23	-3.04	-2.52	-2.17	-3.49	-2.24
BE/ME	coef	-0.06	-0.10	-0.08	-0.05	-0.07	-0.08
	t-stat	-1.66	-3.10	-1.76	-1.25	-2.20	-1.93
EF/A	coef	-0.07	-0.07	-0.10	-0.03	-0.07	-0.11
	t-stat	-1.74	-2.22	-2.41	-0.80	-2.20	-2.76
GS	coef	-0.10	-0.06	-0.08	-0.09	-0.08	-0.04
	t-stat	-2.25	-2.04	-2.13	-2.00	-2.81	-1.28
		$<= 0$	> 0		$<= 0$	> 0	
E/BE	coef	-0.14	-0.07		-0.17	-0.05	
	t-stat	-1.97	-2.74		-2.17	-1.82	
D/BE	coef	-0.14	-0.05		-0.11	-0.06	
	t-stat	-3.02	-1.68		-2.07	-2.32	

Table 11 Purged sentiment index and mutual fund flow

This table reports regression results of mutual fund flow on sentiment variables. We estimate time-series regressions of the form

$$Flow_t = a + bSENT_t + e_t$$

$$Flow_t = a + bSENT_{t-1} + e_t$$

where the dependent variable Flow in the regressions represents a measure of investor inflows into equity-oriented mutual funds scaled by the aggregate capitalization of the U.S. stock market in each month, SENT represents IS-P (purged sentiment) for Specifications (1) and (2) and BW (BW sentiment index) for Specifications (3) and (4). Specifications (1) and (3) provide relations between mutual fund flow and contemporaneous sentiment variables while Specifications (2) and (4) summarize relations between mutual fund flow and lagged sentiment variables. The regression slopes, Newey-West t -statistics, as well as R^2s are reported. The sample period is from January, 1984 to December, 2014.

	(1)	(2)	(3)	(4)
IS-P(t)	0.23			
	[2.89]			
IS-P(t-1)		0.23		
		[2.79]		
BW(t)			-0.02	
			[-0.37]	
BW(t-1)				-0.04
				[-0.74]
Sample year begin	1984	1984	1984	1984
Sample year end	2014	2014	2014	2014
Number of observations	371	371	371	371
R2(%)	3.21	3.00	0.02	0.10

Table 12 Purged sentiment and non-fundamental Q

This table summarizes the relation between purged sentiment and non-fundamental Q. Panel A presents results for the time-series regressions of purged sentiment IS-P on non-fundamental Q (mQ). Non-fundamental Q is defined as the logarithm of the long-run non-fundamental value to the book value. Panel B presents results for the time-series regressions of combined portfolio returns on non-fundamental Q (mQ). Panel C presents results for the time-series regressions of combined portfolio returns, purged sentiment IS-P and non-fundamental Q (mQ) in a multivariate setting. The regression coefficients, Newey-West t -statistics, as well as R^2s are reported. The sample period is from July, 1965 to December, 2014.

	dependent variable		
explanatory variable	IS-P		
Panel A.			
mQ	0.16		
	[4.90]		
Number of observations	539		
R2(%)	2.61		
		dependent variable	
explanatory variable	L-S combine	L combine	S combine
Panel B.			
mQ	-0.01	0.22	0.24
	[-0.18]	[0.91]	[0.78]
Number of observations	539	539	539
R2(%)	0.00	0.20	0.13
		dependent variable	
explanatory variable	L-S combine	L combine	S combine
Panel C.			
IS-P	0.34	-0.62	-0.96
	[3.71]	[-2.89]	[-3.45]
mQ	-0.07	0.32	0.39
	[-0.89]	[1.30]	[1.26]
Number of observations	539	539	539
R2(%)	2.31	1.65	2.18

Table 13 Economic explaination

This table reports estimation results for the bivariate predictive regressions

$$y_{t+1} = \alpha + \beta IS\text{-}P_t + \varphi dp_t + \omega_{t+1}, y = \Delta d_{t+1} \ or \ dp_{t+1}$$

where dp_{t+1} is the annual log dividend-price ratio for each short leg of characteristics based portfolios, Δd_{t+1} is the annual log dividend-growth rate for each short leg of characteristics based portfolios from July of year t to June of year t+1 (in percentage), $IS\text{-}P_t$ is the purged investor sentiment index in June of year t. Δd_{t+1} and dp_{t+1} are constructed following Cochrane (2008, 2011). We report the regression slopes, Newey-West t-statistics, as well as R^2s. The sample period is from 1965 to 2014.

