

Outlier Impacts in Stock Pricing

Jianlong Wang

School of Mathematical Sciences, Faculty of Science and Technology

Universiti Kebangsaan Malaysia

43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Tel: 0-112-880-961 E-mail: wangjianlong7016@gmail.com

Saiful Hafizah Jaaman (Corresponding author)

School of Mathematical Sciences, Faculty of Science and Technology

Universiti Kebangsaan Malaysia

43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Tel: 60-389-213-422 E-mail: shj@ukm.edu.my

Humaida Banu Samsudin

School of Mathematical Sciences, Faculty of Science and Technology

Universiti Kebangsaan Malaysia

43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Tel: 60-389-215-723 E-mail: humaida@ukm.edu.my

Received: September 4, 2014 Accepted: November 27, 2014 Published: February 1, 2015

doi:10.5296/jmr.v7i2.7050 URL: <http://dx.doi.org/10.5296/jmr.v7i2.7050>

Abstract

Outlier is the observation that is not consistent with the rest of observations. It exists not only in stock prices but also in the economic variables. In multifactor asset pricing model, the ordinary least square method (OLS) is commonly used to estimate coefficients. The existence of outliers can lead to inadequate results under the OLS framework. Huber's robust method (HRM) can be used to avoid the bad impacts of outliers and the abnormal problems. Applying both methods to Shanghai stock market, the outlier observations are analyzed to

examine its influence on the results and parameters estimation. The result of this study found that HRM outperforms OLS.

Keywords: Outliers, HRM, OLS, Shanghai stock market

1. Introduction

As it is known, OLS method is to minimize the square of residuals, in which the weight to each factor is the same. If there exist outliers, the regression results will be influenced directly. One approach is to use Least-Trimmed Squares methodology, Knez and Ready (1997) find the Fama and French (1992) size-based risk premium disappears. Hill (2013) develop a robust least squares estimator with heavy tailed errors, and find tail trimming ensure the robustness to heavy tails in both small and large samples. Robust regression model is an alternative to OLS when outliers exist.

Most often the empirical data of a security or portfolio return is not normally distributed but flat tailed. P. Theodossiou (1998) and Bali and Weinbaum (2007) reject the normality assumption for returns as well as some macro-economic factors. McDonald, Michelfelder, and Theodossiou (2009) claim that OLS is not the best estimation method of betas and may lead to erroneous estimates. This is even severe in terms of the outlier and non-normality distributed variables. Alma (2011) indicates the strong adverse effect of outlier observations are not easily noticed and result in bad impacts. By comparing OLS and robust regression method, he concludes that the robust regression method can help detect outliers and provide resistant results in the presence of outliers.

A multitude of studies on how the macroeconomic factors affect the security market are done. (see Yi and Sun (2014), Joslin, Le, and Singleton (2013), Ng and Weight (2013), Gay Jr (2011)) Many of them (see Tian (2006), , Alma (2011)) compare the OLS and robust regression. Considering evidence by Bruner, Eades, Harris, and Higgins (1998) and Graham and Harvey (2001) that most of the companies rely on CAPM to forecast their cost of capital, ignoring the outliers may cause a large-scale misallocation of resources and plus of social cost.

In order to find out the impacts of outliers, the research is done by Martin and Simin (2003), Gray et al. (2005), and Genton and Ronchetti (2008). In an extreme case, even a single outlier may significantly influence the CAPM regression estimations. As a result, misestimated beta will lead to a misestimated cost of capital and an erroneous capital budgeting decisions.

A. Theodossiou, Theodossiou, and Yaari (2009) demonstrate that the outlier returns are not priced in large market portfolios. Thus, the estimated coefficients (alpha and beta) of outlier components should not be included in the systematic risk. Through the comparison of OLS and HRM, Theodossiou and Theodossiou (2014) find the outlier effect exist in empirical study.

In this study, the objective is to ascertain the different results basing on OLS and HRM because of the existence of outlier impacts.

