

The Cross-Section of Expected Stock Returns when Prices deviate from Fundamental Values

Preliminary draft: March 25, 2008

Marianna Caccavaio*

We use evidence from Chinese stock market to study the role of the investor sentiment in the cross-section of expected returns. In particular variables like volume and volatility are used to build a sentiment indicator so that stocks with high past exposure to this factor earn high future average returns. We present the results of trading strategies which exploit the information in the sentiment indicator to suggest the existence of average returns differentials.

JEL classification: G1

Keywords: Investor sentiment, Chinese Stock Market, Bubble, Cross-section of returns

* PhD student, Bocconi University.

Author's address: Via Gobbi 5, 20136 Milan, Italy. Tel: +390258365335, mail: caccavaio@unibocconi.it

1. Introduction

Determination of expected stock returns has been studied by many papers. The literature has mainly considered the US case even though more recently there have been analyses from international markets. Some research has also been conducted with data from emerging markets. There is a consensus about the relevance of multiple factors in the determination of expected stock returns, both for the US and for other markets.

A long debate regards the interpretation of the empirical factors. Some researchers try to provide a structural interpretation in terms of compensation for risk. Others contend that the results cannot only be interpreted from the point of view of risk because market sentiment is also important. Daniel, Hirshleifer and Subrahmanyam (2001) study a market populated by overconfident traders and arbitrageurs and find that expected returns should be cross sectionally explained by the Fama-French factors, some of which, particularly book-to-market, can however be interpreted as proxies for sentiment and security mispricing.

Sentiment has received attention from several authors. Baker and Wurgler (2006) predict that investor sentiment has larger effects on securities whose evaluations are highly subjective and difficult to arbitrage and find that when beginning-of-period sentiment is low (high), subsequent returns are relatively high (low) for small stocks, young stocks, high volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme growth stocks and distressed stocks. Baker and Wurgler build a sentiment indicator combining several variables and then show that the level of this indicator is associated with the subsequent return of stocks grouped in various portfolios. The sentiment story is reinforced by two observations: (a) weak evidence in favour of the effect of the sentiment indicator on time variability of beta and (b) logical inconsistency in explaining switching risk premia on the basis of a one factor model. Baker and Wurgler (2007) show that the sentiment indicator is correlated in the expected way with the returns of portfolios sorted on volatility and also with the market. Moreover in both cases the indicator may be used for forecasting.

Kumar and Lee (2006) use a database of 1.85 million retail investor transactions over 1991-1996 and find that systematic retail trading explains return comovements for stocks with high retail concentration. They build a portfolio-specific indicator on the basis of the buy-sell imbalance of retail investors and then use it to estimate multi-factor time-series models for several portfolios, finding that the indicator has incremental ability to explain returns especially for small firms, lower priced firms, firms with lower institutional ownership and value firms.

In this project we extend our own previous research on the Chinese stock market. The Chinese stock market has been studied by few authors. It is very interesting because of evidence suggesting that irrationalities and/or bubbles may be at work at the aggregate level. Moreover it is characterized by an important presence of retail investors and by a modest importance of macroeconomic variables in determining fundamentals.

As a preliminary analysis we measure the bubble-factor from traded portfolios, exploiting the segmentation existing in the Chinese stock market. In particular in section 2 we consider variables like volume and volatility to build a sentiment indicator. We start with a descriptive analysis aimed at understanding whether different sensitivities to that indicator are reflected in different average stock returns. In doing that we also do some robustness analysis. In section 3 we present the results of trading strategies which exploit the information in these factors to suggest the existence of average returns differentials. Section 4 concludes. Further research is needed to investigate these encouraging findings.

2. Empirical approach and data

A. Investor sentiment

Previous works suggest a number of proxies for sentiment to use as time series conditioning variables. In our own research we have studied volume and volatility as proxies of speculative behavior on the part of Chinese investors. Some of our results are hard to reconcile with market efficiency. Previous empirical analyses of the Chinese market have also suggested the existence of an important speculative component in stock price determination. The theoretical model of Hong, Scheinkman, and Xiong (2007) assumes limited risk absorption on the part of the market, agents with heterogeneous beliefs, overconfidence, insiders and short-sale constraints and show that stock prices should exceed fundamental values due to the presence of an optimism effect (only overoptimistic investors hold stocks while others cannot short them) and of a resale option effect (the possibility to sell stocks to future overoptimistic investors). The model predicts that under certain conditions variables like volume and volatility, should be positively correlated with stock prices. We form a composite index that captures the common component in these two proxies.

Previous researches suggest to include in our index variables like the premium for A shares over B shares and over H shares, price to book, and dividend yields as generic valuation indicators that may be associated with mispricing or rational valuation. However we can not use them in our analysis due to either some structural change in our sample period (from January 1998 to February 2007) or bad quality data. The following figures show the mentioned series.

Data are from Datasatram (dividend yields, market value, price to book, higher and lower price, opening and closing price, price index, return index, turnover by volume) and from Shenzhen GTA Information Technology Co Limited (number of tradeable shares of each company). We have daily data for all the A, B and H shares from the beginning to February 2007.

Figure 1: Average daily Turnover. The figure reports the logarithm of the weighted average daily turnover (million of shares traded on a given day) of the Shanghai and Shenzhen stock markets between January 1998 and February 2007.

