Index Pricing Model Embedding Investor Sentiment: Based on Path Converged Design

YANYUN YAO*
Zhejiang Gongshang University
Research Institute of Econometrics and Statistics
18 Xuezheng Road
Xiasha University Town, Hangzhou
CHINA
nan_xiao@126.com

BING XU
Zhejiang Gongshang University
Research Institute of Econometrics and Statistics
18 Xuezheng Road,
Xiasha University Town, Hangzhou
CHINA
bingxu0@yahoo.com.cn

Abstract: This paper proposes an index pricing method to model the risk premium of market, industry and area for individual stocks. The individual stocks of Shanghai Index 50 in China's stock market are selected. Path converged design is employed to construct three time-varying coefficient semi-parameter regression models for each stock. The three models are base model excluding investor sentiment, path model I including investor sentiment of each individual stock and path model II including investor sentiments of indexes. As the empirical result, the path models explain the fluctuation of individual stocks better than the base model, which means that investor sentiment is an important factor for asset pricing. Furthermore, pricing errors analysis is carried out. The result shows that index pricing model plays an important role in the determination of individual stock's rational prices and it may contribute to trading strategy construction for programming trading. The paper is an effective modeling trial of asset pricing in the nonparametric frame.

Key–Words: Investor sentiment, Index pricing model, Path converged design, Pricing error analysis, Nonparametric estimation

1 Introduction

Asset pricing is one of the important fields of financial research, and is the core of modern portfolio theory. The capital asset pricing model (CAPM), which is proposed by Sharpe (1964), Lintner (1965) and Mossin (1966), is the benchmark of asset pricing models and most empirical studies apply it to calculate asset returns and cost of capital [1]. For the limitations of its critical assumptions, such as perfect competitive and equilibrium market, rational and homogenous investors, many researchers have tried to develop more general asset pricing models by relaxing the assumptions of CAPM. Ross (1976) proposes arbitrage pricing theory (APT), and then many multi-factor models are derived. The famous one of them is three-factor model put forward by Fama and French (1995), the three risk factors are market, size (SIZE) and book-to-market (BM) [2]. But for the model, Daniel and Titman (1997) regards SIZE and BM as characteristics of firms, which reflect the preference of investors, rather than risk factors [3]. Different viewpoints of the same model are due to their different identity theories: rational pricing theory and irrational pricing theory. Rational pricing theory considers that stock returns are influenced by the risks. Irrational pricing theory holds that stock returns are affected by the irrational factors such as cognitive biases and personal preferences of investors. Argument between the two theories is an argument between traditional and behavioral financial theories essentially [4]. Investor sentiment is defined as stock pricing belief of irrational investors in Barberis, Shleifer and Vishny (1998) [5], which means that investor sentiment is an important concept in behavioral financial theory. S. Wang, et al.(2013) concludes that the effect of investor sentiment on real economic activities is an important issue [6]. Brown and Cliff (2005) demonstrates that investor sentiment is always an important factor in predicting future market returns [7]. Similarly, it is generally believed that stock markets are motivated by the irrational investors, due to asymmetric information or their per-

*The author has another address: Department of Mathematics, Shaoxing University,900 Chengnan Avenue, Shaoxing, China sonal knowledge limitation, thus unable to identify the stock prices, and their irrational tradeoffs explain the strong fluctuations beyond the company fundamentals [8]. The recognition of the importance of investor sentiment may be useful in explaining some of the anomalies and empirical results that are defined relative to an asset-pricing model such as the CAPM. For example, Fama and French (1992) find that, even after adjusting for risk, returns are higher on stocks of small-capitalization firms [9]. If small stocks are characterized by a higher sentimental noise trading activity, then systematic sentiment-trader risk can help explain the size effect.

Since the advent of CAPM, characteristics and sentiments of investors have been considered into more and more academic researches on asset pricing, no matter in the traditional approach as well as the behavioral approach. The important works include Kyle (1985), Black (1986), Baker and Wurgler (2006), Basak (2005), Levy et al. (2006), Yoel (2009), and so on [10-15]. The results of Kyle (1985) imply that sentiment does not have a systematic influence on asset prices [10]. But Black (1986) concludes that pricing errors arising from sentiment are systematic across investors [11]. Baker and Wurgler (2006) present empirical evidence of a significant cross-sectional relation between future stock returns and beginning-of-period market sentiment [12]. There is strong merit in studying the influence of investor sentiment on traditional asset-pricing models such as the CAPM. Basak (2005), Levy et al. (2006) and Yoel (2009) consider equilibrium models with investors' heterogeneity by using static CAPM [13-15]. In addition, for the close relation between investor sentiment and liquidity, investor sentiment can be considered in some liquidity-based pricing models, such as Pastor and Stambaugh (2003) [16], Acharya and Pedersen (2005) [17] and Zhou and Zhang (2011) [18].

Up to now, many asset pricing literatures are based on cross section analysis and considering the company's financial information, such as [2, 9, 13, 18], etc. In order to coincide with the frequency of financial data, prices or returns of high frequency are often abandoned, which shows disadvantage to investigate the actual behavior of investors. For the above reason, this article intends to study from the perspective of time series, so that the characteristic of time variation in risk-factor loadings can be explored. The modeling methods in asset pricing literatures can be divided in two types: parameter and nonparameter. Parametric method is used in most of the articles, such as Fama and French (1992, 1995) [2, 9], Zhou and Zhang (2011) [18], etc. But Erdos et al. (2011) [19], Ferreira et al. (2011) [20] and Hua (2010) [21] use nonparametric method. Different from parametric method, nonparametric method without prior assumptions on the model, thus the model specification bias is reduced. For the reason, a nonparametric frame of the model is employed in this article.

