Does investor sentiment predict the near-term returns of the Chinese stock market?

Muhammad A. Cheema, Yimei Man, and Kenneth R. Szulczyk

Abstract

Recent evidence on the relationship between investor sentiment and subsequent monthly market returns in China shows that investor sentiment is a *reliable* momentum predictor since an increase (decrease) in investor sentiment leads to higher (lower) future returns. However, we suggest that momentum predictability of investor sentiment originates from the boom and bust period of 2006-2008 (the bubble period *hereafter*). The bubble period is characterized by several months of sustained optimism followed by several months of sustained pessimism, with the market consequently earning high (low) returns following high (low) sentiment months. Therefore, we find a strong positive association between investor sentiment and subsequent market returns during the bubble period. However, investor sentiment has a negligible impact on subsequent monthly market returns once we exclude the bubble period.

I.   Introduction

Investor sentiment is defined as investor optimism or pessimism about the future stock prices that cannot be explained by the existing facts (Baker and Wurgler, 2006). In the literature, there is a growing consensus that investor sentiment is a contrarian predictor of stock market returns in the long run (e.g., Baker and Stein, 2004, Baker, Wurgler and Yuan, 2012, Brown and Cliff, 2005). However, the evidence of investor sentiment on the predictability of short-run market returns is disputed. For example, Brown and Cliff (2004) find that investor sentiment has a negligible impact on subsequent weekly and monthly

market returns in the U.S. In contrast, Huang, Jiang, Tu and Zhou (2014) show that investor

sentiment is a contrarian predictor at the monthly frequency in the U.S. based on their aligned

sentiment index. Baker and Wurgler (2006) show that the contrarian predictability of investor

sentiment is more pronounced in hard to value stocks, e.g., small size and high volatile firms.

Subsequently, Baker and Wurgler (2007) suggest that investor sentiment will be a contrarian

predictor of subsequent market returns if all the stock prices are positively related to the

sentiment. However, the impact of sentiment on subsequent market returns is weakened if

hard (easy) to value stocks are positively (negatively) related to the sentiment.

Recently, Han and Li (2017) hypothesize that investor sentiment is likely to forecasts

positive market returns in the short run in China because of the unique features of the Chinese

market (i.e., highly speculative market dominated by local retail investors, short sale

constraints), which differs from the U.S. and other developed markets.[1] Consequently, they

show that investor sentiment is a *reliable* momentum predictor of subsequent monthly market

returns and propose a profitable trading strategy based on investor sentiment in China.

However, their study is limited by their relatively small sample size (16.5 years) which

includes the Chinese boom and bust period of 2006-2008 (the bubble period *hereafter*).[2]

During the "bubble period", Shanghai Stock Exchange Composite Index rose from 1200 at

the beginning of the 2006 to a peak of 6092 in October 2007, and then plummeted to just

---

[1] Arguably, the unique features of Chinese market make it more prone to the prolonged boom and bust periods since short sale constraints and investors' speculative trading are important determinants of a stock market bubble (e.g., Cochrane, 2002, Lamont and Thaler, 2003, Ofek and Richardson, 2003)

[2] It is well documented in the literature that investor sentiment is strong in bubble periods, i.e., high (low) during boom (bust) periods (e.g., Baker and Wurgler, 2006, Huang et al., 2014). Some event, expectation, or development leads to a rise in stock prices that grow into a bubble. However, overconfidence, herding behaviour, etc. hastens a bubble's growth (e.g., Akerlof and Shiller, 2010). During the bubble or boom periods, investors become overconfident because they attribute the increase in stock prices to their skills (Daniel, Hirshleifer and Subrahmanyam, 1998). Consequently, they continue buying stocks, and even latecomers to the market cannot resist anymore as they witness the rising wealth of others. Eventually, a bubble burst since stock price is supported only by expectations of further price increases and it cannot continue forever (e.g., Akerlof and Shiller, 2010).

below 2000 in October 2008 (e.g., Yao and Luo, 2009, Liang and Willett, 2015, Chang, Chen, Waggoner and Zha, 2016).