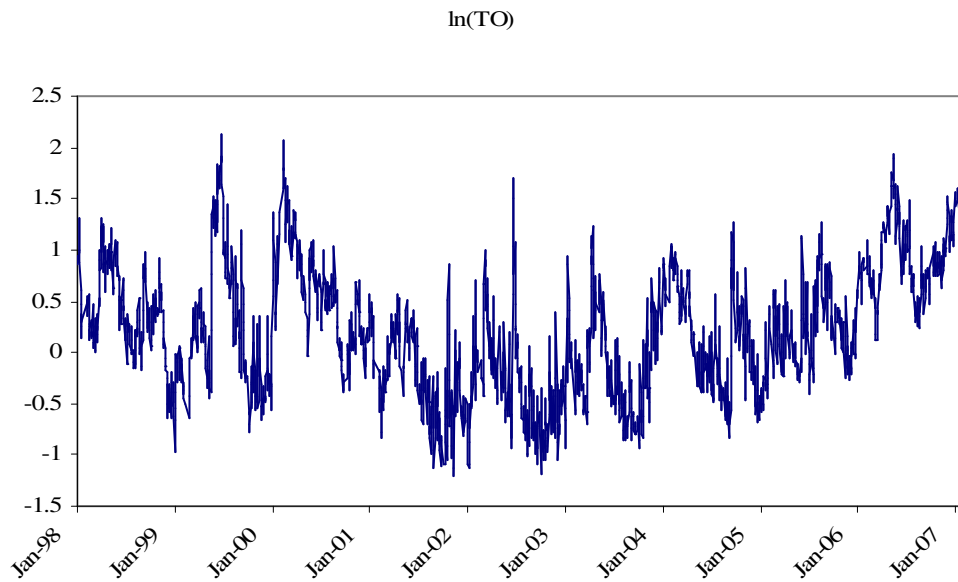


Figure 2: Average daily Price Range. The figure reports the weighted average daily Price range ((Price High - Price Low)/Price Low) of the Shanghai and Shenzhen stock markets between January 1998 and February 2007.

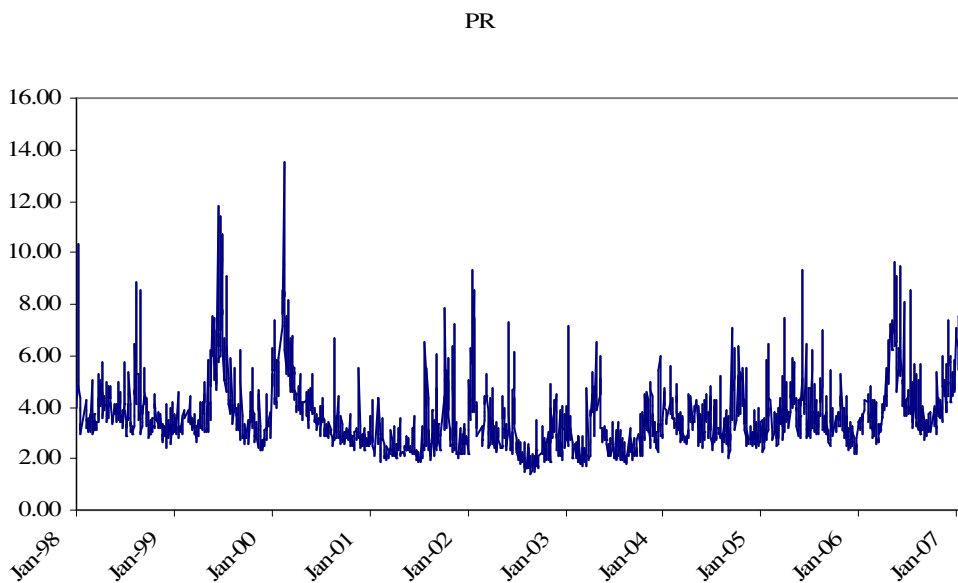


Figure 3: Average daily Price to Book Value. The figure reports the weighted average daily Price to Book value of the Shanghai and Shenzhen stock markets between January 1998 and February 2007.



Figure 4: Average daily Dividend Yield. The figure reports the inverse of weighted average daily dividend yield of the Shanghai and Shenzhen stock markets between January 1998 and February 2007.



Figure 5: Average daily AB premium. The figure shows the AB premium plotted at a daily frequency. The sample period is January 1998 and February 2007.