Many assumptions of CAPM can't satisfy in real markets, but it is still the benchmark of asset pricing models, market is still the most important risk factor. In addition, industry and area are two important factors for empirical researches in China's stock market. For example, Li et al. (2009) and Zhu et al. (2011) build three-factor parametric model to investigate the China's stock market, the factors are market, industry and area risks [22-23]. Li et al. (2009) suggests that the political and economic characteristics of the provincial regions contribute to the segmentation of China's A-share market and hence affect the pricing of A-share stocks, which leads to that area effect should be considered to asset pricing in China's stock market [22]. As a result of Zhu et al. (2011), the stock returns are close related to industry, area and market information, and industry information dominates stock price movement [23]. Modeling stock market is to provide valuable information about possible models of the return series, so as to understand the behavior of the stock market [24]. As we know, industrial and regional development is an important strategy for China's economy, which is another reason for industry and area effects considered in China's stock market study. For the above reasons, three factors -market, industry and area risks are considered in our model. How to model the three factors - market, industry and area risks from the perspective of time series? An index pricing model is taken into account. Index pricing means making a price to an individual or a portfolio by using indexes information, it is from the perspective of information economics and market transaction, rather than the traditional cost pricing perspective. In fact, when the market portfolio is a market index, CAPM is a single index pricing model. When market, industry and provincial area indexes are selected to represent the three risk factors, a three-factor index pricing framework is built.

In order to consider investor sentiment, its measurement is a key step. For it often can not be measured directly, many researches model it in an indirect way. Wang and Zhang (2012) indicate that some liquidity measurement including trading volumes or turnovers can be the proxy of investor sentiment [8]. So, the liquidity measurements mentioned in Zhou and Zhang (2011)[18] is selected. In addition, Zhou and Zhang (2011) concludes that, liquidity has intrinsic correlation with BM and SIZE of company and two-factor model based on liquidity and market risks still has a good explanatory ability to China's stock market [18], which provides a support to our model frame.

The current paper reviews the asset pricing models and investor sentiment starting from CAPM briefly. We derive a testable model that incorporates the influence of investor sentiment on asset prices. The main contribution of this paper is presenting a new pricing model named index pricing model from the perspective of information economics and market transaction, which is in the nonparametric framework, rather than parametric method. The remainder of this paper is organized as follows: Section 2 outlines the index pricing model incorporating investor sentiment. Section 3 is an implication of our model in China's stock market. Conclusions about the index pricing model and the empirical study are drawn in section 4.

2 Index Pricing Model

2.1 Model Construction

The capital asset pricing model (CAPM) proposed by Sharpe (1964), Lintner (1965) and Mossin (1966) is described by:

$$R_{j,t} = \alpha_j + r_{f,t} + \beta_j(R_{M,t} - r_{f,t}) + \varepsilon_{j,t}, \quad (j = 1, 2, \ldots, N, t = 1, 2, \ldots, T) \tag{1}$$

where $R_{j,t}$ and $R_{M,t}$ denote the return on asset j and market return from time $t-1$ to t, $r_{f,t}$ denotes the risk-free interest rate. Denote $R_{j,t} - r_{f,t}$and$R_{M,t} - r_{f,t}$ as $r_{j,t}$and$r_{M,t}$, then formula (1) is:

$$r_{j,t} = \alpha_j + \beta_j r_{M,t} + \varepsilon_{j,t}. \tag{2}$$

In CAPM, market is the only source of systematic risk, which is often not corresponding with the actual markets.

Fama and French (1995) [2] improved the CAPM and put forward the famous three-factor model, it is represented as:

$$r_{j,t} = \alpha_j + \beta_{1,j} r_{M,t} + \beta_{2,j} SML_t + \beta_{3,j} HML_t + \varepsilon_{j,t}, \tag{3}$$

where SML_t and HML_t reflect the SIZE and BM factors.

As to China's stock market, a three-factor model demonstrated in Li et al. (2009) and Zhu et al. (2011) follows:

$$R_{j,t} - r_{f,t} = \alpha_j + \beta_M(R_{M,t} - r_{f,t}) + \beta_I(R^I_{j,t} - r_{f,t}) + \beta_A(R^A_{j,t} - r_{f,t}) + \varepsilon_{j,t}, \tag{4}$$

where $R^I_{j,t}$ and $R^A_{j,t}$ represent the industry and area returns on asset j from $t-1$ to t. Generally, $R^I_{j,t} - r_{f,t}$ and $R^A_{j,t} - r_{f,t}$ will commove with $R_{M,t} - r_{f,t}$, which means multicollinearity in formula (4). An alternative method is eliminating the influence of $R_{M,t}$ from $R^I_{j,t}$ and $R^A_{j,t}$ by CAPM.

Assume $R^I_{j,t} - r_{f,t} = \hat{\beta}^I_j(R_{M,t} - r_{f,t}) + \varepsilon^I_{j,t}$, and denote $r^I_{j,t} = R^I_{j,t} - r_{f,t} - \hat{\beta}^I_j(R_{M,t} - r_{f,t})$ as the industry return in excess of market. Similarly, the area return in excess of market is $r^A_{j,t} = R^A_{j,t} - r_{f,t} - \hat{\beta}^A_j(R_{M,t} - r_{f,t})$. So, formula (4) turns into:

$$r_{j,t} = \alpha_j + \beta_M r_{M,t} + \beta_I r^I_{j,t} + \beta_A r^A_{j,t} + \varepsilon_{j,t}. \tag{5}$$

Obviously, (5) is a parametric model with static beta coefficients. But stock markets often suffer from event shocks and show time-varying behavior of long-run [25]. Thus, stock markets are often taken as dynamic systems, such as in [26]. In order to explore the characteristic of time variation in risk-factor loadings, we propose our first model named base model:

$$r_{j,t} = \alpha_1(t) r_{M,t} + \alpha_2(t) r^I_{j,t} + \alpha_3(t) r^A_{j,t} + F(T_{j,t}) + \varepsilon^1_{j,t}. \tag{6}$$

where $\alpha_1(t)$, $\alpha_2(t)$ and $\alpha_3(t)$ are time-varying beta coefficients corresponding with $r_{M,t}$, $r^I_{j,t}$ and $r^A_{j,t}$, $F(T_{j,t})$ is an unknown function and $T_{j,t}$is a time variable.

Is (6) an appropriate model? We don't know, for we don't understand the latent structure of stock returns. Xu (2010) proposes a modeling method bypassing model specification bias, named path converged design. In the method, three models: latent structure, base model and path model are built. These models are connected by path factors, equivalent with the 1 probability under the large sample [27]. In empirical study, base model and path model are considered, by comparison of the two models to draw conclusions. In this sense, we call (6) as base model by using path converged design.