Motivated by the recent literature that cast doubts on the credibility of empirical findings in finance (e.g., Harvey, Liu and Zhu, 2016, Hou, Xue and Zhang, 2017, Chordia, Goyal and Saretto, 2017), we suggest that the "bubble period" is likely to drive the positive association between investor sentiment and subsequent market returns in Han and Li (2017). The "bubble period" is characterized by several months of sustained optimism followed by several months of sustained pessimism, with the market consequently earning high (low) returns following high (low) sentiment months. Therefore, we hypothesize that investor sentiment is not a *reliable* momentum predictor in the Chinese stock market once we exclude the 'bubble period' from the sample. Consistent with our hypothesis, we find that investor sentiment *fails* to predict subsequent market returns with sentiment registering an insignificant OLS and bias-corrected slope coefficients.

Our paper adds two important contributions to the literature. First, it provides further support to the growing skepticism (e.g., Harvey et al., 2016, Hou et al., 2017, Chordia et al., 2017) about the validity of empirical findings in finance from the largest emerging market. Second, it suggests caution in making firm inferences based on small samples especially when the sample includes a bubble period. Therefore, robustness checks of regressions are necessary to determine whether a financial crisis or a bubble has an undue influence on a regression.

II.   Data and Methods

We collect the value-weighted monthly market returns (market returns *hereafter)* of A-shares listed on Shanghai and Shenzhen stocks exchanges and risk-free monthly rate from China Stock Market and the Accounting Research (CSMAR) database. Han and Li (2017) calculate the market returns based on the return index (RI) data of the individual stocks

collected from DataStream. We suggest that CSMAR forms a more reliable market return than DataStream since individual researchers estimate the later. Individual researchers impose screening processes, and the constant updating nature of the data from DataStream can cause difficulties to replicate and verify empirical findings. Nonetheless, our results remain robust when we estimate the market returns based on RI data from DataStream. Furthermore, our results remain the same when we use returns from the Shanghai Stock Exchange Composite Index, Shenzhen Stock Exchange Composite Index, and SSE 50 Index.

Following, Han and Li (2017), we develop a sentiment index using three sentiment proxies: price-earnings ratio (PE), turnover ratio (TO), and the number of newly opened individual investor accounts (IA) of Shanghai Stock Exchange.[3] We regress each proxy on a set of macroeconomic variables, namely, the growth of industrial production (IP), growth of money supply (MS2), 30-day inter-bank offer rate and foreign exchange rate (RMB/USD). Subsequently, we build the sentiment index (STM) using principal component analysis on the residuals from PE, TO, and IA. We collect all relevant sentiment and macroeconomic data from the CEIC database with the sample period starting in December 1996 and ending in September 2017.[4]

The resulting sentiment index is:

$$STM_t = 0.543PE_t^o + 0.590TO_t^o + 0.598IA_t^o \qquad (1)$$

The first principal component explains 76.41% of the sample variance. The subscript O denotes that all the three components in equation (1) are orthogonalized to the macroeconomic variables.

---

[3] Following Baker and Wurgler (2007), Chen, Chong and She (2014) and Han and Li (2017), we scale each individual sentiment proxy by its prior six-month moving average.

[4] We collect data from December 1996 since we scale each individual sentiment proxy by its prior six-month moving average.

To examine the predictability of sentiment on subsequent market returns, we estimate the following regression:

$$EMR_{t+1} = \alpha + \beta STM_t + \mu_{t+1} \tag{2}$$

As a robustness measure, we specify an alternative regression in (3) to examine the predictability of sentiment on subsequent market returns after controlling for the 2006-2008 bubble period.

$$EMR_{t+1} = \alpha + \beta \cdot STM_t + \gamma \cdot D_t + \delta \cdot D_t \cdot STM_t + \mu_{t+1} \tag{3}$$

where EMR_{t+1} is the market returns in excess of the risk-free rate in $t+1$ month; STM is the lagged sentiment index in month t.[5] D is a dummy variable that is equal to one for the bubble period (2006-2008), and zero otherwise. To test the predictability of investor sentiment, we use both the conventional OLS estimation and multivariate augmented regression method (mARM *hereafter*) of Amihud, Hurvich and Wang (2009). The mARM reduces the bias of the parameter estimates and standard errors of sentiment for the well-known Stambaugh (1999) small-sample bias.

