行政院國家科學委員會專題研究計畫　成果報告

市場情緒和基本面對短期股價可預測性之影響

計畫類別：個別型計畫
計畫編號：NSC92-2416-H-004-043-
執行期間：92 年 08 月 01 日至 93 年 07 月 31 日
執行單位：國立政治大學國際貿易學系

計畫主持人：郭維裕

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 93 年 11 月 4 日

行政院國家科學委員會補助專題研究計畫成果報告

市場情緒和基本面對短期股價可預測性之影響

計畫類別：個別型計畫
計畫編號：NSC 92－2416－H－004－043
執行期間：民國九十二年八月一日至九十三年七月三十一日

計畫主持人：郭維裕

成果報告類型(依經費核定清單規定繳交)：精簡報告

執行單位：國立政治大學國際貿易學系

中　華　民　國　九十三　年 十 月 三十一 日

中文摘要

本研究主要在探討利用基本面變數和市場情緒變數來預測台灣股市中短期股票報酬率的可行性。本研究所採用的基本面變數主要是根據 Pesaran and Timmermann（1995，2000）的研究而來，至於市場情緒變數則是參考 Brown and Cliff（2001a，2001b）的研究成果。我們用來預測短期股票報酬率的計量方法是遞迴線性迴歸模型，Pesaran and Timmermann（1995，2000）也是採用這個模型來進行預測。相對地，Brown and Cliff（2001a，2001b）則是採用主成分法搭配傳統的線性迴歸模型來進行預測。雖然 Brown and Cliff（2001a，2001b）發現：利用市場情緒變數來預測短期報酬率的效率不如預期的好，我們認為他們所發現的低效率很可能源自於其所採用的計量方法。這是因為，他們的方法只能透過 R^2 來呈現出市場情緒變數在樣本期間中的平均預測能力，而無法顯示出這些變數在不同時點上對短期報酬率的預測能力。也就是說，市場情緒變數很可能在不同的市場情況下，會具備不同的短期報酬率預測能力；亦即，有時會預測的準，但有時不準，而平均的結果卻是令人失望的低 R^2。Pesaran and Timmermann（1995，2000）所建議的遞迴線性迴歸模型剛好可以針對這個可能性進行檢驗，這也是本研究的主要貢獻之一。本研究的另一個貢獻則是在遞迴線性迴歸模型同時加入基本面變數和市場情緒變數，以比較兩種變數在不同時點對短期股票報酬率的預測效能。最後，我們還採用兩種方法來衡量本研究的資料採礦（data mining or data snooping）風險。第一種是 Leamer（1983）所建議的極端界限分析（extreme bounds analysis），第二種則是利用模型的預測值來實際進行投資組合操作，並檢驗此操作能否產生顯著的超額報酬。

本研究發現許多有關市場情緒和市場基本面的變數具有對短期股價的預測能力。就市場情緒變數而言，券資比、零股賣量和買量比、大盤週轉率、認購權證發行總數以及三大法人淨買超等變數皆具有顯著的預測能力。而在市場基本面變數方面，對大盤短期變動具有顯著預測能力的變數則有一個月期的定期存款利率、工業生產指數以及市場現金股利殖利率。雖然這些變數往往可以預測短期大盤變動的方向，但是它們對大盤指數的影響卻會隨著時間而變。例如，券資比常被不同的模型選為顯著變數，可是它的係數卻是正負反覆不定。也就是說，在某些樣本期間內，當前一個月的市場券資比升高時，下個月的大盤指數會上漲的機率較高；可是在其他的樣本期間內，我們卻發現相反的結果。這表示投資人在選擇最佳的預測模型時，所面臨的不確定性很高，當然這也反映出預測股市短期變動的高難度。此外，我們也發現：一般而言，根據不同篩選標準所挑中的最佳預測模型都具有一個共同的特性，它們在整體股市的波動度變大時，具有較佳的預測能力，這個結果與 Pesaran and Timmermann（1995，2000）的結果相一致。最後，本研究的結果顯示：所有最佳模型的預測能力有隨著時間下降的趨勢。
關鍵詞：短期股票報酬率之可預測性、基本面變數、市場情緒變數、遞迴線性迴歸模型

Abstract

This study investigates the forecasting power of fundamentals variables and market sentiment proxies on short-term stock return in Taiwan stock market. Since the forecasting power of both fundamental variables and market sentiment variables may vary with time, we decide to adopt the recursive regression model as proposed by Pesaran and Timmermann (1995, 2000) to explore this possibility. Our results indicate that several fundamental variables and market sentiment variables have significant predictive ability on short-term stock returns in Taiwan stock market. These variables include the one-month time deposit interest rate, the industrial production index, the market dividend yield, the ratio of the short sales to margin purchase, the market turnover ratio, the number of warrant issues, and the total net purchase of institutional investors, including the investment trusts, the qualified foreign institutional investors, and the security firms. Although these variables can predict short-term stock returns significantly, their forecasting power are actually time-varying in terms of both the degree of significance and the forecasted direction. Take the ratio of the short sales to margin purchase for example, even though it has significant predictive ability most of the time, it does lose this ability during certain sample periods. Besides, during those periods when it can predict short-term stock returns effectively, its coefficient can also be positive or negative. This result is puzzling and reflects the great difficulty in constructing a forecasting model that can predict short-term stock returns correctly and consistently. We also find that the best forecasting models selected by various criteria have better performance when the stock market is volatile. Finally, we also show that the overall predictive power of these best models actually deteriorate with time. In other words, it is getting more and more difficult to successfully forecast the short-term variation of stock market.