Furthermore, investor sentiment is considered in the framework of (6). Firstly, investor sentiment of individual stock is taken into account. Two liquidity measurements: turnover rate and illiquidity indicator are selected as proxy variables. Turnover rate can be obtained from Internet or trading terminals.

Illiquidity indicator is defined as Zhou and Zhang (2011) [10]. It measures the price varying range caused by turnover in a unit of time and it can reflect the trade impact on prices more directly. Illiquidity indicator is defined as:

$$IL_{j,t} = \frac{\frac{Hp_{j,t} - Lp_{j,t}}{Op_{j,t}}}{M_{j,t}} \tag{7}$$

where $Hp_{j,t}$, $Lp_{j,t}$, $Op_{j,t}$ and $M_{j,t}$ represent the highest price, lowest price, opening price and turnover on asset j from $t-1$ to t respectively. Denote turnover rate as $Tr_{j,t}$ and illiquidity indicator as $IL_{j,t}$, path model I is acquired:

$$\begin{aligned} r_{j,t} = \beta_1(t) r_{M,t} + \beta_2(t) r^I_{j,t} + \beta_3(t) r^A_{j,t} \\ + G(T_{j,t}, Tr_{j,t}, IL_{j,t}) + \varepsilon^2_{j,t}. \end{aligned} \tag{8}$$

where $G(.)$ is an unknown function, $Tr_{j,t}$ and $IL_{j,t}$ are path factors, reflecting the individual stock's investor sentiment.

Secondly, investor sentiment of market, industry and area indexes are considered. We employ illiquidity indicator defined as (7) to measure the investor sentiment of the indexes, and sign them as M_IL_t, $I\_IL_{j,t}$ and $A\_IL_{j,t}$. Then we get the path model II:

$$\begin{aligned} r_{j,t} = \gamma_1(t) r_{M,t} + \gamma_2(t) r^I_{j,t} + \gamma_3(t) r^A_{j,t} \\ + H(T_{j,t}, M\_IL_t, I\_IL_{j,t}, A\_IL_{j,t}) + \varepsilon^3_{j,t}. \end{aligned} \tag{9}$$

In the framework of path converged design, in the case of large sample, base model, path model I and path model II are equivalent. So, comparisons between (8) and (6), (9) and (6) can be used to analyze the influence of path factors: investor sentiment. Furthermore, comparison between (8) and (9), the efficiency of index pricing method can be tested by some design of simulated trading strategy.

2.2 Nonparametric Estimation of the Model

(1) Partial Residuals Estimation

In this paper, formula (6), (8) and (9) are called as base model, path model I and path model II respectively, and they can be considered as semi-parametric models. Referring to Li and Ye (2000) [28], we use partial residual error estimation to solve them. The solving method can be decomposed as three steps. (Take formula (6) as the example.)

Step 1: Estimate the parametric part, i.e. estimate

$$r_{j,t} = \alpha_1(t) r_{M,t} + \alpha_2(t) r^I_{j,t} + \alpha_3(t) r^A_{j,t} + \varepsilon^1_{j,t}, \tag{10}$$

thus the estimators of $\alpha_{i,t} (i = 1, 2, 3)$ are gotten, which are regarded as their preset values.

Step2: Estimate the nonparametric part $\hat{F}(T_{j,t})$. Note

$$u_{j,t} = r_{j,t} - \hat{\alpha}_1(t) r_{M,t} - \hat{\alpha}_2(t) r^I_{j,t} - \hat{\alpha}_3(t) r^A_{j,t},$$

then

$$\hat{F}(T_{j,k}) = \sum_{t=1}^{N_j} W_{N_j,t}(T_{j,k}) u_{j,t},$$

the weight function

$$W_{N_j,t}(T_{j,k}) = \frac{K(\frac{T_{j,k} - T_{j,t}}{h})}{\sum_{l=1}^{N_j} K(\frac{T_{j,k} - T_{j,l}}{h})},$$

where $K(\cdot)$ represents kernel function, h is the bandwidth of $\{T_{j,t}\}$.

Step 3: Estimate $\alpha_{i,t} (i = 1, 2, 3)$ again based on the known $\hat{F}(T_{j,t})$. Based on

$$\begin{aligned} r_{j,t} - \hat{F}(T_{j,t}) = \alpha_1(t) r_{M,t} + \alpha_2(t) r^I_{j,t} \\ + \alpha_3(t) r^A_{j,t} + \varepsilon^1_{j,t}, \end{aligned} \tag{11}$$

$\alpha_{i,t} (i = 1, 2, 3)$ are estimated by Step1, accompanied by the variance estimator of the model

$$\hat{\sigma}^2_j = \frac{1}{N_j} \sum_{t=1}^{N_j} (\varepsilon^1_{j,t})^2.$$

In the process of model solving, the kernel function $K(\cdot)$ is Gaussian, the bandwidth of $\{T_{j,t}\}$ is taken as constant

$$h_{1,j} = 1.06 \times \sigma_j \times N_j^{-\frac{1}{5}},$$

where σ_j is the standard deviation of $\{T_{j,t}\}$, N_j is the sample size.

Due to the significant deviating from uniform, turnover rate and illiquidity indicator use time-varying bandwidths. According to [29] and [30], the time-varying bandwidth is:

$$h_n(x) = [\frac{d \cdot \sigma^2(x)}{2^d n \pi^{\frac{d}{2}} f(x) tr^2\{H_m(x)\}}]^{\frac{1}{4+d}}$$

where d is the dimension, n is the sample size, $f(x)$ is the kernel density function, $H_m(x)$ is Hessen matrix of the latent true function $m(x)$, and $tr\{\cdot\}$ is the trace operation.

(2) Estimation of Time-varying Coefficient Model

In the upper segment, Step 1 and Step 3 are related to an estimation of time-varying coefficient model. Taking (10) as the example, we give the solving process.