2. Methodology

2.1 Sample data

In this study, monthly equity prices and microeconomic data are collected from Sina Finance. National macroeconomic factors are collected from National Bureau of Statistics of China

(NBSC). The global stock indexes, gold price and crude oil price are collected from World Bank. There are 300 firms 34 factors during period year 2000 to 2011 included in this study. (refer Appendix 1)

In multiple factor models

$$R_t = \alpha + F_t' \beta + \varepsilon \quad (1)$$

The dependent variables are the firms' excess stock returns ($R_t = r_t - r_{ft}$), where r_{ft} is the risk free rate, using 3-month interest rate collected from China Bank,

$$r_t = \log(p_t / p_{t-1}) \quad (2)$$

where p_t is the stock price at time t . The independent variables are factors defined as,

$$F_t = \log(f_{it} / f_{it-1}) \quad (3)$$

where f_{it} is the collected present value of factor i at time t .

2.2 Robust regression

Different with OLS (Equation 4), HRM is similar to a form of weighted and reweighted least squares regression.

$$\text{Minimize: } \sum_{i=1}^n (R_i - F_i' \beta)^2 \quad (4)$$

As shown in the Equation (5), the robust regression weighs the observations differently based on the observations' behavior. The widely used method of robust regression is M-estimation (HRM), proposed by P. J. Huber (1964). Consider the model in Equation (1), the general HRM is to minimize the objective function presented in Equation (5).

$$\text{Minimize: } \sum_{i=1}^n \rho(R_i - F_i' \beta) \quad (5)$$

Where $\rho(\cdot)$ satisfy:

1) $\rho(\cdot) \geq 0$; 2) $\rho(0) = 0$; 3) $\rho(-e) = \rho(e)$; 4) For any $|e_i| > |e_j|$, $\rho(e_i) > \rho(e_j)$

Let $\varphi = \rho'$ be the derivative of ρ . Thus

$$\sum_{i=1}^n \varphi(R_i - F_i' \beta) F_i' = 0 \quad (6)$$

Let $w(e) = \varphi(e) / e$ and $w_i = w(e_i)$. The estimation Equation (6) is to minimize $\sum w_i^2 e_i^2$, which is the weighted least squares problem. The weights, however, depend on the residuals, which depend on the estimated coefficients, with the estimated coefficients depend upon the weights. An iterative solution (called iteratively reweighted least-squares, IRLS) is sufficient in solving this problem stated below:

- 1) Select initial estimates $b^{(0)}$, such as the least-squares estimates.
- 2) At each iteration t , calculate residuals $e_i^{(t-1)}$ and associated weights $w_i^{(t-1)} = w(e_i^{(t-1)})$ from the previous iteration.
- 3) Solve for new weighted-least-squares estimates

$$b^{(t)} = [FX'W^{(t-1)}F]^{-1} F'W^{(t-1)}y \quad (7)$$

Here, F is the model matrix, with F_i as its i th row, and $W^{(t-1)} = \text{diag}\{w_i^{(t-1)}\}$ is the current weight matrix. Steps 2 and 3 are repeated until the estimated coefficients converge. The asymptotic covariance matrix of b is

$$v(b) = \frac{E(\varphi^2)}{[E(\varphi')]^2} X'X^{-1} \quad (8)$$

Using $\sum [\varphi(e_i)]^2$ to estimate $E(\varphi^2)$ and $[\sum \varphi'(e_i) / n]^2$ to estimate $[E(\varphi')]^2$ produces the estimated asymptotic covariance matrix, $\hat{v}(b)$ (which is not reliable in small samples).

Table 1. Methods Weight Comparison

Objective Function	
OLS	$\rho_{LS}(e) = e^2$
Huber	$\rho_H(e) = \begin{cases} \frac{1}{2}e^2 & \text{for } e \leq k \\ k e - \frac{1}{2}k^2 & \text{for } e > k \end{cases}$
Weight Function	
OLS	$w_{LS}(e) = 1$
Huber	$w_H(e) = \begin{cases} 1 & \text{for } e \leq k \\ \frac{k}{ e } & \text{for } e > k \end{cases}$