Keywords: the predictability of short-term stock return 、 fundamental variables 、 market sentiment variables 、 recursive linear regression

1. Introduction

Whether stock return can be predicted or not is a question that very much intrigues both practitioners and academics for decades and yet still needs to be answered. Before 1970s, stock price is deemed to follow a random walk process that implies the unpredictability of stock return and the market efficiency of capital markets, e.g., Working (1934), Granger and Morgenstern（1963）, Samuelson（1965）, Mandelbrot（1966）and Fama（1965，1970）. In contrast, in 1980s there are more and more empirical evidence indicating that both in the long run and short run, stock price does contain a predictable component, e.g., Shiller（1984）, DeBondt and Thaler（1985）, Summers（1986）, Fama and French（1988）, Conrad and Kaul（1988，1989）, Jegadeesh（1990，1991）and Boudoukh, Richardson, and Whitelaw（1994）. Although the statistical significance of the empirical evidence is still subject to serious scrutiny, the argument that stock return is predictable seems to gain its popularity gradually.

Several macroeconomic variable and firm characteristics variables have been used to forecast stock return in the vast amount of empirical studies on the predictability of stock return. These variables include interest rates, monetary growth rates, changes in industrial production, inflation rates, international oil prices, term structure variables for interest rate, earnings-price rations, dividend yields, firm sizes, market-to-book ratio and lagged stock returns. Recently, in the literature of behavioral finance many variables germane to investor behavior and market sentiment have also been proposed for predicting the variation of stock returns. Brown and Cliff（2001）adopt several direct and indirect proxies of market sentiment to forecast both short-term and long-term stock returns in U.S. stock markets. These parameters include the percentage of bullish investors minus the percentage of bearish according to the survey conducted by the American Association of Individual Investors, the ARMS index also known as the Trading Index or the short-term trading index, the percent change in margin borrowing, the percent change in short interest, the ratio of short sales to total sales, the ratio of CBOE equity put to call trading volume, the change in the net position of non-commercial traders and of small traders and other proxies for market sentiment such as the closed-end fund discount, the net purchases of mutual funds, the proportion of fund assets held in cash, the number of initial public offerings, the initial public offering first day returns, the returns on large stocks and small stocks. They find that although these sentiment parameters have different degree of statistical significance in predicting stock returns, generally they perform better in forecasting long-term stock returns than short-term stock returns.

Our study focuses on employing the recursive regression model of Pesaran and

Timmermann (1995, 2000) to reexamine the marginal power of market sentiment variables, conditional on the macroeconomic variables and firm characteristics variables, in predicting short-term stock returns in Taiwan stock market. In particular, we argue that the low R^2 reported by Brown and Cliff measures only the overall average forecasting ability of sentiment parameters during the whole sample period rather than the forecasting ability of each parameter at each individual time point within the sample period. Hence, using R^2 to judge the predictability of short-term stock returns based on the sentiment variables would be misleading, especially when the predicting ability of the sentiment variables is actually time varying. To explore this possibility, we therefore adopt the recursive regression model to study the power of the market sentiment variables in forecasting short-term stock returns.

2. Methodology

Following Pesaran and Timmermann (1995, 2000), we assume that investors have no specific preference in choosing models. Investors would choose the best set of variables as long as they can predict stock returns. In other words, the forecasting models may evolve over time in order to reflect the learning process of the investor. Suppose that, at each point of time t, an investor utilizes a base set of k variables to construct one-period-ahead forecasts of excess returns according to data available at the time. In order to simulate the investor's real-time searching for a forecasting model, we choose the best model recursively at each point of time from all possible models estimated by including all possible k factors $\{x_1, x_2, ... x_k\}$ in the base set. This best model is chosen according to several statistical criteria for model selection discussed below. This procedure generates a total of 2^k different models at each point of time and the selected variable(s) can be specified by a $k \times 1$ selection vector, v_i, composed of 1, -1, and 0 where a 1 in its j_{th} element means that the j_{th} regressor included in the model has positive influence in the stock return while a -1 means that the j_{th} regressor included in the model has negative influence in the stock return and a 0 in its j_{th} element means that the j_{th} regressor is excluded from the model. At each point in time, τ, our forecasting model i (denoted by M_i) is estimated by means of linear regressions