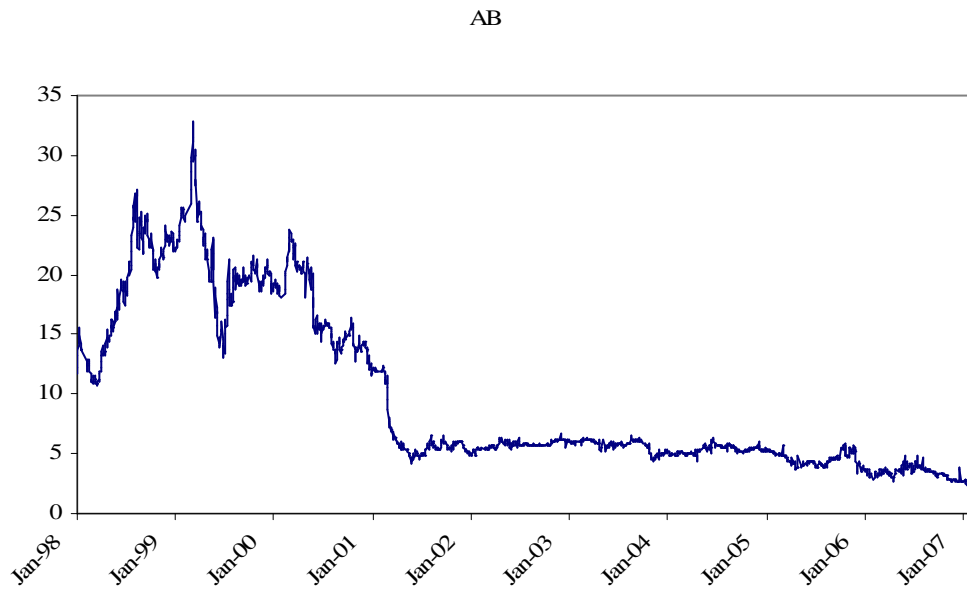
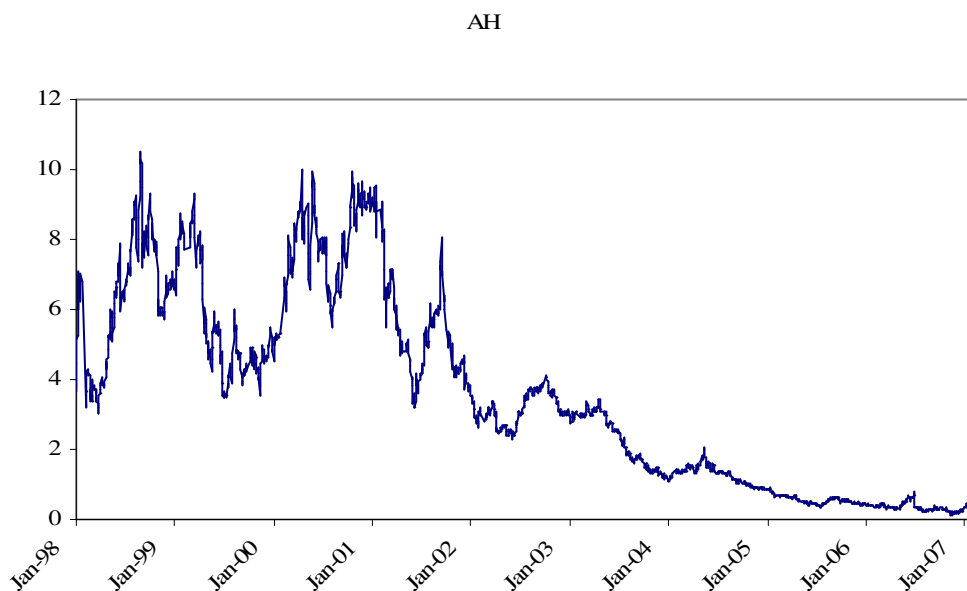


Figure 6: Average daily AH spread. The figure shows the AH spread plotted at a daily frequency. The sample period is January 1998 and February 2007.



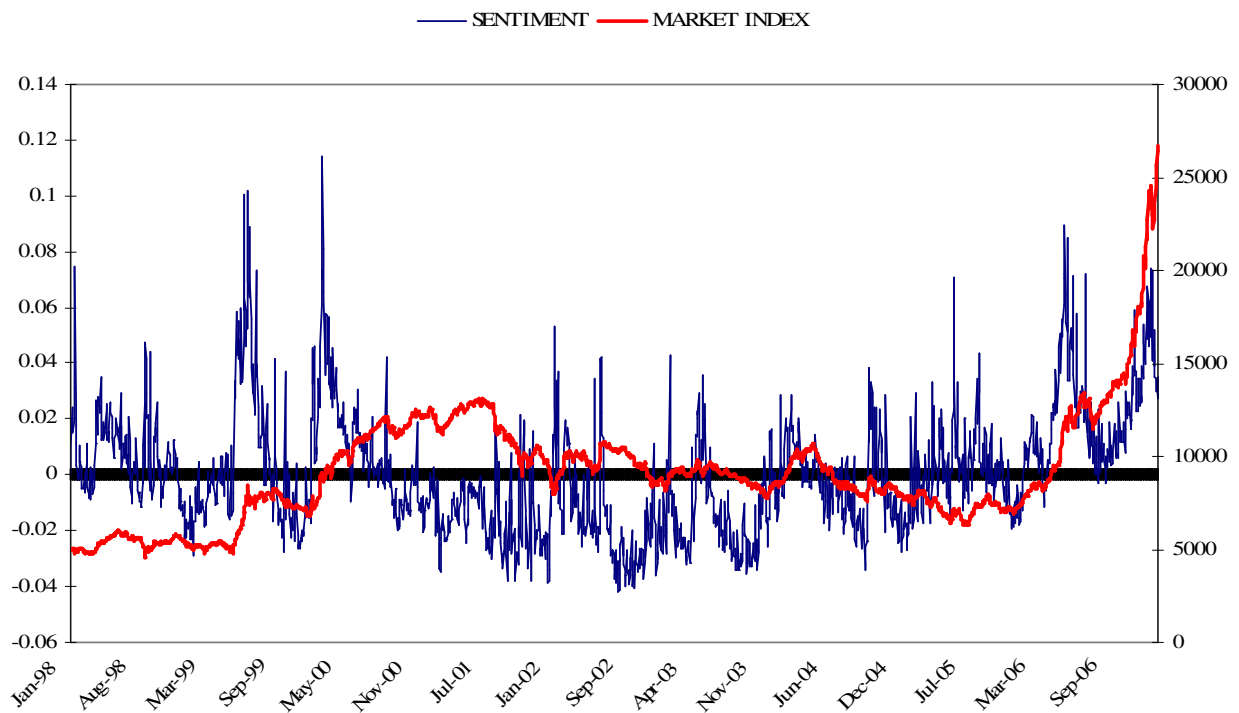
Following Baker and Wurgler (2006) we start by estimating the first principal component of the two proxies and their lags. This gives us a first-stage index with four loadings, one for each of the current and lagged proxies. We then compute the correlation between the first-stage index and the current and lagged values of each of the proxies. Finally, we define SENT as the first principal component of the correlation matrix of two variables – each respective proxy’s lead or lag, whichever has higher correlation with the first-stage index – rescaling the coefficient so that the index has unit variance.