For the first order Taylor expansion at t_0,

$$\alpha_i(t) = a_i(t_0)+b_i(t_0)(t-t_0)+o(t-t_0), (i = 1, 2, 3)$$

because $\lim_{t\to t_0} o(t-t_0) = 0$, then

$$\alpha_i(t) \approx a_i(t_0) + b_i(t_0)(t - t_0).$$

Formula (10) can be represented approximately as:

$$\begin{aligned} r_{j,t} = &[a_1(t_0) + b_1(t_0)(t - t_0)]r_{M,t} \\ &+[a_2(t_0) + b_2(t_0)(t - t_0)]r^I_{j,t} \\ &+[a_3(t_0) + b_3(t_0)(t - t_0)]r^A_{j,t} + \varepsilon^1_{j,t}, \end{aligned} \tag{12}$$

Note $X_j =$

$$\begin{bmatrix} r_{M,1} & (1-t_0)r_{M,1} & r^I_{j,1} & (1-t_0)r^I_{j,1} & r^A_{j,1} & (1-t_0)r^A_{j,1} \\ r_{M,2} & (2-t_0)r_{M,2} & r^I_{j,2} & (2-t_0)r^I_{j,2} & r^A_{j,2} & (2-t_0)r^A_{j,2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{M,N_j} & (N_j-t_0)r_{M,N_j} & r^I_{j,N_j} & (N_j-t_0)r^I_{j,N_j} & r^A_{j,N_j} & (N_j-t_0)r^A_{j,N_j} \end{bmatrix}$$

$$B_j = [a_1(t_0)\ \ b_1(t_0)\ \ a_2(t_0)\ \ b_2(t_0)\ \ a_3(t_0)\ \ b_3(t_0)]^T,$$
$$Y_j = [r_{j,1},\ \ r_{j,2}, \cdots , r_{j,N_j}]^T,$$

thus $Y_j = X_jB_j$ is the matrix representation of (12). Employ weighted least square (WLS) method, we get the estimation:

$$\hat{B}_j = \arg\min_{B_j}(Y_j - X_jB_j)^T W_j(Y_j - X_jB_j), \tag{13}$$

where W_j is kernel weighted matrix,

$$W_j = \begin{bmatrix} K_h(1-t_0) & & & \\ & K_h(2-t_0) & & \\ & & \ddots & \\ & & & K_h(N_j-t_0) \end{bmatrix},$$

h is the bandwidth vector of X_j.

According to the first-order conditions of the optimal problem, we obtain

$$\hat{B}_j = (X^T_j W_jX_j)^{-1}X^T_j W_jY_j. \tag{14}$$

$\hat{B}_j$ is a function of the expansion point t_0. When $t_0 = 1, 2, \cdots , N_j$ in turn, $\hat{B}_j(1), \hat{B}_j(2), \cdots , \hat{B}_j(N_j)$ is obtained successively. When $t_0 = t_k$, the time-varying coefficients are obtained:

$$\hat{\alpha}_i(t_k) = a_i(t_k) + b_i(t_k)(t_k - t_k) = a_i(t_k)\ .$$

3 Empirical Study

3.1 Data

As we know, the sample stocks of Shanghai Index 50 have large scale and good liquidity, and they can reflect the overall situation of the most influential stocks with high quality in the Shanghai stock market. Thus, the sample stocks of Shanghai Index 50 are selected in this paper. The time spans from August 4, 2010 to April 10, 2013. In order to keep consistency, all prices are disposed by forward answer authority method. The Shanghai composite index is used as the market index. The industry sector indexes corresponding to the 50 sampled stocks are selected as the industry indexes, including 19 sectors. Area indexes are selected from the province sector indexes where the 50 sampled stocks registered, including 15 provinces, see Table 1. The selected data include the opening price, the highest and the lowest price, closing price, trading volume, turnover and turnover rate. They are all daily data. (All data are obtained from the Qianlong securities investment and analysis software.)

The stock returns are using the logarithmic yield, that is:

$$\begin{aligned} &R_{j,t} = (lnP_{j,t} - lnP_{j,t-1}) \times 100, \\ &(j = 1, 2, \cdots , 50, t = 1, 2, \cdots , N_j) \end{aligned} \tag{15}$$

where $P_{j,t}$ denotes the closing price of the j stock at the time t, N_j denotes the sample size of the j stock.

The stock illiquidity is defined by formula (7). Measurements of the returns and illiquidity of market, industry and area indexes are similar to the individual stocks. To achieve the comparability of illiquidity and return, range transformation is implemented to the illiquidity indicator. Time variable $T_j = [1, 2, \cdots , N_j]$ is implemented by range transformation too. For a series $\{X_t\}_{t=1:n}$, range transformation is:

$$Y_t = \frac{X_t - \min_{1\le t\le n}(X_t)}{\max_{1\le t\le n}(X_t) - \min_{1\le t\le n}(X_t)}, t = 1, 2, \cdots , n. \tag{16}$$

The risk-free return $r_{f,t}$ is using the date yield of one-year deposit interest rate, its data are from CSMAR database in China.

3.2 Correlation Analysis

Firstly, we calculate the Pearson correlation coefficients of the variables $r_{M,t}$, $r^I_{j,t}$, $r^A_{j,t}$, $Tr_{j,t}$, $IL_{j,t}$, T_j, $M\_IL_{j,t}$, $I\_IL_{j,t}$ and $A\_IL_{j,t}$. Correlation results