III. Empirical Findings

Panels A and B of Table 1 reports the descriptive statistics and correlation matrix, respectively. We find that the market excess return from CSMAR at 0.67% per month exceeds the market excess return from DataStream at 0.58% per month. We suggest that the relatively lower market excess return from DataStream originates from DataStream carrying forward zero monthly returns of non-trading stocks from the previous period.[6] Panel B

---

[5] Results of predictive regressions are based on CSMAR data. However, we find similar results based on DataStream data.

[6] DataStream carries forward its RI even after the firm ceases trading. To identify and eliminate these dummy records, following Ince and Porter (2006) we delete all monthly RI observations from the end of the sample to the first non-zero return. However, we include zero returns of nontrading months where the firm has not ceased its trading.

reports the correlation between sentiment index and market excess returns. Consistent with the literature (e.g., Baker and Wurgler, 2007, Han and Li, 2017), we find that sentiment index is positively related to the contemporaneous market excess return both from CSMAR and DataStream. As expected, we find that the market excess return from CSMAR and DataStream are highly correlated (0.998) which adds assurance to our estimate of market return from DataStream.

Panels C to H of Table 1 reports the parameter estimates of the sentiment index where standard errors and t-statistics are adjusted using Newey and West (1987). Furthermore, we report the empirical p-values from the wild bootstrap procedure as in Huang et al. (2014), and also the p-values based on the bias-correction procedure of Amihud et al. (2009).

Panels C to H report the results based on equation (2). Panel C reports the results for the full sample (1997-2017). We find that the lagged sentiment positively relates to the subsequent market returns with a bias-corrected slope coefficient of 1.79% (t-stat=4.14) that corroborates Han and Li (2017). However, we suggest that the bubble period in the sample had induced these results. Therefore, we exclude the 2006-2008 period from our sample and report the results in Panel D. We find that investor sentiment is no longer a valid momentum predictor in the Chinese stock market after excluding the bubble period even at the 10% significance level.[7]

For additional robustness checks, we divide our sample into three sub-samples: the pre-bubble period, 1997-2005; the bubble period, 2006-2008; and the post-bubble period, 2009-2017, which Panels E, F, and G report respectively. Panels E and G show that the lagged sentiment is not associated with subsequent market returns before and after the bubble

---

[7] Our results remain similar when we use sentiment index data of Han and Li (2017). For example, we find that the OLS slope coefficient is 0.51% (t-stat = 0.79) and a bias-corrected slope coefficient is 0.64% (t-stat = 1.12) once we exclude the 2006-2008 period from Han and Li (2017) sample. We are thankful to Han and Li (2017) for providing their sentiment index data for robustness purposes.

period. However, consistent with our expectations, Panel F shows a strong positive association between investor sentiment and subsequent market returns during the bubble period with a bias-corrected slope coefficient of 5.12% (t-stat=8.64).[8]

Finally, Equation (3) tests whether our results remain robust when a dummy variable controls the effect of the bubble period.[9] We report the results in Panel H. The results in Panel H show that there is a strong positive association between investor sentiment and subsequent market returns when we interact bubble dummy with sentiment since the OLS coefficient (δ) is 3.87% (t-stat=3.47), and reduced-biased coefficient is 3.95% (t-stat=5.78) that is consistent with our results for the bubble period in Panel F.[10] Furthermore, the coefficient on sentiment (β) is insignificant for both OLS and reduced-biased estimators, consistent with our hypothesis.

In sum, our results show that the positive association between investor sentiment and subsequent market returns is limited to the bubble period which results from high (low) stock market returns following several months of continued high (low) investor sentiment since investor sentiment is high in boom and low in bust periods. Consequently, investor sentiment fails to reliably predict market returns in China once the sample excludes the bubble period.