$$SENT_t = 0.01187 \ln(TO)_t + 0.01187 PR_{t-1} \quad (1)$$

Each individual proxy enter in the index with the expected positive sign.

We compare the sentiment indicator with a market index which includes both shares traded in the Shanghai Stock Exchange and shares traded in the Shenzhen Stock Exchange. We need to compute a market index by considering the actual float of each company. This is important in view of the large difference between float and capitalization caused by the existence of nontradeable shares. A capitalization index would include the quantity of both tradeable shares and nontradeable shares to compute the weights assigned to the various stocks and would provide a measure not reflecting current market conditions.

Figure 7: Sentiment and Market Index. The figure reports the daily index of the Shanghai and Shenzhen stock markets and the sentiment index between January 1998 and February 2007.



The sentiment indicator is positive in the first part of the sample period, negative between 2001 and 2003, almost neutral between 2003 and 2006. After that it remains positive until the end of the sample period, February 2007. A similar pattern is apparent consistent with the dynamics in the Chinese Stock Market.

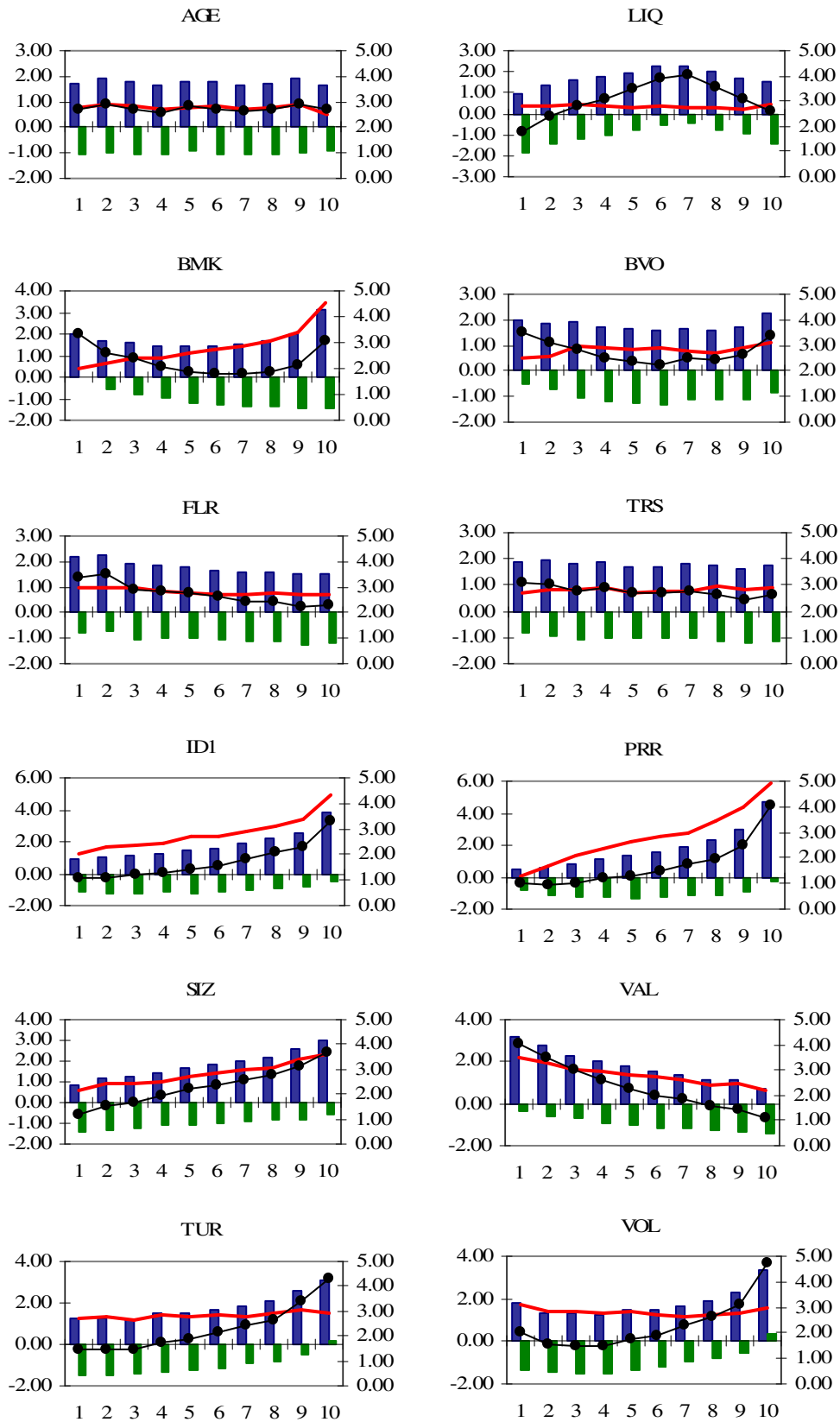
B. Sorts

We look for conditional characteristics effect in a simple, nonparametric way. The characteristics analysed are:

- AGE: age of companies in trading days
- BMK: market beta
- BVO: beta to the volume factor
- FLR: floating ratio
- ID1: idiosyncratic volatility computed over one month
- LIQ: liquidity computed as in Pastor and Stambaugh,
- PR: price range
- SIZ: log market capitalization for firms within the portfolio
- VAL: book to market ratio
- TUR: turnover by volume
- VOL: log volume over log market value

We place each monthly return observation into a bin according to the decile rank that a characteristic takes at the beginning of that month, and then according to the average level of SENT of the same month. We compute the equal-weighted average monthly return for each bin and look for patterns.

Figure 8: Two way sort. Figure reports monthly returns by sentiment index and firm characteristics 1998-2007. The blue bars are returns in positive sentiment periods, and the green bars are returns in negative sentiment periods. The black line is the average. The red line is the difference (right scale).



Then, we identify time series change in cross-sectional effects from the conditional difference of average returns across deciles characteristics. We experiment with several risk pricing models for the Chinese stock market. We consider a simple market model, and an extended Wang-Xu model including the market, a size portfolio, a floating ratio portfolio, a liquidity portfolio. Wang and Xu (2004) propose including a floating ratio portfolio as a proxy for risk of bad governance and expropriation of holders of tradeable shares. They also suggest that book-to-market is unlikely to play an important pricing role because of poor accounting quality in the Chinese stock market. The factor replicating portfolios have been built following the methodology described by Fama and French (1996). We use the Shenzhen GTA Information Technology Co Limited data in order to build a float-weighted market index and float-weighted risk factors.

$$R_{X_{it=High,t}} - R_{X_{it=Low,t}} = \beta_0 + \beta_{SENT} SENT_t + \beta_{MKT} MKT_t + \varepsilon_t \quad (2)$$

$$R_{X_{it=High,t}} - R_{X_{it=Low,t}} = \beta_0 + \beta_{SENT} SENT_t + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{FLR} FLR_t + \beta_{LIQ} LIQ_t + \varepsilon_t \quad (3)$$

Table 1: Time series regression of characteristics portfolios return. Table reports regression of long-short portfolio returns.

		CAPM		WXL			CAPM		WXL	
		Coeff	P-value	Coeff	P-value		Coeff	P-value	Coeff	P-value
Constant	AGE	0.00	0.01	0.00	0.61	LIQ	0.00	0.84		
Beta_SENT		0.01	1.46	0.00	0.38		-0.01	1.69		
Beta_MKT		0.02	3.99	-0.02	3.33		0.00	0.74		
Beta_SIZ				-0.13	5.36					
Beta_FLR				0.62	23.00					
Beta_LIQ				-0.16	5.84					
R-squared		0.01		0.37			0.00			
Constant	BMK	0.00	2.59	0.00	2.48	BVO	0.00	0.76	0.00	0.90
Beta_SENT		-0.03	5.04	-0.03	5.84		0.00	0.40	0.00	0.29
Beta_MKT		0.30	56.98	0.28	38.05		0.06	18.76	0.07	13.56
Beta_SIZ				-0.29	8.64				-0.06	2.60
Beta_FLR				0.22	6.03				-0.11	4.55
Beta_LIQ				0.15	4.09				0.03	1.19
R-squared		0.45		0.50			0.08		0.09	
Constant	FLR	0.00	1.20			TRS	0.00	1.24	0.00	0.07
Beta_SENT		0.01	1.52				0.01	1.51	0.00	0.64
Beta_MKT		0.05	13.63				0.04	11.21	0.00	2.55
Beta_SIZ									0.11	13.71
Beta_FLR									0.95	105.32
Beta_LIQ									0.00	0.16
R-squared		0.05					0.03		0.88	
Constant	ID1	0.00	3.02	0.00	2.82	PRR	0.00	3.01	0.00	2.75
Beta_SENT		-0.01	2.51	-0.01	2.89		-0.02	2.84	-0.02	3.49
Beta_MKT		0.09	16.96	0.07	9.86		0.20	33.07	0.17	20.57
Beta_SIZ				0.05	1.49				0.02	0.42
Beta_FLR				0.32	8.61				0.53	12.89
Beta_LIQ				-0.04	0.93				0.01	0.30
R-squared		0.07		0.11			0.21		0.30	
Constant	SIZ	0.00	0.09			VAL	0.00	0.48	0.00	0.70
Beta_SENT		-0.01	1.81				-0.01	1.40	-0.01	2.06
Beta_MKT		-0.04	8.64				0.14	27.73	0.13	18.65
Beta_SIZ									-0.44	13.60
Beta_FLR									-0.13	3.84
Beta_LIQ									-0.06	1.62
R-squared		0.02					0.16		0.25	
Constant	TUR	0.00	1.52	0.00	1.28	VOL	0.00	0.85	0.00	0.73
Beta_SENT		-0.01	2.67	-0.01	2.49		-0.01	2.07	-0.01	1.75
Beta_MKT		0.12	24.08	0.12	16.60		0.11	24.26	0.11	17.07
Beta_SIZ				0.30	9.53				0.06	2.11
Beta_FLR				0.24	6.98				0.01	0.24
Beta_LIQ				0.05	1.26				0.20	6.08
R-squared		0.13		0.17			0.13		0.15	