Table 1: Individual stocks, industry sectors and area sectors

No.	Individual	Industry	Area	No.	Individual	Industry	Area
1	BGGF	Steel	Neimenggu	26	SDHJ	Non-ferrous metal	Shandong
2	BGXT	Non-ferrous metal	Neimenggu	27	SQJT	Automobile	Shanghai
3	BLDC	Real estate	Guangdong	28	WKFZ	Trade and services	Beijing
4	BGGF	Steel	Shanghai	29	XYYH	Bank	Fujian
5	BJYH	Bank	Beijing	30	YQMY	Coal industry	Shanxi
6	CJDL	Power	Beijing	31	YLGF	Food and beverage	Neimenggu
7	DQTL	Transportation and logistics	Shanxi	32	ZSYH	Bank	Guangdong
8	FZZQ	Security and insurance	Hunan	33	ZGBC	Machinery	Beijing
9	GZB	Construction	Hubei	34	ZGJZ	Construction	Beijing
10	GSYH	Bank	Beijing	35	ZGLT	Communication	Shanghai
11	GDYH	Bank	Beijing	36	ZGLY	Non-ferrous metal	Beijing
12	GHNY	Comprehensive	Xinjiang	37	ZGNC	Machinery	Beijing
13	GZMT	Wine brewing	Guizhou	38	ZGPA	Security and insurance	Guangdong
14	HLSN	Building material	Anhui	39	ZGRS	Security and insurance	Beijing
15	HNXJ	Auto parts	Hainan	40	ZGSH	Coal industry	Beijing
16	HTZQ	Security and insurance	Shanghai	41	ZGSH	oil and gas	Beijing
17	HXYH	Bank	Beijing	42	ZGSY	oil and gas	Beijing
18	JXTY	Non-ferrous metal	Jiangxi	43	ZGSD	Construction	Beijing
19	JTYH	Bank	Shanghai	44	ZGTB	Security and insurance	Shanghai
20	JMGF	Non-ferrous metal	Shanxi	45	ZGZG	Machinery	Beijing
21	LAHN	Coal industry	Shanxi	46	ZJHJ	Non-ferrous metal	Beijing
22	MSYH	Bank	Beijing	47	ZMNY	Coal industry	Beijing
23	NYYH	Bank	Beijing	48	ZXZQ	Security and insurance	Guangdong
24	PFYH	Bank	Shanghai	49	ZJKY	Non-ferrous metal	Fujian
25	SYZG	Machinery	Beijing	50	YZMY	Coal industry	Shandong

of the 50 individual stocks show the similar characteristics. Take No.1 stock - BGGF as example, the correlation matrix is shown in Table 2.

From Table 2, all variables have low correlations to the market, industry and area indexes' returns, but the correlations between$Tr_{j,t}$, $IL_{j,t}$, T_j, $M\_ IL_{j,t}$, $I\_ IL_{j,t}$ and $A\_ IL_{j,t}$ show high. The result implies that there may exist multicollinearity problem in the variables $Tr_{j,t}$, $IL_{j,t}$, T_j, $M\_ IL_{j,t}$, $I\_ IL_{j,t}$ and $A\_ IL_{j,t}$, some information of them is redundant and they may be alternative. This result may provide support for the construction of path model I (formula (8)) and path model II (formula (9)) respectively.

3.3 Coefficients Estimation and Goodness of Fit

Partial residuals estimation described in section 2.2 is used to solve model (6), (8) and (9), then the time-varying coefficients and the nonparametric parts are obtained. Firstly, statistical analysis is implemented to the coefficients on average. We compare the strength of the market, industry and area effect. There are six types: $M > I > A$, $M > A > I$, $I > M > A$, $I > A > M$, $A > M > I$, $A > I > M$. For example, $M > I > A$ means that market effect is larger than industry effect and industry effect is larger than area effect. The results are shown in Table 3. Obviously, $M > I > A$ and $I > M > A$ account for the most percent, 36% and 34% respectively. This result is different from Zhu et al. (2011) [23], in which their conclusion is that industry effect is stronger than market and area effects and industry information leads to the change of stock price.

Furthermore, we study the dynamic characteristics of the coefficients. (Take No.3 stock as example.) The coefficients of base model, path model I and path model II are displayed in the upper part of Figure 1. From Figure 1, the coefficients of base model fluctuate most violently, then path model II and path model I successively. Whether these coefficients have significant time-varying characteristics? We plot the histogram of the coefficients of path model I with gentlest fluctuation in the lower part of Figure 1. It is clear that the coefficients span widely and they are significantly time-varying. Other stocks have the similar conclusion about the coefficients.

Generally there are three indicators to measure

Table 2: Correlation matrix of explanatory variable

	$r_{M,t}$	$r^I_{j,t}$	$r^A_{j,t}$	$Tr_{j,t}$	$IL_{j,t}$	T_j	$M\_IL_{j,t}$	$I\_IL_{j,t}$	$A\_IL_{j,t}$
$r_{M,t}$	1.0000	0.0002	0.0004	0.1197	-0.0856	-0.0091	-0.0326	-0.0844	-0.0325
$r^I_{j,t}$	0.0002	1.0000	-0.0028	0.1438	-0.0515	-0.0682	-0.0226	-0.0685	-0.0301
$r^A_{j,t}$	0.0004	-0.0028	1.0000	-0.0466	-0.1038	-0.0376	-0.1786	-0.1369	-0.2070
$Tr_{j,t}$	0.1197	0.1438	-0.0466	1.0000	-0.4783	-0.4579	-0.3157	-0.4264	-0.3741
$IL_{j,t}$	-0.0856	-0.0515	-0.1038	-0.4783	1.0000	0.5034	0.5513	0.6843	0.5980
T_j	-0.0091	-0.0682	-0.0376	-0.4579	0.5034	1.0000	0.2998	0.4557	0.3975
$M\_IL_{j,t}$	-0.0326	-0.0226	-0.1786	-0.3157	0.5513	0.2998	1.0000	0.8014	0.8760
$I\_IL_{j,t}$	-0.0844	-0.0685	-0.1369	-0.4264	0.6843	0.4557	0.8014	1.0000	0.7802
$A\_IL_{j,t}$	-0.0325	-0.0301	-0.2070	-0.3741	0.5980	0.3975	0.8760	0.7802	1.0000

Note: Correlation matrixes of other individual stocks are similar to this table.

Table 3: Comparison of market, industry and area effect

Type	Amount	Percentage(%)	Stock number
$M > I > A$	18	36	4, 7, 8, 14, 16, 18, 29, 35, 36, 38, 39, 40, 43, 44, 45, 47, 48, 50
$M > A > I$	9	18	9, 12, 15, 20, 21, 30, 33, 37, 49
$I > M > A$	17	34	1, 3, 5, 11, 17, 19, 22, 23, 24, 25, 26, 27, 31, 32, 41, 42, 46
$I > A > M$	2	4	10, 28
$A > M > I$	2	4	2, 13
$A > I > M$	2	4	6, 34

the goodness of fit for modeling. They are R-square (R2), average relative error (ARE) and root-mean-square error (RMSE). R2 represents the proportion of the explanatory variable changes can be explained by the response variable. ARE and RMSE represent the deviation of fitting values and real values. After solving model (6), (8) and (9), we also get R2, ARE and RMSE of them, see Figure 2.