---

[8] We are thankful to the anonymous reviewer for pointing out about the 2008-2009 financial crisis period. We also examine the relationship between investor sentiment and subsequent monthly market returns during the 2008-2009 financial crisis period. We find a positive association between investor sentiment and subsequent monthly market returns during the 2008-2009 financial crisis period. For example, the OLS coefficient is 4.76% (t-stat=1.80) and reduced-biased coefficient is 6.74% (t-stat=4.57). Most importantly, we still find that investor sentiment is not a *reliable* momentum predictor once we exclude the 2006-2009 bubble and financial crisis periods.

[9] We thank the anonymous reviewer for suggesting this test.

[10] For the sake of brevity, we do not tabulate Newey-West t-statistics, Newey-West p-values, and intercept dummy variable, γ. The Newey-West p-values are almost same as empirical p-values. The intercept dummy variable, γ, is statistically insignificant.

IV.   Concluding Remarks

Recent evidence on the relationship between investor sentiment and subsequent monthly market returns shows that investor sentiment is a reliable momentum predictor in China. We hypothesize that the 2006-2008 boom and bust period (the bubble period *hereafter*) had created the positive association between investor sentiment and subsequent monthly market returns, which is characterized by several months of sustained optimism followed by several months of sustained pessimism with the market consequently earning high (low) returns following high (low) sentiment months. Our results fully support our hypothesis as the results show that investor sentiment has a negligible impact on subsequent monthly market returns once we exclude the bubble period.

Table 1: Predictive Regressions

Panel A (B) reports the descriptive statistics (correlation matrix) of the sentiment index (STM), excess value-weighted market returns from CSMAR (CSM-Ret) and excess value-weighted market returns from DataStream (DSM-Ret). Panels C-G report the estimation results for the following predictive regression,

$$EMR_{t+1} = \alpha + \beta \cdot STM_t + \mu_{t+1}$$

Panel H reports the estimation results for the following predictive regression,

$$EMR_{t+1} = \alpha + \beta \cdot STM_t + \gamma \cdot D_t + \delta \cdot D_t \cdot STM_t + \mu_{t+1}$$

Where EMR_{t+1} is the excess value-weighted market returns at month $t+1$. STM is the level of sentiment index in month t. D is dummy variable that is equal to one for the bubble period (2006-2008), and zero otherwise. We report the slopes from both OLS and mARM methods. Table includes the Newey and West (1987) adjusted t-statistics (NW t-stat) and p-values (NW p-value), empirical p-values (E p-value) using the wild bootstrap as in Huang et al. (2014) on the Newey-West t-statistic, and the p-values (AH p-value) based on the bias-correction procedure of Amihud et al. (2009).

Panel A: Descriptive Statistics					
	Obs.	Mean	Median	Minimum	Maximum
STM	243	0.00	-0.16	-1.81	5.73
CSM-Ret (%)	243	0.67	0.74	-25.36	33.65
DSM-Ret (%)	243	0.58	0.70	-25.81	33.72

Panel B: Correlation Matrix			
	STM	CSM-Ret	DSM-Ret
STM	1.000		
CSM-Ret	0.605	1.000	
DSM-Ret	0.597	0.998	1.000

Panels C-G: Predictive Regressions						
Method	β. STM	NW t-stat	NW p-value	E p-value	AH p-value	R^2
Panel C: Full-Sample (July 1997-September 2017)						
OLS	0.0170	2.40	0.017	0.015	-	4.05%
mARM	0.0179	4.14	0.000	0.000	0.001	45.60%
Panel D: Full-Sample excluding 2006-2008						
OLS	0.0045	0.81	0.420	0.423	-	0.29%
mARM	0.0053	1.41	0.159	0.154	0.359	43.15%
Panel E: 1997-2005 period						
OLS	0.0020	0.41	0.693	0.691	-	0.07%
mARM	0.0036	1.21	0.230	0.224	0.640	49.49%
Panel F: 2006-2008 period						
OLS	0.0429	4.30	0.000	0.002	-	22.66%
mARM	0.0512	8.64	0.000	0.000	0.001	54.93%
Panel G: 2009-2017						
OLS	0.0062	0.72	0.474	0.468	-	0.47%
mARM	0.0083	1.08	0.282	0.294	0.359	38.17%