C. The pre-formation regression

Like before, the empirical model that we examine are:

$$r_t^i = \beta_0 + \beta_{MKT}^i MKT_t + \beta_{SENT}^i SENT_t + \varepsilon_t^i \quad (4)$$

And the Wang and Xu with liquidity replicating portfolio

$$r_t^i = \beta_0 + \beta_{MKT}^i MKT_t + \beta_{SENT}^i SENT_t + \beta_{SMB}^i SMB_t + \beta_{FLR}^i FLR_t + \beta_{LIQ}^i LIQ_t + \varepsilon_t^i \quad (5)$$

We sort firms on *SENT* loadings over the past month using the regression (4) or (5) with daily data.

At the end of each month, we sort stocks into quintiles, based on the value of the realized β_{SENT} coefficients from lowest (quintile 1) to highest (quintile 5) over the past 2 years. Results for the market model are reported in the following table.

Table 2: Portfolio sorted by exposure to sentiment index

Rank	Quintile_1	Quintile_2	Quintile_3	Quintile_4	Quintile_5
GI	0.65%	0.68%	0.75%	0.89%	1.16%
Std Dev	7.314	7.370	7.493	7.957	8.705
% Mkt share	22.2%	20.0%	19.6%	18.7%	19.5%
alpha_CAPM	-0.017	-0.024	-0.026	-0.027	-0.020
alpha_WXL	-0.042	-0.049	-0.050	-0.052	-0.050
AGE	1298	1150	1018	1060	1411
BMK	1.052	1.040	1.044	1.033	1.005
BVO	1.423	1.385	1.439	1.413	1.361
FLR	-0.012	-0.007	-0.005	0.000	0.015
ID1	1.795	1.762	1.780	1.828	1.922
ID2	1.843	1.811	1.824	1.874	1.969
LIQ	-0.266	-0.288	-0.239	-0.261	-0.354
PR	3.678	3.624	3.654	3.717	3.822
SIZ	6.612	6.560	6.565	6.551	6.586
TRS	0.355	0.357	0.359	0.364	0.382
TUR	2609	2491	2544	2498	2351
VAL	0.300	0.297	0.274	0.263	0.252
VOL	0.996	0.997	1.001	1.002	0.995

The statistics in the rows labelled GI and Std Dev are measured in daily percentage terms and apply to total simple returns. The % Mkt share represents the percentage of market value within the quintile. The alpha_CAPM and alpha_WXL rows report Jensen's alpha with respect to the CAPM and the Wang and Xu with liquidity replicating portfolio model. Other rows report the weighted average characteristics for firms within the portfolio. The sample period is form January 1998 to

February 2007. Table 2 demonstrates that the high average returns to stocks with high past sensitivity to investor sentiment cannot be explained by firm characteristics.

D. A factor mimicking investor sentiment

We create the mimicking factor MIMIC by estimating the coefficient b in the following regression:

$$SENT_t = c + b' X_t + u_t \quad (6)$$

where X_t represents the returns on the base assets. The return on the portfolio, bX_t , is the factor MIMIC. We use the quintile portfolios sorted on past β_{SENT} as the base assets X_t .

E. The price of investor sentiment

Since this evidence supports the case that sensitivity to investor sentiment is a priced risk factor in the cross-section of stock returns, the next step is to estimate the factor premium λ_{SENT} on the investor sentiment.

$$r_t^i = \beta_0 + \beta_{MKT}^i MKT_t + \lambda_{SENT}^i SENT_t + \varepsilon_t^i \quad (7)$$

$$r_t^i = \beta_0 + \lambda_{MKT}^i MKT_t + \lambda_{SENT}^i SENT_t + \lambda_{SMB}^i SMB_t + \lambda_{FLR}^i FLR_t + \lambda_{LIQ}^i LIQ_t + \varepsilon_t^i \quad (8)$$

We use the 25 $\beta_{MKT} \times \beta_{SENT}$ base assets to estimate factor premiums in equations (7) and (8) following the two-step procedure of Fama-MacBeth (1973).

We conduct Fama and MacBeth test with this sentiment factor as one of the candidate factor variables. Under the null hypothesis of the existence of the risk premium on each possible factor, we find that the risk premium of for sentiment factor is significant.

Figure 9 and 10 show the results for the market model and the Wang and Xu model with liquidity replicating portfolio.