$$M_i : \rho_{\tau+1} = \beta'_i X_{\tau,i} + \varepsilon_{\tau+1,i} \quad \tau = 1,2,\ldots,t\text{-}1$$

where $\rho_{\tau+1}$ is the excess stock return at time τ+1 and $X_{\tau,i}$ is a $(k_i + 1) \times 1$ vector of regressors under model M_i, obtained as a subset of the base set of regressors, X_τ, which is determined by the investor before the predicting process, plus a vector of ones for the intercept term. Conditional on model M_i and given the observations of $\rho_{\tau+1}$ and $X_{\tau,i}$, τ = 1, 2,…, t-1 (with $t \geq k + 2$), parameters of model M_i can be estimated by the ordinary least square (OLS) method. Denoting these estimates by $\hat{\beta}_{t,i}$ we have

$$\hat{\beta}_{t,i} = (\sum_{\tau=0}^{t-1} X_{\tau,i} X'_{\tau,i})^{-1} \sum_{\tau=0}^{t-1} X_{\tau,i} \rho_{\tau+1}\,, \quad \text{for} \quad t = k+2, k+3, \ldots, T, \text{ and } i = 1,2,\ldots 2^k .$$

The OLS estimates are simple to compute and according to the Gauss-Markov Theorem, are reasonably robust even in the presence of nonnormal errors in the excess stock return equation.

The particular choice of $X_{\tau,i}$ to be used in predicting $\rho_{\tau+1}$ is dependent on a number of statistical model selection criteria including the $\overline{R}^2$, Akaike's Information criterion [AIC, (Akaike, 1974)], Schwarz's Bayesian Information Criterion [BIC, (Schwarz, 1978)], the Fisher Information Criterion [FIC, (Wei, 1992)], the Posterior Information Criterion [PIC, (Philips and Ploberger, 1996)], and the sign criterion (SC). The $\overline{R}^2$ criterion is based on the optimality principle as minimizing the estimate of the error variance σ^2. It balances the fit and parsimony by taking into account the trade-off between the gain in explanatory power and loss in degrees of freedom. The AIC, BIC, FIC, and PIC select the best model based on the mean squared error multiplied by a certain penalty factor depending on the model complexity that is measured by the number of regression coefficients to be estimated. The SC determines a model as the best model that has the highest directional accuracy of the forecasts. The $\overline{R}^2$, AIC, and BIC are chosen on the basis of their popularity whereas the FIC and PIC are chosen because of their robustness in the existence of unit-root nonstationarities. Regarding the SC, it is adopted to complement the above criteria.

The $\overline{R}^2$ criterion is given by

$$\overline{R}^2 = 1 - \frac{\tilde{\sigma}^2_{t,i}}{S^2_{\rho,t}}$$

where $\tilde{\sigma}^2_{t,i}$ is the unbiased estimator of σ^2, which is given by $\tilde{\sigma}^2_{t,i} = \sum_{\tau=0}^{t} (\rho_{\tau+1} - X'_{\tau,i}\hat{\beta}_{t,i})^2 / (t - k_i - 1)$, $S^2_{\rho,t} = \sum_{\tau=1}^{t} (\rho_\tau - \overline{\rho}_\tau)^2 / (t-1)$ is the sample variance for the first t observations on ρ, and $\overline{\rho}_t = \sum_{\tau=1}^{t} \rho_\tau / t$. The AIC and BIC selection criteria are likelihood-based and assign different weights to "parsimony" and "fit" of the models. The "fit" is measured by the maximized value of the log-likelihood function (LL) as defined below, and the parsimony by the number of freely estimated coefficients. The AIC, BIC, FIC, and PIC can be defined as

$$LL_{t,i} = \frac{-t}{2}\{1 + \log(2\pi\tilde{\sigma}^2_{t,i})\}$$

$$\text{AIC} = LL_{t,i} - (k_i + 1)$$

$$\text{BIC} = LL_{t,i} - \frac{1}{2}(k_i + 1)\log(t)$$

$$\text{FIC} = SSE_{t,i}\frac{t}{t-k_i} + \frac{SSE_t}{t-k_i}\log\left(\frac{|X_{t,i}X'_{t,i}|}{\frac{SSE_{t,i}}{t-k_i}}\right)$$

$$\text{PIC} = SSE_t\left(\frac{SSE_{t,i}}{SSE_t} - 1\right) + \frac{SSE_t}{t-k_i}\log\left(\frac{|X_{t,i}X'_{t,i}|}{\frac{SSE_t}{t-k_i}}\right)$$

where $SSE_{t,i}$ indicates the residual sum of squares, $\sum_{\tau=0}^{t}(\rho_{\tau+1} - X'_{\tau,i}\hat{\beta}_{t,i})^2$, SSE_t is the largest residual sum of squares among model M_i at time t, k_i is the number of regressors, and t presents the number of observations within the sample period.