The objective functions and corresponding weight functions for OLS and HRM are presented in Table 1. Both the least-squares and the Huber objective functions increase without bound as residual e departs from 0. The least squares objective function increases more rapidly. Least-squares, as known, gives equal weight to each observation; the weights for Huber method decrease for $|e| > k$. The value k for the HRM is called a tuning constant; smaller values of k produce more resistance to outliers (reduce the impact from outliers), but at the expense of lower efficiency. The tuning constant is generally picked to give reasonably high efficiency in the normal case; particularly, $k = 1.345\sigma$ for the HRM (σ is the standard deviation of the errors) produce 95 percent efficiency and still offer protection against outliers. Common approach in estimating the standard deviation of residuals takes $\hat{\sigma} = MAR / 0.6745$. (MAR is the median absolute residual.) The constant value 0.6745 makes the estimate unbiased for normal distribution. (refer P. Huber (1981))

2.3 Comparison between OLS and HRM

HRM is mathematically related to the OLS method. In that case, comparison between the OLS and robust regression enables to check the impact of outliers on beta estimation. Following Roll (1988), Theodossiou and Theodossiou (2014) build a mixed return process (MRP) model in describing stock returns with regular and outlier observations. This process fit for the distributions with outliers. Even the regular components can be assumed as normally distributed, given an appropriate parameterization of the outlier component, the

return as a whole can distributed anomalous. Thus, the MRP is consistent with the real stock returns.

The MRP can be expressed as:

$$R_{it} = K_{it} + d_{it}H_{it} \quad (9)$$

where R_{it} is the stock returns for firm i during period t . Regular component (K_{it}) and outlier component ($d_{it}H_{it}$) constitute the stock returns (R_{it}), which is made up by linear function of market return (R_{Mt}). d_{it} is a Bernoulli random variable shown in:

$$p = \begin{cases} q, & d_{it} = 1 \\ 1 - q, & d_{it} = 0; \text{outlier} \end{cases} \quad (10)$$

The regular and outlier components can be written as:

$$K_{it} = Z_t \theta_{kt} + e_{it} \quad (11)$$

and

$$d_{it}H_{it} = Z_t \theta_{ht} + v_{it} \quad (12)$$

where $Z_t = [1 \ R_{Mt}]$ is a row vector of the number one, related to the intercept, and the market return. $\theta_{kt} = [\alpha_{kt} \ \beta_{kt}]'$ is a column vector of the alpha (intercept) and beta (slope) of the regular return equation. $\theta_{ht} = [\alpha_{ht} \ \beta_{ht}]'$ is a column vector of alpha and beta of the outlier return equation. e_{it} and v_{it} are white noise error terms to the regular and outlier equations respectively.

$R_t = [R_{it}]$ is a $T \times 1$ vector of trimmed returns of stock i and $Q_Z = (Z'Z)^{-1}Z'$ is the OLS projection operator. Thus, $R_t = R_t - H_t$

$$\hat{\theta}_{k,t} = Q_Z R_t = Q_Z (R_t - H_t) = Q_Z R_t - Q_Z H_t = \hat{\theta}_{OLS,t} - \hat{\theta}_{r,t} \quad (13)$$

Where $\hat{\theta}_{OLS,t}$ are the OLS estimators for the alpha and beta, and $\hat{\theta}_{r,t}$ are the OLS estimators for the intercept and slope of the outlier component of stock returns for stock i .

It follows from equation (13), that

$$\hat{\theta}_{OLS,t} = \hat{\theta}_{k,t} + \hat{\theta}_{r,t} \quad (14)$$

That implies that the OLS estimators of the stock return intercept (alpha) and slope (beta) are constituted by their relative HRM and outlier regression estimators. Theodossiou and Theodossiou (2008) prove that outlier returns are not priced in large market portfolios. This indicates that the alpha and beta of the outlier component of stock returns do not contribute to the systematic risk. From this perspective, OLS estimated intercept and slope are biased measures of stock's alpha and beta. Therefore, the OLS biases for alpha and beta are

$$OLS \text{ bias} = \hat{\theta}_{OLS,t} - \hat{\theta}_{k,t} = \hat{\theta}_{r,t} \quad (15)$$

Which are captured by the corresponding intercept and slope of the outlier regression model. The details of the derivation of the model are provided by Theodossiou and Theodossiou (2014).

3. Empirical Results

3.1 Outliers Description

Since the different result of the OLS and HRM, further study and comparison are done in this section. Different from the APT model, model in this part is the individual stock company return that is applied as the dependent variable and market return is the independent variable. This enables to know whether the outlier effect is company specific or not.