Figure 9: Plot of the factors premium for the market model

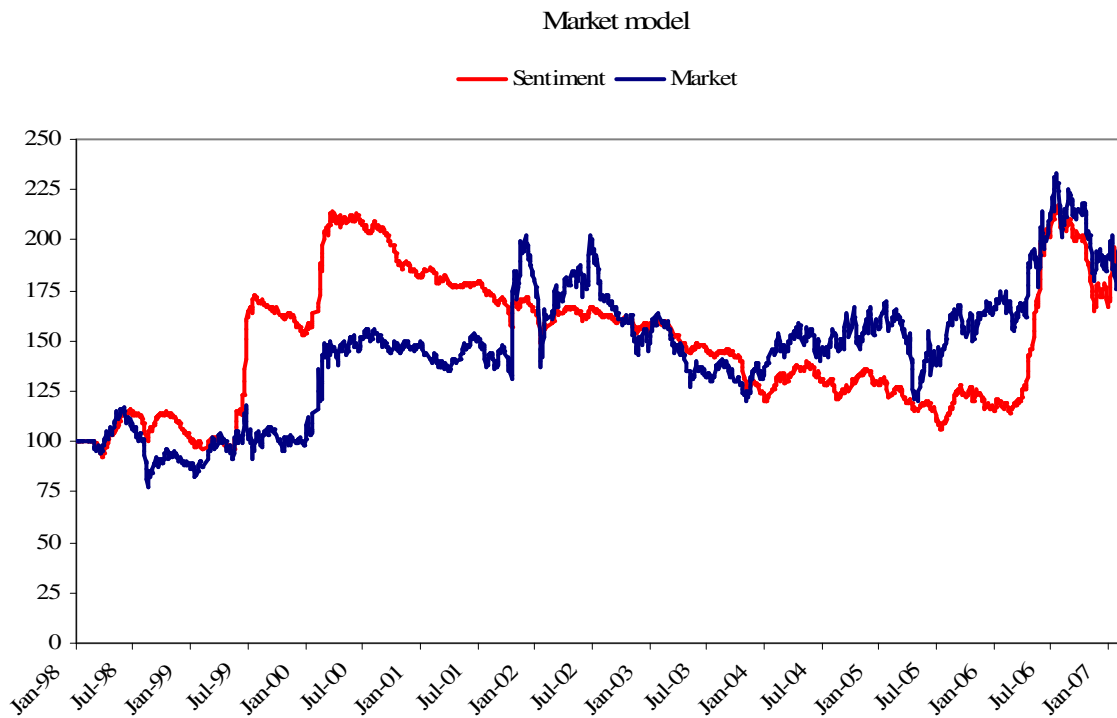
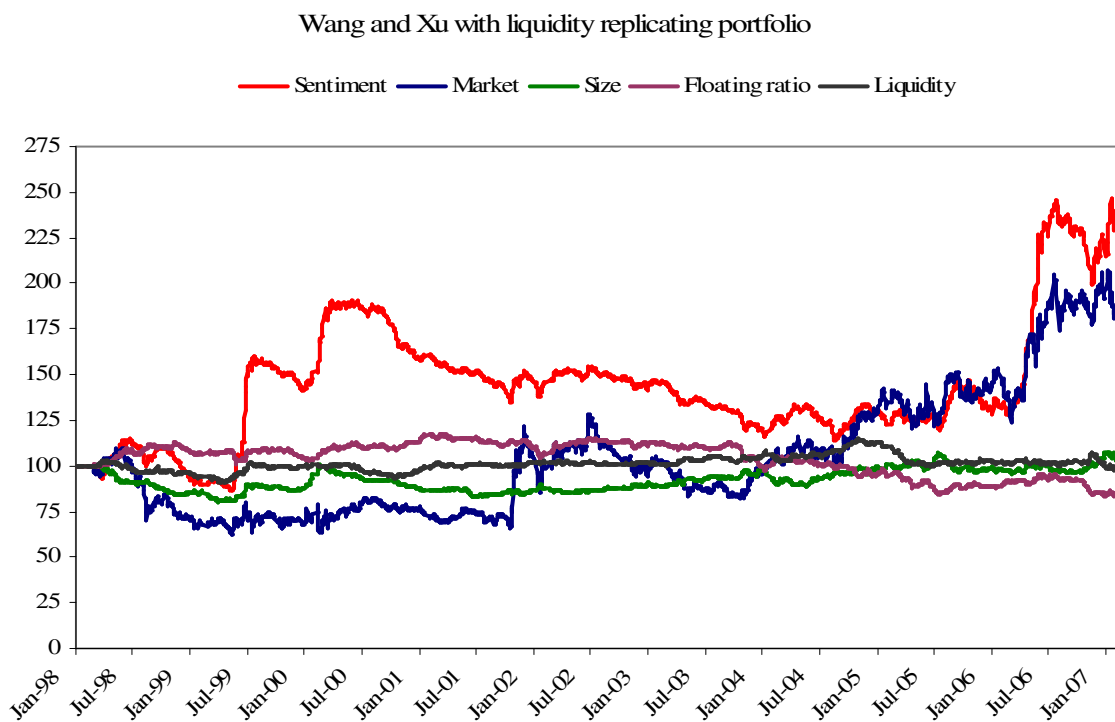


Figure 10: Plot of the factors premium of the Wang and Xu with liquidity replicating portfolio