The criteria such as $\overline{R}^2$, FIC, and PIC select the model with the highest value as the best model while the AIC, and BIC prefer the model with the lowest one. The SC criterion is suggested on the grounds that investors in practice often are interested in predicting the changes in the sign of excess return function but not necessarily the magnitude of changes in the excess return. Using it to choose the best model involves two steps. In the first step, one finds the set of regressors that maximize the proportion of correctly predicted signs of the excess returns given by

$$SC_{t,i} = \frac{1}{t}\sum_{\tau=1}^{t}\{I(\rho_\tau)I(\hat{\rho}_{\tau,i}) + (1 - I(\rho_\tau))(1 - I(\hat{\rho}_{\tau,i}))\}$$

where $I(\rho_\tau)$ is an indicator function that takes the value of unity if $\rho_\tau > 0$, and zero otherwise, and $I(\hat{\rho}_{\tau,i})$ is the forecast of ρ_τ based on model M_i. If two or more models correctly predict the same (maximum) proportion of signs of excess returns, the second step selects a model recursively according to the $\overline{R}^2$ criterion.

The above six model selection criteria will be applied to the linear regressions. In the context of forecasting stock returns where the “true” model of the correct list of regressors is clearly unknown and may be changing over time, the consistency property of a model selection criterion is not as important as it may appear at first. Under the very weak assumption, data-generating process is not necessarily fixed throughout the sample period, and we simply use different model selection criteria to obtain appropriate forecasting equations at each point of time. With this procedure, all forecasting models under consideration would be equally likely selected and a particular model chosen at time t does not necessarily remain unchanged in the next period.

In view of the data availability in Taiwan stock market and the empirical evidence in the existent literature, we would like to include the following variables in our base set of regressors: the interest rate on one-month time deposit quoted by First Commercial Bank, the consumer price index, the industrial production index, the M1A money supply index, the price to earnings ratio, the dividend yield, the ratio of the short sales to margin purchase, the odd-lot trading volume, the net trading volumes of mutual funds, qualified foreign investment institutes, and security dealers, the one period lagged excess stock return, the closed-end fund discount, the turnover ratio, the ratio of the number of advancing issues to declining issues, the ARMS index, the number of equity initial public offerings, the number of initial public offering of warrants, the cash holding of mutual funds, and the ratio of net fund redemption. We do not consider this base set exhaustive but we believe that it is reasonably wide to encompass possible forecasting models meaningful to investors. We expect to collect the data for these 21 variables from the data bank of Taiwan Economic Journal and various data sources such as the data bank kept by Securities Investment Trust & Consulting Association of R.O.C. and the annual reports issued by Taiwan Stock

Exchange Corporation.

Because of the large set of relevant variables and the recursive nature of model estimation, we need to estimate a very large number of possible models and then select the best model according to different model selection criteria. This exhaustive specification search process may induce the risk of data mining or data snooping. To gauge this risk, we decide to conduct the extreme bound analysis suggested by Leamer (1983, 1985). Furthermore, following Pesaran and Timmermann (1995), we perform asset allocation strategies according to the forecast of the selected model and examine whether such strategies generate significant abnormal returns. Specifically, the strategy switches fund between the broad stock market index and the time deposit account. If this strategy brings investors significant abnormal returns, the risk of data mining is controlled at a negligible level.

3. Empirical Results

In this study, we collect data of two groups of variables: fundamental variables and market sentiment variables. We collect six macroeconomic indicators that have been shown to have certain forecasting power on short-term stock market returns. These variables include the first Commercial Bank interest rate on 1-month time deposit (IR), the Consumer Price Index (CPI), the Industrial Production Index (IPI), the money supply index (M1A), the price to earning ratio (P/E), and the dividend yield (YSI). Early studies on the predictability of stock returns are not always clear about the appropriate time lag between the changes in these fundamental variables and stock returns. Here, we include a one month lagged value for experiment, and all macroeconomic indicators are taken from Taiwan Economic Journal Data Base.

We collect monthly data of the sentiment variables from several sources. From Taiwan Economic Journal Data Bank (TEJ), we have the ratio of the short sales to margin purchase (SMR), the ratio of odd-lot sales to purchases (ODDLOT), the monthly net purchase from the local investment trusts (IT), the monthly net purchase from the qualified foreign institutional investors (QFTT), the monthly net purchase from the dealer department of local security firms (SD), the one-month lagged market excess return (LagR), the level of discount on closed-end funds (CEFD), the market turnover ratio (TURNOVER), the ratio of the number of advancing issues to declining issues (ADV/DEC), the modified ADV/DEC indicator which incorporates volumes (ARMS)[1], and the number of warrant issues (Warrant). The monthly aggregate cash holding of mutual funds (Fundcash) and the ratio of net fund redemption to total fund size (Fundflow) are available from Securities Investment Trust & Consulting Association of R.O.C. (SITCA) and the number of initial public offerings (IPON) is taken from the annual reports of Taiwan Stock Exchange Corporation. The sample period for both groups of variables is from January 1995 to December 2001.