According to Figure 2, path model I incorporating liquidity information of individual stock capture the fluctuation of returns best. The average of R2 is 0.9951, ARE 0.0850 and RMSE 0.1401, which indicates that path model I fits the data well. As a whole, path model I and II are both better than base model, which reflects that investor sentiment represented by liquidity is indeed an important factor in asset pricing.

3.4 Pricing Error Analysis

In order to investigate the efficiency of index pricing, we perform pricing error analysis. According to the idea of path converged design, path model converges to base model in probability 1 when $t \to \infty$. We define path premium of liquidity about path model I and base model as:

$$LLP^1_{j,t} = \varepsilon^1_{j,t} - \varepsilon^2_{j,t}.$$

Similarly, path premium of liquidity about path model II and base model is defined as:

$$LLP^2_{j,t} = \varepsilon^1_{j,t} - \varepsilon^3_{j,t}.$$

Furthermore, we define index pricing error $LP_{j,t}$ as:

$$LP_{j,t} = LLP^1_{j,t} - LLP^2_{j,t}. \tag{17}$$

$LP_{j,t}$can be considered as a bubble of individual stock relative to the market environment because of liquidity. Based on this idea, we construct simulated trading strategy according to series $\{LP_{j,t}\}$. After observation, we find more than 80% of the previous day's closing price is between the lowest and highest prices on the next day, so we suppose that the previous day's closing price would appear on the next day, which means that it will make a bargain at bidding by the previous days' closing price on the next day. In our trading strategy, one unit asset is considered and we all bid by the previous day's closing price early on the next day. The trading strategy is as follows:

(I) If$LP_{j,t} \geq 0$, then we sell the stock or maintain the short position.

(II) If $LP_{j,t} < 0$ and we have no stock at that time, then we buy one unit stock.

(III) If $LP_{j,t} < 0$ and we have stock already at that time, then we continue to hold it without any dealing.

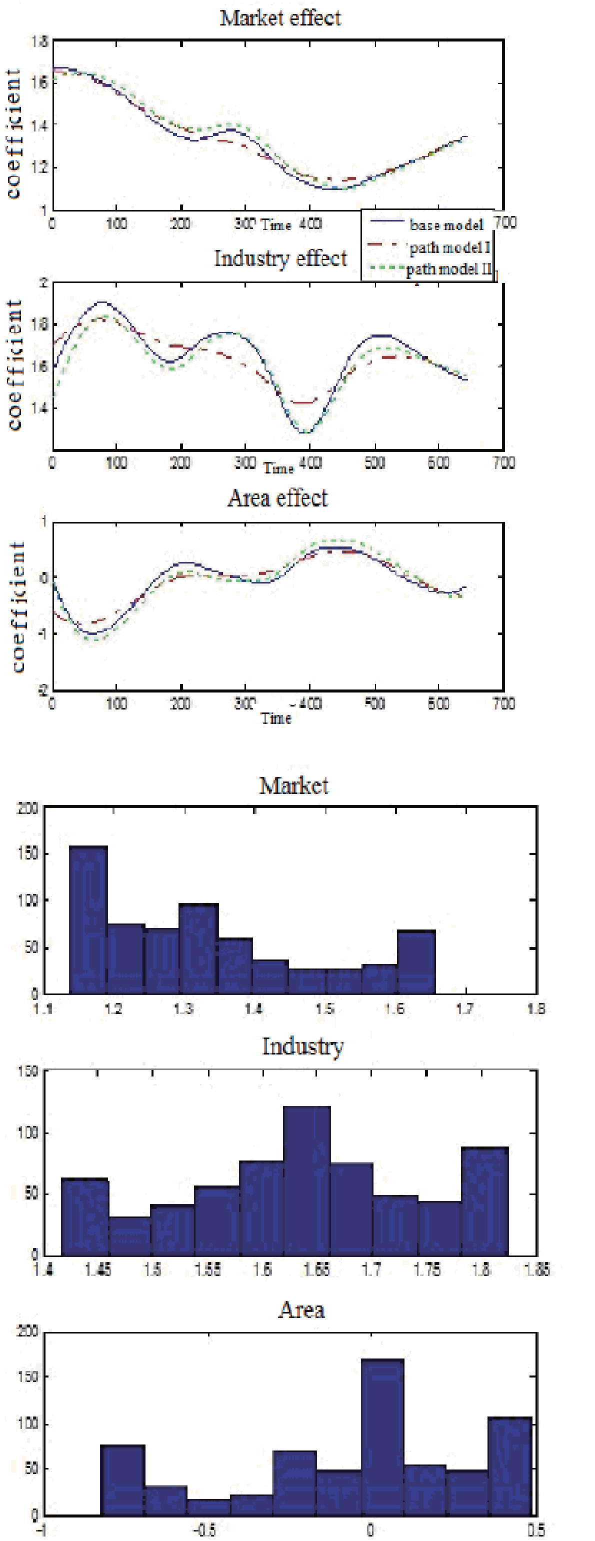

Figure 1: Coefficients and histogram of coefficients of No.3 stock

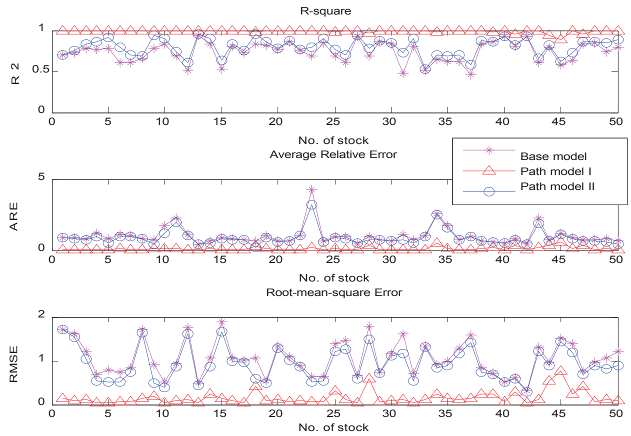

Figure 2: R2, ARE and RMSE of the three models

Denote the state variable to describe whether owing stock j at day t:

$$TS_{j,t} = \begin{cases} 1, & owing\ stock\ j\ at\ day\ t, \\ 0, & otherwise, \end{cases} \quad (18)$$

then

$$TS_{j,t+1} = \begin{cases} 1, & if\ LP_{j,t} < 0, \\ 0, & if\ LP_{j,t} \geq 0. \end{cases} \quad (19)$$

Suppose $TS_{j,1} = 0$, then the series $\{TS_{j,t}\}_{t=1:T}$ are obtained. Ignoring transaction costs, the capital accumulation at day $t+1$ is:

$$M_{j,t+1} = M_{j,t} \cdot (1 + TS_{j,t} \times R_{j,t}). \quad (20)$$

At the end of considering period, the total yield on our trading strategy is $r_j = M_{j,N_j} - 1$.