3. Pricing investor sentiment in the Cross-Section

A. Trading strategies

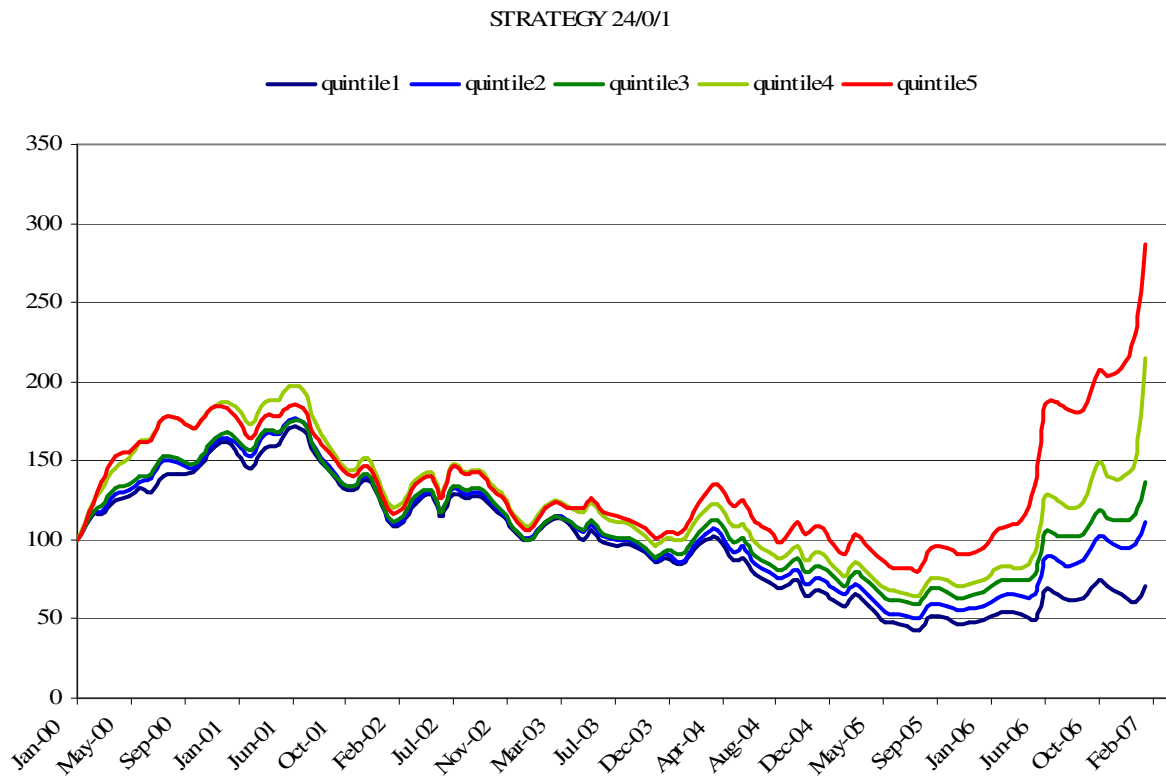
To examine trading strategies based on investor sentiment, we describe portfolio formation strategies based on an estimation period of L months, a waiting period of M months, and a holding period on N months. We describe an $L/M/N$ strategy as follows. At month t , we run regression (4) on daily data over an L month period from month $t-L-M$ to month $t-M$. At time I , we construct value-weighted portfolios based on the β_{SENT}^i and hold these portfolios for N months. We present result for analysis on 24/0/1 strategy, in which we simply sort stocks into quintile portfolios based on their level of beta computed using daily data returns over the past two years, and we hold these value-weighted portfolios for one month. The portfolios are rebalanced each month.

The average monthly return for each quintile are the following:

- Quintile1: - 0.23%
- Quintile2: + 0.57%
- Quintile3: + 1.01%
- Quintile4: + 1.93%
- Quintile5: + 2.46%

Figure 11 shows the performance of the five portfolios.

Figure 11: Portfolios strategy



The findings are provocative. Even if the statistics in tables 1 and 2 seem to support these results we need to examine whether the phenomena persist if we control for a large number of cross sectional effects that the literature has identified either as potential risk factors or anomalies.

4. Conclusions and further analysis

In classical finance theory, investor sentiment does not play any role in the cross-section of stock prices and expected returns. We use simply theoretical arguments to demonstrate that investor sentiment has significant cross sectional effects. We have studied volume and volatility as proxies of speculative behavior on the part of Chinese investors and we have found that stocks with high past exposure to investor sentiment earn high future average returns.

Such strong patterns in stock returns definitely call for further research.

First of all we plan to extend the analysis to include the period between February 2007 and March 2008 examining the cause, the characteristics, and behavior of the Chinese asset price bubble.

Moreover we want to repeat the exercise in a more efficient market. We plan to analyze the US stock market. It is interesting to see if NASDAQ stocks are the ones with high exposure to investor sentiment during the Internet bubble period. We already have data for both Chinese and US markets.

REFERENCES

- Ang, Andrew, Robert Hodrick, Yuhang Xing and Xiaoyan Zhang. *Journal of Finance*
- Baker, Malcolm and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645-1680
- Baker, Malcolm and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, NBER Working paper series no. 13189
- Beltratti, Andrea and Marianna Caccavaio, 2007, Asst float and stock prices: Evidence from the Chinese market, mimeo
- Daniel, Kent D., David Hirshleifer and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 66, 921-965
- Fama, Eugene F. and James D. MacBeth, 1973, Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy*, Vol. 81, Issue 3, pp. 607-636
- Fama, Eugene, and Kenneth French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84
- Hong, Harrison, José Scheinkman, and Wei Xiong, 2006, Asset float and speculative bubbles, *Journal of Finance* 61, 1073-1117.
- Kumar, Alok and Charles M.V.C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance*, 61, 2451-2486
- Pastor, Lubos, and Robert Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 113, 642-685.
- Wang, Fenghua and Yexiao Xu, 2004, What determines Chinese stock returns?, *Financial Analysts Journal* 60, 65-77.