[1] The ARMS is. This measure is the ratio of the number of advances to declines standardized by their respective volumes

$$ARMS_t = \frac{Adv_t / AdvVol_t}{DEC_t / DecVol_t}$$

Table Ⅰ shows summary statistics of our independent variables including fundamental and sentiment factors. From Table Ⅰ, the odd-lot trading volume has changed considerably over time and the three institutional investors also have the same characteristics that the standard deviations of their holding positions are relatively higher than other indicators. Furthermore, the QFII, on average, is on the position of net buy in Taiwan stock market and seems to have the opposed trading strategies with other two institutional investors, IT, and SD. During the period from 1995 to 2001, the CEFD is traded at a discount of about 16% with relatively small standard deviation. As far as the fundamental variables are concerned, the short-term interest rate has stayed at the level of about 4.78% and the dividend yield has also substantially increased from 2.5% in 1995 to about 6.5% in 2001. Maybe this is why the QFII is on the net buying position.

Table Ⅱ presents sample correlation coefficients among our independent and dependent variables. The most interesting feature in Table II is that there exist only weak correlations between the market excess returns and all independent variables. In addition, the correlations between most of our explanatory variables are reasonably low except fund related indices. For example, the correlation coefficient between the CEFD and FUNDCASH is 0.63 implying that the cash position of fund managers would adjust positively with the fund discount. The negative correlation between the IT and the FUNDFLOW also confirms that the net redemption ratio would drop when the investment trusts promote their shareholdings.

Although Table Ⅱ indicates that the correlations are small between most variables and market excess returns, it does not imply that our variables have no explanatory power in predicting market returns. On the contrary, the multicolinearity will not be the issue and since these variables are not highly correlated to the market excess returns in the next period based on the whole sample, our recursive regression form six selection criterion can explore the time-varying correlation and volatility between both fundamental and market sentiment variables and market excess returns.

Based on the recursive regression model, we separate our independent variables into two categories as fundamental and sentiment variables. Our purpose is to focus on the time-varying relations between these variables and the future market excess returns. We estimate in total 1,048,576 models over the period from July, 1998 to December, 2001, and investigate the influences of these two categories of variables, respectively. Notice that we use the period from January, 1995 to June, 1998 as the estimation period for the recursive regression models and evaluate the forecasting performance of the best models selected by six different model selection criteria during the period from July, 1998 to December, 2001. In other words, the performance of selected models is actually evaluated in a out-of-sample fashion.

Table Ⅲ presents the percentage of periods when a specific variable is included in the selected models. The variables that are chosen most frequently include IT, SD and YSI for models selected according to AIC, BIC, and $\overline{R}^2$ selection criteria. On average, the $\overline{R}^2$ criterion tends to select more variables than the AIC and BIC criteria. The former criterion imposes a lower penalty for the inclusion of an additional variable than the latter criteria. Specifically, IT, SD, and YSI usually have an inclusion rate of more than 50% across AIC, BIC, and $\overline{R}^2$. According to $\overline{R}^2$ criterion, IT has a proportion of 92.9% being included as a regressor in the estimation period, and the inclusion proportion of SD is 100%, and the YSI has an inclusion rate of 97.6%. Furthermore, based on the $\overline{R}^2$ selection criterion, SMR, CEFD, and P/E also have relatively higher percentages of inclusion, 92.9%, 59.5%, and 57.1%, respectively. Except for the AIC, BIC, and $\overline{R}^2$, almost twenty variables are frequently selected by the FIC selection criterion.

We also focus individually on the 14 sentiment variables and 6 fundamental variables to see which variables would be frequently chosen for predicting future market returns. The results are reported in the Panels B and Panel C on Table III, respectively. The Panel B shows that, according to $\overline{R}^2$ criterion, there are three additional variables, ODDLOT, Turnover rate, and Warrant, in a very close relation with market returns. The Panel C of Table III shows the relation between six fundamental indicators and market returns. According to the AIC, BIC, and $\overline{R}^2$ selection criteria, we find that the IR, viewed as a bearish indicator, is also in a close relation with the market returns. It implies that when short-term interest rate decreases, investors may take aggressive trading strategies. Under the $\overline{R}^2$ criterion, another important indicator, IPI, also has an inclusion percentage as high as 64.3%. The industrial production index has been chosen from early 1999 to mid 2001 and can be considered as a bullish predictor.

4. Conclusions

In recent studies, some researchers suggested that some popular measures of investor sentiment were closely related to stock market returns and some others emphasized the importance of fundamental variables. In this article, we reconcile these two streams of literature to examine the influences of these two groups of variables over one-period-ahead market returns and also take an approach to simulate the behavior of an investor in real time.