y_{t+1}	β	t-stat	φ	t-stat	R^2(%)
Panel A. Combination portfolio					
dp	0.04	1.51	0.89	14.54	82.78
Δd(%)	-6.65	-3.38	-12.82	-2.47	27.93
Panel B. Individual portfolio					
ME					
dp	0.05	2.02	0.92	19.78	87.8
Δd(%)	-5.40	-2.38	-9.82	-2.09	14.6
age					
dp	0.04	1.44	0.92	19.93	88.2
Δd(%)	-7.88	-3.49	-8.95	-1.85	18.7
sigma					
dp	0.06	1.67	0.91	16.92	87.3
Δd(%)	-6.61	-2.01	-11.87	-2.24	13.4
D/BE					
dp	-0.02	-0.30	0.68	7.06	50.2
Δd(%)	-17.30	-2.90	-36.82	-3.92	28.2
PPE/A					
dp	0.05	1.78	0.92	20.28	88.8
Δd(%)	-8.02	-2.67	-11.72	-2.04	9.4
RD/A					
dp	0.02	0.59	0.91	16.95	86.4
Δd(%)	-7.73	-2.52	-10.07	-1.94	15.7
BE/ME					
dp	-0.01	-0.18	0.91	19.36	87.7
Δd(%)	-13.62	-3.90	-13.77	-2.90	34.0

Table 14 Purged sentiment and alternative survey based sentiment measures

This table compares purged sentiment IS-P with alternative survey based sentiment measures. Panel A presents results for the time-series regressions of purged sentiment IS-P as dependent variable and alternative survey based sentiment measures as explanatory variable. Panel B presents results for the time-series regressions of L-S combination portfolio returns as dependent variable and different sentiment measures as explanatory variable, controlling Fama French three factors and Carhart's momentum factor. Panel C presents results for the time-series regressions of long-leg combination portfolio returns and different sentiment measures as explanatory variable, controlling four factors. Panel D presents results for the time-series regressions of short-leg combination portfolio returns and different sentiment measures as explanatory variable, controlling four factors. AIres is residual component of anxious index. ICSres is residual component in Michigan University consumer sentiment. AAIIres is residual component of individual investor sentiment. GAres is residual component in rescaled Gallup survey data. IS-P is purged sentiment for which sample period corresponds to the sample period of different survey based sentiment measures respectively. The regression coefficients and Newey-West t-statistics are reported.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A. purged sentiment												
AIres	0.11											
	[2.59]											
ICSres		-0.09										
		[-2.60]										
AAIIres			-0.07									
			[-1.66]									
GAres				-0.07								
				[-1.17]								
R2(%)	1.23	1.18	0.79	0.65								
Panel B. L-S combine												
AIres	-0.01								-0.03			
	[-0.25]								[-0.59]			
ICSres		0.02								0.04		
		[0.31]								[0.55]		
AAIIres			-0.14								-0.12	
			[-1.63]								[-1.38]	
GAres				0.00								-0.02
				[-0.01]								[-0.18]
IS-P(AI)					0.19				0.19			
					[3.71]				[3.80]			
IS-P(ICS)						0.27				0.28		
						[4.13]				[4.11]		
IS-P(AAII)							0.32				0.32	
							[3.16]				[3.06]	
IS-P(GA)								0.41				0.41
								[3.31]				[3.32]
Panel C. L combine												
AIres	0.00								0.00			
	[-0.05]								[0.00]			
ICSres		0.02								0.02		
		[0.32]								[0.29]		
AAIIres			0.10								0.10	
			[1.56]								[1.54]	
GAres				0.22								0.22
				[2.42]								[2.37]
IS-P(AI)					-0.03				-0.03			
					[-0.58]				[-0.58]			
IS-P(ICS)						-0.03				-0.02		
						[-0.41]				[-0.39]		
IS-P(AAII)							-0.01				0.00	
							[-0.11]				[-0.02]	
IS-P(GA)								0.06				0.05
								[0.56]				[0.48]
Panel D. S combine												
AIres	0.01								0.03			
	[0.14]								[0.40]			
ICSres		0.00								-0.02		
		[-0.03]								[-0.19]		
AAIIres			0.24								0.22	
			[1.92]								[1.75]	
GAres				0.23								0.24
				[1.12]								[1.44]
IS-P(AI)					-0.21				-0.22			
					[-2.90]				[-2.97]			
IS-P(ICS)						-0.30				-0.30		
						[-2.84]				[-2.82]		
IS-P(AAII)							-0.33				-0.32	
							[-2.06]				[-1.94]	
IS-P(GA)								-0.34				-0.35
								[-1.75]				[-1.82]