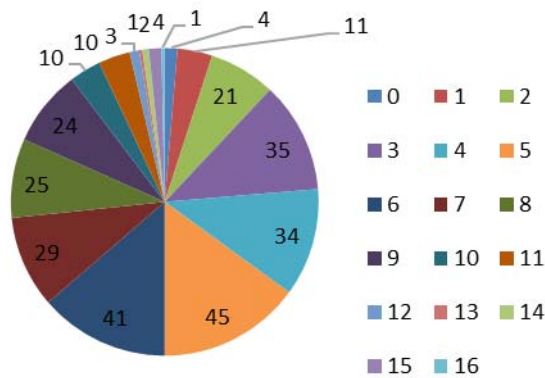


Figure 1. Outlier Description

The outliers are verified existing in the stock return variables and in the factor variables as well. The number of company in each number of outlier is plotted in Figure 1. Different colors indicate the number of outliers.

As can be seen, there are one company with the most of outlier (16) and 4 companies without outlier. The most frequency of outlier are 5 with 45 companies. Further, the remarkable increasing from outlier of counting 0 to 5 and decreasing trend from outlier of counting 5 to 16. The weight average of outlier is 0.58. These indicate that each monthly company return variable contains around 5 to 6 extreme values from 2001 to 2011. Figure 2 gives the boxplot of outliers for all 34 factors. The outliers are also widely existing in economic factor variables. The average number of missing values are 7. There are only two factors (import and national revenue) without outlier.

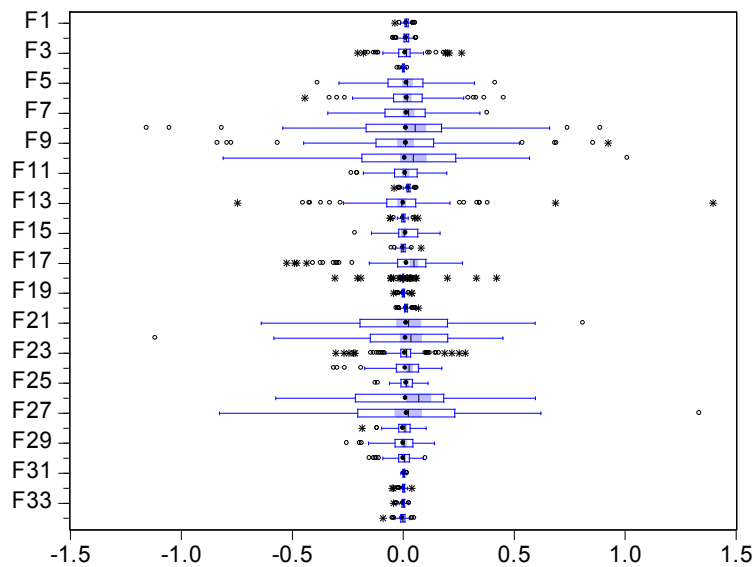


Figure 2. Outliers of Factors

3.2 Comparison and Outlier Effects

Table 2 offers the estimated slope values from OLS and HRM estimation method as well as the standard errors for both methods.

The second column is the number of outlier of each company. Outliers do exist in almost all companies (except four companies). The existence of outliers indicates the influence of policy and unexpected company-specific events. There is an increasing trend of the differences of both standard errors with the increasing of the number of outlier (Figure 3). This tells that HRM is able to mitigate with the increasing of outlier numbers. The dramatically decreasing after the outlier number of 14 indicate this mitigation is limited to a restricted range and, in addition, might be affected by the outliers in regressors as well.

Table 2. Descriptive to Comparison

Code	Outlier	SE OLS	SE HRM	α_{ols}	α_{hrm}	$ \alpha_{ols} - \alpha_{hrm} $
Mean	5.82	0.0488	0.0292	-0.0045	-0.0044	0.0054
SD	3.01	0.0121	0.0068	0.0143	0.0111	0.006
Min	0	0.0275	0.0157	-0.0448	-0.0352	0.0001
10%	2	0.0381	0.0219	-0.02	-0.0186	0.0007
25%	4	0.0428	0.0251	-0.0137	-0.0122	0.0017
50%	5.5	0.0478	0.0287	-0.0063	-0.004	0.0037
75%	8	0.0537	0.0323	0.0043	0.0027	0.0071
90%	10	0.0585	0.0365	0.0134	0.009	0.0111
Max	16	0.1709	0.07	0.0909	0.028	0.0629

The third and fourth column present the standard error for both models. OLS method takes greater mean standard error than that of HRM. Additionally, all values from the minimum to the maximum, OLS standard error are greater than HRM. This shows that HRM estimation method outperforms OLS method.