We adopt several statistical model selection criteria to determine the best models for predicting excess market returns through both fundamental and sentiment variables. The choice set consists of linear models with varying numbers of predictors. Most of significant variables we find are in a consistent relation with the market excess returns, except CEFD and P/E. According to $\overline{R}^2$ criterion, it is interesting to notice that several other variables, except IT, SD, and YSI, are selected by the best models. As far as the sentiment variables are concerned, we show that the short sale to margin purchase (SMR) could be reasonably considered as a bearish indicator for market returns. When focusing only on the sentiment variables, the ODDLOT, Turnover, and Warrants become very effective in predicting market returns. Regarding the fundamental variables, the IR and IPI are also occasionally selected in the best models for prediction. All of the frequently selected variables chosen to predict market excess returns are closely related to political regime switches, which happened around the Taiwan 2000 presidential election campaign according to the models selected based on $\overline{R}^2$ criterion.

The results suggest that investors may make their decisions based not only upon fundamental information but also on the signals of market sentiment. In particular, the net buy (sell) of institutional investors, such as SD and IT, are persistently chosen for predicting market returns throughout the whole period and the dividend yields (YSI), which are usually referred to be a fundamental factor, is almost constantly picked for prediction over long periods. In Taiwan stock market, the individual investors' trading, on average, has dominated more than 80% of market trading volume. Most individual investors may not only refer to the information about the status of the business cycle but also follow the trading behavior of institutional market participants. Our result suggests that individual investors should keep in mind the effects of trading strategies of institutional investors such as IT and SD who are usually considered as informed traders and are frequently able to capture the market sentiment.

However, our research displays that the degree of fitness of the best models decrease in unsettled periods such as some big political events. We also find that the degree of fitness improves from mid 2000, the beginning of global recession and the internet bubbles, which suggests that the degree of model fitness would positively vary with business cycle. As a result, even the focus of our analysis is to simulate investors' search in 'real time' for a model to forecast stock returns, the robustness of predicting performance would be seriously influenced by the major events such as political unsettlement and business cycle.

Reference

Akaike, H., 1974, A new look at the statistical identification model, IEEE Trans. Auto. Control 19, 716-723.

Brown, G. W., and M. T. Cliff, 2001, Investor sentiment and the near-term stock market, Working Paper, University of North Carolina at Chapel Hill.

Boudoukh, J., M. P. Richardson, and R. F. Whitelaw, 1994, A tale of three schools: Insights on autocorrelations of short-horizon security returns, Review of Financial Studies 7, 539-573.

Conrad, J., and G. Kaul, 1988, Time-variation in expected returns, Journal of Business 61, 409-425.

Conrad, J., and G. Kaul, 1989, Mean reversion in short-horizon expected returns, Review of Financial Studies 2, 225-240.

DeBondt, W., and R. Thaler, 1985, Does the stock market overreact?, Journal of Finance 40, 793-805.

Jegadeesh, N., 1990, Evidence of predictable behavior of security returns, Journal of Finance 45, 881-898.

Jegadeesh, N., 1991, Seasonality in stock price mean reversion: Evidence from the U.S. and the U.K., Journal of Finance 46, 1427-1444.

Fama, E. F., 1965, The behavior of stock market prices, Journal of Business 38, 34-105.

Fama, E. F., 1970, Efficient capital markets: A review of theory and empirical work, Journal of Finance 25, 383-417.

Fama, E. F., and K. R. French, 1988, Permanent and temporary components of stock prices, Journal of Political Economics 96, 246-273.

Granger, C. W. J., and O. Morgenstern, 1963, Spectral analysis of New York stock market prices, Kyklos 16, 1-27.

Leamer, E. E., 1983, Let's take the con out of econometrics, American Economic Review 73, 31-43.

Leamer, E. E., 1985, Sensitivity analyses would help, American Economic Review 75, 308-313.

Mandelbrot, B., 1966, Forecasts of future prices, unbiased markets, and 'martingale' models, Journal of Business 39, 394-419.

Pesaran, M. H., and A. Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, Journal of Finance 60, 1201-1228.

Pesaran, M. H., and A. Timmermann, 2000, A recursive modeling approach to predicting U.K. returns, Economic Journal 110, 159-191.

Phillips, P. C. B., and W. Ploberger, 1996, Posterior odds testing for a unit root with

data-based model, Econometrica 64, 381-412.
Samuelson, P. A., 1965, Proof that properly anticipated prices fluctuate randomly, Industrial Management Review 6, 41-49.
Schwartz, G., 1978, Estimating the dimension of a model, Annals of Statistics 6, 461-464.
Shiller, R. J., 1984, Stock prices and social dynamics, Brookings Papers on Economic Activity 2, 457-497.
Summers, L. H., 1986, Does the stock market rationally reflect fundamental values?, Journal of Finance 41, 591-601.
Wei, C., 1992, On predictive least square principles, Annals of Statistics 20, 1-42.
Working, H., 1949, The investigation of economic expectations, American Economic Review 39, 150-166.

Table I : Summary Statistics

The table shows summary statistics for the data used in the analysis. The full monthly sample contains 84 observations from January 1995 through December 2001.