Panels H: Predictive Regression with Bubbles Dummy Variable (Full-Sample)						
Method	β. STM	E p-value	AH p-value	δ. D STM	E p-value	AH p-value	R^2
OLS	0.0043	0.441	-	0.0387	0.002	-	9.01%
mARM	0.0054	0.128	0.395	0.0395	0.000	0.001	47.79%

References

Akerlof, G. A. and R. J. Shiller (2010), Animal spirits: How human psychology drives the economy, and why it matters for global capitalism: Princeton University Press.

Amihud, Y., C. Hurvich and Y. Wang (2009), 'Multiple-Predictor Regressions: Hypothesis Testing', Review of Financial Studies, 22, 413-34.

Baker, M. and J. C. Stein (2004), 'Market liquidity as a sentiment indicator', Journal of Financial Markets, 7, 271-99.

Baker, M. and J. Wurgler (2006), 'Investor sentiment and the cross‐section of stock returns', The Journal of Finance, 61, 1645-80.

Baker, M. and J. Wurgler (2007), Investor sentiment in the stock market. National Bureau of Economic Research Cambridge, Mass., USA.

Baker, M., J. Wurgler and Y. Yuan (2012), 'Global, local, and contagious investor sentiment', Journal of Financial Economics, 104, 272-87.

Brown, G. W. and M. T. Cliff (2004), 'Investor sentiment and the near-term stock market', Journal of Empirical Finance, 11, 1-27.

Brown, G. W. and M. T. Cliff (2005), 'Investor sentiment and asset valuation', The Journal of Business, 78, 405-40.

Chang, C., K. Chen, D. F. Waggoner and T. Zha (2016), 'Trends and cycles in China's macroeconomy', NBER Macroeconomics Annual, 30, 1-84.

Chen, H., T. T. L. Chong and Y. She (2014), 'A principal component approach to measuring investor sentiment in China', Quantitative Finance, 14, 573-79.

Chordia, T., A. Goyal and A. Saretto (2017), 'p-hacking: Evidence from two million trading strategies'.

Cochrane, J. H. (2002), Stocks as money: convenience yield and the tech-stock bubble. National bureau of economic research.

Daniel, K. D., D. Hirshleifer and A. Subrahmanyam (1998), 'Investor psychology and security market under- and overreactions', The Journal of Finance, 53, 1839-85.

Han, X. and Y. Li (2017), 'Can investor sentiment be a momentum time-series predictor? Evidence from China', Journal of Empirical Finance, 42, 212-39.

Harvey, C. R., Y. Liu and H. Zhu (2016), '… and the cross-section of expected returns', The Review of Financial Studies, 29, 5-68.

Hou, K., C. Xue and L. Zhang (2017), Replicating Anomalies. National Bureau of Economic Research.

Huang, D., F. Jiang, J. Tu and G. Zhou (2014), 'Investor sentiment aligned: A powerful predictor of stock returns', The Review of Financial Studies, 28, 791-837.

Ince, O. S. and R. B. Porter (2006), 'Individual equity return data from Thomson Datastream: Handle with care!', Journal of Financial Research, 29, 463-79.

Lamont, O. A. and R. H. Thaler (2003), 'Can the market add and subtract? Mispricing in tech stock carve-outs', Journal of political Economy, 111, 227-68.

Liang, P. and T. D. Willett (2015), 'Chinese stocks during 2000–2013: Bubbles and busts or fundamentals?', The Chinese Economy, 48, 199-214.

Newey, W. K. and K. D. West (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', Econometrica: Journal of the Econometric Society, 703-08.

Ofek, E. and M. Richardson (2003), 'Dotcom mania: The rise and fall of internet stock prices', The Journal of Finance, 58, 1113-37.

Stambaugh, R. F. (1999), 'Predictive regressions', Journal of Financial Economics, 54, 375-421.

Yao, S. and D. Luo (2009), 'The economic psychology of stock market bubbles in China', *The World Economy,* 32**,** 667-91.