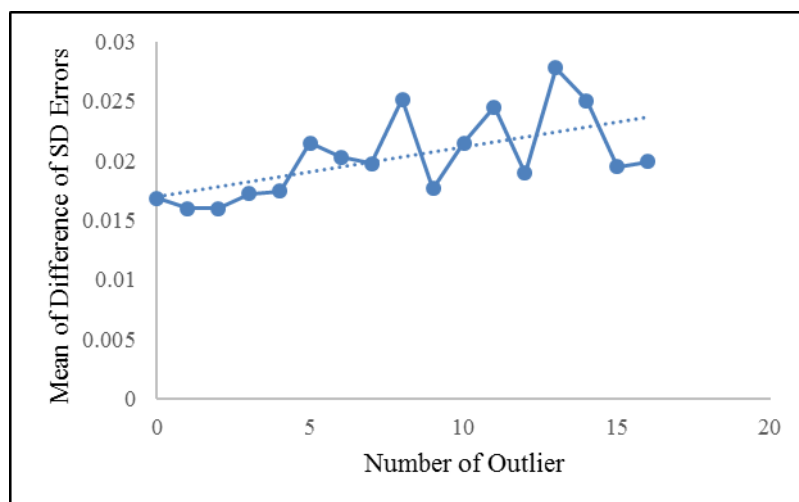


Figure 3. Difference of Standard Error

The last three column shows the estimated alpha from both estimation methods and absolute values of their differences. The minimum, median and maximum of the absolute values are 0.0001, 0.0054 and 0.0629. Approximately, 25% of companies take the difference value greater than 0.0071 and 10% take the difference value greater than 0.0111.

4. Conclusion

In order to find out whether the outlier effect exists, the comparison between OLS and HRM is based on the regression with individual company as dependent variable and the microeconomic and macroeconomic factors as independent variables. Results asserts that the OLS estimation method provides higher standard error than HRM. Hence, HRM helps to estimate coefficients with less error. There is a significant difference between the estimated OLS and HRM alphas. As a whole, HRM performs better in estimating coefficients when outliers exist.

Acknowledgement

The authors would like to acknowledge the research grants: FRGS/1/2013/SG/04/UKM/02/6 and GGPM-2012-007.

References

Alma, Ö. G. (2011). Comparison of Robust Regression Methods in Linear Regression. *International Journal of Contemporary Mathematical Sciences*, 6(9), 409-421.

Bali, T. G., & Weinbaum, D. (2007). A conditional extreme value volatility estimator based on high-frequency returns. *Journal of Economic Dynamics and Control*, 31(2), 361-397. <http://dx.doi.org/10.1016/j.jedc.2005.10.002>

Bruner, R. F., Eades, K. M., Harris, R. S., & Higgins, R. C. (1998). Best practices in estimating the cost of capital: survey and synthesis. *Financial Practice and Education*, 8, 13-28.

- Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), 427-465. <http://dx.doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Gay Jr, R. D. (2011). Effect of macroeconomic variables on stock market returns for four emerging economies: Brazil, Russia, India, and China. *International Business & Economics Research Journal (IBER)*, 7(3).
- Genton, M. G., & Ronchetti, E. (2008). *Robust prediction of beta Computational Methods in Financial Engineering* (pp. 147-161), Springer. http://dx.doi.org/10.1007/978-3-540-77958-2_8
- Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics*, 60(2), 187-243. [http://dx.doi.org/10.1016/S0304-405X\(01\)00044-7](http://dx.doi.org/10.1016/S0304-405X(01)00044-7)
- Gray, S., Hall, J., Bowman, J., Brailsford, T., Faff, R., & Officer, R. (2005). The performance of alternative techniques for estimating equity betas of Australian firms. Report prepared for the Energy Networks Association (May).
- Hill, J. B. (2013). Least tail-trimmed squares for infinite variance autoregressions. *Journal of Time Series Analysis*, 34(2), 168-186. <http://dx.doi.org/10.1111/jtsa.12005>
- Huber, P. (1981). *Robust Statistics*: John Wiley, New York. <http://dx.doi.org/10.1002/0471725250>
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73-101. <http://dx.doi.org/10.1214/aoms/1177703732>
- Joslin, S., Le, A., & Singleton, K. J. (2013). JFEC Invited Paper: Gaussian Macro-Finance Term Structure Models with Lags. *Journal of Financial Econometrics*, 11(4), 581-609. <http://dx.doi.org/10.1093/jjfinec/nbt012>
- Knez, P. J., & Ready, M. J. (1997). On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions. *The Journal of Finance*, 52(4), 1355-1382. <http://dx.doi.org/10.2307/2329439>
- Martin, R. D., & Simin, T. T. (2003). Outlier-resistant estimates of beta. *Financial Analysts Journal*, 56-69. <http://dx.doi.org/10.2469/faj.v59.n5.2564>
- McDonald, J. B., Michelfelder, R. A., & Theodossiou, P. (2009). Evaluation of Robust Regression Estimation Methods and Intercept Bias: A Capital Asset Pricing Model Application. *Multinational Finance Journal*, 12, 293-321.
- Ng, S., & Wright, J. H. (2013). *Facts and challenges from the great recession for forecasting and macroeconomic modeling*: National Bureau of Economic Research. <http://dx.doi.org/10.3386/w19469>
- Roll, R. (1988). R2. *The Journal of Finance*, 43(3), 541-566. <http://dx.doi.org/10.2307/2328183>

Theodossiou, A., Theodossiou, P., & Yaari, U. (2009). *Beta estimation with stock return outliers: The case of US pharmaceutical companies*. Available at SSRN 1410371. <http://dx.doi.org/10.2139/ssrn.1410371>

Theodossiou, A. K., & Theodossiou, P. (2014). Stock return outliers and beta estimation: The case of US pharmaceutical companies. *Journal of International Financial Markets, Institutions and Money*, 30, 153-171. <http://dx.doi.org/10.1016/j.intfin.2014.02.002>

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management science*, 44(12-part-1), 1650-1661.

Theodossiou, P., & Theodossiou, A. (2008). Impact of Outliers on Stock Return Models: Implications for Event Studies and the Pricing of Risk: Working Paper, Rutgers University and Cyprus University of Technology.

Tian, D.-w. (2006). BIRR Model and the Use of a Macroeconomic Factor Model in China's Stock Market. *Journal of Xi'an Institute of Finance and Economics*, 19(5), 40-45.

Yi, R., & Sun, Z. (2014). Comparative Study on the Functional Forms of Stock Pricing. *Journal of Jishou University (Social Science Edition)*, 35(1).

Glossary

HRM: Huber robust regression method

OLS: Ordinary least square

CAPM: Capital asset pricing model

Appendix

Appendix 1. Factor Description

Symbol	Factors	Symbol	Factors
National Macroeconomic Factors			
R	Index returns	F13	Industrial growth
F1	Money and Quasi money (M2) Supply	F14	Producer purchase index (PPI)
F2	Money (M1) supply	F15	Total retail sales of consumer goods
F3	Supply of Currency (M0) in circulation	F16	Special drawing right to RMB
F4	Consumer price index (CPI)	F17	Government deposit in central bank

F5	Total import and export value	F18	Creditor's rights to financial firms
F6	Export value	F19	Ex-factory price of industrial product
F7	Import value	F20	Savings deposit
F8	Completed Investment in Fixed Assets	F21	Original insurance income
F9	Investment in Real Estate Development	F22	Original insurance expenses
F10	Actually used foreign direct investment	F23	Total freight
F11	Shanghai mutual fund Index	F26	National revenue
F12	Foreign exchange reserve	F27	National fiscal expenditure (excluding debt)
Microeconomic Factors			
F31	Book-to-market value	F33	Liability
F32	Size	F34	Cash flow
Global Macroeconomic Factors			
F24	Global crude oil price	F29	Stock index Nasdaq
F25	World gold price	F30	Stock index D&J
F28	Stock index S&P		

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).