	Mean	Std-dev	N
SMR	8.95	4.15	84.00
ODLOT	52.47	29.27	84.00
IPON	2.10	1.81	84.00
FUNDFLOW(9609-0112)	0.13	3.20	64.00
FUNDCASH(9705-0112)	7.67	1.64	56.00
IT	-21.58	81.50	77.00
QFII	8.80	17.75	84.00
SD	-7.00	47.79	84.00
LagR	-4.68	9.25	84.00
CEFD	-16.24	2.82	84.00
TURNOVER	20.82	8.37	84.00
ADV/DEC	2.22	3.74	84.00
ARMS	0.18	0.56	84.00
WARRANT(9709-0112)	3.48	3.95	52.00
IR	4.78	0.87	84.00
CPI	1.59	1.57	84.00
IPI	3.76	6.88	84.00
M1A	2.45	6.49	84.00
P/E	29.28	10.78	84.00
YSI	3.62	1.01	84.00
EXCESS RETURNS	-4.43	9.20	84.00

Table II : Correlation Coefficients

Pairwise correlations for independent variables used in the analysis and those with market excess returns.

	SMR	ODDLOT	IPON	FUNDFLOW	FUNDCASH	IT	QFII	SD	LAGR	CEFD	TURNOVER	ADV/DEC	ARMS	WARRANT	IR	CPI	IPI	M1A	P/E	Dividend	E.S. Return
SMR	1.00	-0.25	-0.19	-0.14	0.25	0.04	-0.02	0.18	0.00	0.25	-0.24	0.03	0.13	0.05	-0.15	-0.03	0.03	-0.25	-0.31	0.59	0.17
ODDLOT	-0.25	1.00	-0.22	0.06	0.29	-0.21	0.41	0.14	0.45	0.12	0.47	0.36	0.05	0.69	-0.34	-0.19	-0.26	0.36	0.67	-0.22	0.07
IPON	-0.19	-0.22	1.00	-0.04	-0.32	0.05	-0.14	0.04	-0.15	-0.09	-0.06	0.02	0.09	-0.28	0.19	-0.14	0.02	-0.11	-0.19	-0.05	-0.02
FUNDFLOW	-0.14	0.06	-0.04	1.00	-0.09	-0.48	0.05	0.01	0.01	-0.21	0.21	-0.09	-0.03	0.02	0.01	-0.08	-0.02	-0.05	0.08	-0.01	0.15
FUNDCASH	0.25	0.29	-0.32	-0.09	1.00	-0.01	0.16	-0.13	-0.10	0.63	-0.34	-0.14	-0.15	0.40	-0.42	-0.07	-0.02	-0.02	0.05	0.38	0.01
IT	0.04	-0.21	0.05	-0.48	-0.01	1.00	-0.02	0.07	0.10	0.14	-0.27	0.04	0.13	-0.25	-0.01	0.10	-0.19	-0.21	-0.24	0.16	-0.20
QFII	-0.02	0.41	-0.14	0.05	0.16	-0.02	1.00	0.41	0.47	0.27	-0.01	0.22	0.09	0.45	-0.38	-0.12	-0.10	-0.10	0.29	0.12	0.13
SD	0.18	0.14	0.04	0.01	-0.13	0.07	0.41	1.00	0.40	0.09	0.08	0.25	0.10	0.27	-0.09	-0.08	0.01	-0.01	0.16	-0.01	0.26
LAGR	0.00	0.45	-0.15	0.01	-0.10	0.10	0.47	0.40	1.00	-0.03	0.44	0.68	0.47	0.29	-0.21	-0.01	-0.23	0.05	0.30	-0.11	0.06
CEFD	0.25	0.12	-0.09	-0.21	0.63	0.14	0.27	0.09	-0.03	1.00	-0.37	-0.02	-0.07	0.25	-0.40	-0.07	-0.03	-0.27	-0.14	0.47	-0.07
TURNOVER	-0.24	0.47	-0.06	0.21	-0.34	-0.27	-0.01	0.08	0.44	-0.37	1.00	0.40	0.10	-0.03	0.32	-0.04	-0.19	0.33	0.38	-0.55	-0.05
ADV/DEC	0.03	0.36	0.02	-0.09	-0.14	0.04	0.22	0.25	0.68	-0.02	0.40	1.00	0.75	0.30	-0.12	-0.12	-0.19	0.02	0.12	-0.06	0.01
ARMS	0.13	0.05	0.09	-0.03	-0.15	0.13	0.09	0.10	0.47	-0.07	0.10	0.75	1.00	0.10	-0.16	-0.11	-0.07	-0.20	-0.07	0.14	0.02
WARRANT	0.05	0.69	-0.28	0.02	0.40	-0.25	0.45	0.27	0.29	0.25	-0.03	0.30	0.10	1.00	-0.52	-0.11	-0.14	0.21	0.48	0.07	0.12
IR	-0.15	-0.34	0.19	0.01	-0.42	-0.01	-0.38	-0.09	-0.21	-0.40	0.32	-0.12	-0.16	-0.52	1.00	0.13	-0.02	0.06	-0.15	-0.51	-0.26
CPI	-0.03	-0.19	-0.14	-0.08	-0.07	0.10	-0.12	-0.08	-0.01	-0.07	-0.04	-0.12	-0.11	-0.11	0.13	1.00	0.02	0.07	-0.02	-0.10	-0.17
IPI	0.03	-0.26	0.02	-0.02	-0.02	-0.19	-0.10	0.01	-0.23	-0.03	-0.19	-0.19	-0.07	-0.14	-0.02	0.02	1.00	0.25	0.08	-0.18	-0.01
M1A	-0.25	0.36	-0.11	-0.05	-0.02	-0.21	-0.10	-0.01	0.05	-0.27	0.33	0.02	-0.20	0.21	0.06	0.07	0.25	1.00	0.60	-0.64	-0.04
P/E	-0.31	0.67	-0.19	0.08	0.05	-0.24	0.29	0.16	0.30	-0.14	0.38	0.12	-0.07	0.48	-0.15	-0.02	0.08	0.60	1.00	-0.59	0.04
Dividend	0.59	-0.22	-0.05	-0.01	0.38	0.16	0.12	-0.01	-0.11	0.47	-0.55	-0.06	0.14	0.07	-0.51	-0.10	-0.18	-0.64	-0.59	1.00	0.22
E.S. Return	0.17	0.07	-0.02	0.15	0.01	-0.20	0.13	0.26	0.06	-0.07	-0.05	0.01	0.02	0.12	-0.26	-0.17	-0.01	-0.04	0.04	0.22	1.00