Positive index pricing error means investors excessive optimism at the stock, which leads to excessive risk premium. Bubbles will burst, prices will gradually return to rational, which is the basic idea of our trading strategy. "Reasonable prices" of stock are associated with market condition and can be represented by the returns and investor sentiments of market, industry and area indexes, which is our modeling basis. Selling or buying each stock under our trading strategy, the accumulative earnings at the end of sample period are obtained. At the same time, we compute the earnings of individual stock, market index, industry index and area index under buy-and-hold strategy respectively. The earnings are shown in Table 4.

From Table 4, we calculate the average earnings of the individual stocks. Under our trading strategy,

Table 4: Comparison of market, industry and area effect

No	Trade Strategy	Individual	Market	Industry	Area	No	Trade Strategy	Individual	Market	Industry	Area
1	2.68	50.00	-15.06	-29.72	18.84	26	46.63	-13.09	-15.00	-16.34	-13.10
2	54.79	95.57	-15.06	-9.84	18.84	27	26.15	5.83	-15.06	-12.27	-20.17
3	105.37	53.25	-15.06	-5.34	-9.08	28	44.05	-8.69	-15.06	-13.22	-11.26
4	48.23	-15.74	-15.06	-29.72	-20.17	29	20.97	23.13	-15.06	-5.62	-3.93
5	1.22	-21.59	-15.06	3.73	-11.26	30	-32.97	-17.78	-15.06	-16.64	-14.69
6	5.51	-5.66	-15.06	-6.26	-11.26	31	32.46	81.16	-15.06	-7.87	19.49
7	36.34	-7.94	-15.06	-37.86	-14.69	32	24.66	-6.46	-15.06	-5.62	-9.08
8	27.53	16.75	-13.77	2.84	-26.75	33	2.27	-15.05	-15.06	-13.96	-11.26
9	-28.89	-58.44	-19.18	-18.18	-27.60	34	-0.49	-6.01	-15.06	-10.14	-11.26
10	34.84	7.39	-15.06	-5.62	-11.26	35	14.52	-33.89	-15.06	-29.51	-20.17
11	31.94	-12.10	-17.18	-5.66	-12.32	36	-45.65	-58.82	-15.06	-14.78	-11.26
12	37.87	65.95	-15.29	-24.45	-42.93	37	0.99	-19.25	-15.06	-13.96	-11.26
13	16.23	32.05	-15.06	-5.71	9.36	38	34.59	-13.80	-16.17	-15.55	-12.64
14	156.76	49.92	-15.06	-10.20	-3.82	39	7.49	-24.74	-15.06	-18.54	-11.26
15	-42.86	-51.97	-20.26	-33.14	-26.04	40	44.12	-0.83	-15.06	-16.64	-11.26
16	53.66	4.37	-15.06	-18.54	-20.17	41	16.17	-10.19	-15.06	-18.24	-11.26
17	-20.11	-10.86	-15.06	-5.62	-11.26	42	1.27	-11.10	-15.06	-18.24	-11.26
18	9.23	-19.78	-15.06	-14.78	-20.99	43	-2.82	-29.25	-6.37	3.21	-2.21
19	-9.11	-16.46	-15.06	-5.62	-20.17	44	36.64	-13.77	-15.06	-18.54	-20.17
20	4.22	-16.27	-15.06	-14.78	-26.65	45	49.16	3.70	-15.06	-13.96	-11.26
21	57.34	-0.70	-15.06	-16.64	-14.69	46	12.04	-12.46	-15.06	-14.78	-11.26
22	67.16	95.69	-15.06	-5.62	-11.26	47	-22.89	-25.16	-15.06	-16.64	-11.26
23	19.53	7.11	-15.06	-5.62	-11.26	48	-1.01	1.02	-15.06	-18.54	-9.08
24	32.23	-7.53	-15.06	-5.62	-20.17	49	-12.90	-11.79	-15.06	-14.78	-3.93
25	33.12	49.78	-15.06	-13.96	-11.26	50	30.21	1.32	-15.06	-16.64	-12.18

it is 21.21%. Under the buy-and-hold strategy, it is 1.34%. Obviously, profit of our trading strategy is much larger than buy-and-hold strategy. Furthermore, as to the 50 stocks, under our trading strategy, there are 38 stocks earn more than under their own buy-and-hold strategy, 44 outperform market index, 41 outperform industry index, 41 outperform area index and 31 stocks outperform all other situations, at 62%. The sample stocks of Shanghai Index 50 are selected from the stocks with good profitability and our trading strategy is constructed based on the comparison with the path model I and the path model II, so we can draw the conclusion that index pricing model - path model II provides important reference to ”rational” price of stock, the idea of index pricing is practical. Quan et al. (2012) [31] indicates that investor attention is a scarce resource, investors often show the state of selective attention for the information in the market, the selective attention state will eventually has a significant impact on stock price and its dynamic change. Because stock price indexes can reflect the overall change direction and degree of study object and continuous index sequences can also reflect the trend of development, investors will pay more attention to the information of indexes. Especially, the indexes will earn more attention of institutional investors. The institutional investors have advantages of information and technology, so that their trading will be more professional and their pricing and expectation will be more rational. Thus, the price and fluctuation of indexes provide guidance to individual investors.

4 Conclusion

This paper has put forward an index pricing model embedding investor sentiment for individual stocks. Path converged design are employed to construct the model under a nonparametric frame. The sample stocks of Shanghai Index 50 are selected as the study object. After empirical study, we can draw the conclusions for China’s stock market:

(1) The idea of index pricing is practical, some indexes may provide important reference to ”rational” price of individual stocks.

(2) Investor sentiment is an important factor for asset pricing.

(3) Market, industry and area indexes influence the returns of individual stocks significantly, with the

characteristic of time-varying. Influences of market and industry are stronger than of area.