Table III : Percentage of Periods Where a Regressor is Included in Forecasting Equations

The results reports the inclusion frequency of both sentiment and fundamental indicators used in the linear regression based on six selection criteria. All the results are estimated recursively over the period from 1998(7) to 2001(12).

Panel A

All twenty indicators are input into the selection criteria for the linear regressions

Selection Criterion	**Percentages**									
	SMR	ODDLOT	IPON	FUNDFLOW	FUNDCASH	IT	QFII	SD	LAGR	CEFD
AIC	11.9%	0.0%	0.0%	0.0%	0.0%	61.9%	0.0%	83.3%	0.0%	2.4%
BIC	31.0%	0.0%	0.0%	0.0%	0.0%	83.3%	0.0%	97.6%	0.0%	4.8%
$\overline{R}^2$	92.9%	14.3%	16.7%	11.9%	30.9%	92.9%	35.7%	100.0%	16.7%	59.5%
SC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
FIC	95.2%	100.0%	100.0%	100.0%	90.5%	33.3%	100.0%	0.0%	100.0%	100.0%
PIC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	TURNOVER	ADV/DEC	ARMS	WARRANT	IR	CPI	IPI	M1A	P/E	Dividend
AIC	0.0%	0.0%	0.0%	7.1%	21.4%	0.0%	0.0%	7.1%	0.0%	45.2%
BIC	0.0%	0.0%	0.0%	11.9%	11.9%	0.0%	0.0%	11.9%	2.4%	59.5%
$\overline{R}^2$	4.8%	28.6%	11.9%	50.0%	35.7%	0.0%	26.2%	33.3%	57.1%	97.6%
SC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
FIC	97.6%	100.0%	100.0%	78.6%	26.2%	100.0%	57.1%	100.0%	97.6%	14.3%
PIC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.7%	0.0%	0.0%	0.0%

Panel B

The inclusion percentages of fourteen sentiment variables used in linear regression

Selection Criterion	Percentages													
	SMR	ODDLOT	IPON	FUNDFLOW	FUNDCASH	IT	QFII	SD	LAGR	CEFD	TURNOVER	ADV/DEC	ARMS	WARRANT
AIC	0.0%	0.0%	0.0%	2.4%	0.0%	54.8%	0.0%	97.6%	0.0%	0.0%	0.0%	0.0%	0.0%	14.3%
BIC	0.0%	0.0%	2.4%	0.0%	0.0%	73.8%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	14.3%
$\overline{R}^2$	30.9%	90.5%	0.0%	38.1%	2.4%	100.0%	2.4%	100.0%	0.0%	28.6%	100.0%	45.2%	2.4%	97.6%
SC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
FIC	100.0%	100.0%	100.0%	92.9%	90.5%	21.4%	95.2%	0.0%	100.0%	100.0%	95.2%	100.0%	100.0%	61.9%
PIC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Panel C

The inclusion percentages of six fundamental variables used in linear regression

Selection Criterion	Percentages					
	IR	CPI	IPI	M1A	P/E	Dividend
AIC	57.1%	0.0%	0.0%	0.0%	0.0%	0.0%
BIC	61.9%	0.0%	0.0%	0.0%	4.8%	0.0%
$\overline{R}^2$	73.8%	14.3%	64.3%	12.2%	42.9%	59.5%
SC	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
FIC	35.7%	97.6%	64.3%	100.0%	100.0%	40.5%
PIC	0.0%	0.0%	0.0%	0.0%	2.4%	0.0%