However, the proposed index pricing method is not applicable to every stock and our trading strategy can't always win, which may be due to the partialness for liquidity as the proxy of investor sentiment. In addition, the trading strategy based on path premium of liquidity is too simplified. In fact, the bubble will not immediately burst, the transaction may be in haste under our strategy. Suitable conditions of index pricing model need further research.

Acknowledgements: This work is supported by the Major Project of Zhejiang Provincial Science Foundation of China(No.Z6110519), the Plan of National Statistical Science Project(Grant No: 2013LZ15), the Zhejiang Provincial Key Research Base for Humanities and Social Science Research (Statistics 1020XJ3314010 & Applied Economics 1020XJ3314018G) and Research Project of Shaoxing University in 2013.

References:

[1] Y. Shih, S. Chen, et al., The Evolution of Capital Asset Pricing Models, *Rev Quant Finan Acc.* doi: 10.1007/s11156-013-0348-x, 2013.

[2] E.F. Fama and K.R. French, Size and Book-to-market Factors in Earnings and Returns, *Journal of Finance.* 50, 1995, pp. 131-155.

[3] D. Kent and S. Titman, Evidence on the Characteristics of Cross-sectional Variation in Common Stock Returns, *Journal of Finance.* 52, 1997, pp. 1-33.

[4] S. Wu and N. Xu, A Comparative Study on the Rational Asset Pricing Model and Irrational Asset Pricing Model: Evidence from Stock Market in China, *Economic Research Journal.* 6, 2006, pp. 105-116.

[5] B. Nicholas, A. Shleifer, and R. Vishny, A Model of Investor Sentiment, *Journal of Financial Economics.* 3, 1998, pp. 307-343.

[6] S. Wang, F. Zhao and D. Wang, The Effect of Stock Market Mispricing on Investment-Evidence from China, *WSEAS Transactions on Systems.* 12, 2013, pp. 593-603.

[7] B.W. Gregory and M.T. Cliff, Investor Sentiment and Asset Valuation, *Journal of Business.* 2, 2005, pp. 405-440.

[8] C. Wang and W. Zhang, Research Review on the Impact of Investor Sentiment on the Stock Market, *Financial Theory & Practice.* 8, 2012, pp. 98-102.

[9] E.F. Fama and K.R. French, The Cross-section of Expected Stock Returns, *Journal of Finance.* 47, 1992, pp. 427-465.

[10] A.S. Kyle, Continuous Auctions and Insider Trading, *Econometrica.* 53, 1985, pp. 1315-1336.

[11] F. Black, Noise, *Journal of Finance.* 41, 1986, pp. 529-543.

[12] B. Malcolm, J. Wurgler, Investor Sentiment and the Cross-section of Stock Returns, *Journal of Finance.* 61, 2006, pp. 1645-1680.

[13] S. Basak, Asset Pricing with Heterogeneous Beliefs, *Journal of Bank & Finance.* 11, 2005, pp. 2849-2881.

[14] H. Levy, M. Levy and G. Benita, Capital Asset Prices with Heterogeneous Beliefs, *Journal of Business.* 79, 2006, pp. 1317-1353.

[15] K. Yoel, Asset Pricing with Status Risk, *Working paper,* 2009.

[16] L. Pastor and R. Stambaugh, Liquidity Risk and Expected Stock Returns, *Journal of Political Economics.* 3, 2003, pp. 642-685.

[17] V.V. Acharya and L.H. Pedersen, Asset Pricing with Liquidity Risk, *Journal of Financial Economics.* 77, 2005, pp. 375-410.

[18] F. Zhou and W. Zhang, Study of Liquidity Risk Premium in China's Stock Market, *Journal of Financial Research.* 5, 2011, pp. 194-206.

[19] P. Erdos, M. Ormos and D. Zibriczky, Non-parametric and Semi-parametric Asset Pricing, *Economic Modeling.* 28, 2011, pp. 1150-1162.

[20] E. Ferreira, J. Gil-Bazo and S. Orbe, Conditional Beta Pricing Models: A Nonparametric Approach, *Journal of Banking & Finance.* doi: 10.1016/j.jbankfin.2011.05.016, 2011.

[21] J. Hua, Nonparametric Estimation of Beta Coefficient in CAPM Model, *Contemporary Economics.* 9, 2010, pp. 126-128.

[22] J. Li, H. Shen and Q. Jin, An Empirical Study of Comovement of Locations and Stock Returns, *China Industrial Economics.* 2, 2009, pp. 109-119.

[23] H. Zhu, L. Chen, and N. Pan, Industry, Local and Market Information, Who Dominates Price Movement in Chinese Stock Market? *Chinese Journal of Management Science.* 4, 2011, pp. 1-8.

[24] A. Bougioukou, Modeling Cyprus Stock Market, *WSEAS Transactions on Mathematics.* 12, 2012, pp. 1076-1084.

[25] P. Ceretta, M. Righi, A. Costa, et al., Time-varying Behavior of Long-run among Stock Markets in Western Europe, *WSEAS Transactions on Business and Economics.* 37, 2014, pp. 398-404.

[26] J. Shao and J. Wang, Analysis of Chain Reaction Between Two Stock Indices Fluctuations by Statistical Physics Systems, *WSEAS Transactions on Mathematics.* 11, 2010, pp. 830-839.

[27] B. Xu, Path Converged Design Application to Production Efficiency of FDI in Regions, *Economic Research Journal.* 2, 2010, pp. 44-54.

[28] Z. Li and A. Ye, *Advanced Econometrics,* Tsinghua University Press, Beijing, pp. 116-117, 2000.

[29] A. Ye, R.J. Hyndman and Z. Li, *Local Linear Multiple Regression With Variable Bandwidth in the Presence of Heteroscedasticity,* http://www.wise.xmu.edu.cn/seta2006/download /ORAL%20PRESENTATION/Ye%20Azhong /Heteroscedasticitynewn.pdf.

[30] A. Ye, *Nonparametric Econometrics,* Nankai University Press, Tianjin, 2003.

[31] X. Quan, T. Hong and S. Wu, Effect of Selective Attention, Ostrich and Market Anomalies, *Journal of Financial Research.* 3, 2012, pp. 109-123.