Panel data model for estimating capital cost and empirical test for Japanese stock market

Koji INUI[1]

May 29th, 2017

Abstract

In this study, we explore a capital cost model proposed Lyle and Wang (2015) regarding as a return prediction model, and propose modified procedure for the sake of consistency to the original model assumptions based on panel data analysis. As our results, *one-dimensional random-effect model* is selected through statistical tests, and we confirm that the predictive power on the expected stock returns in Japanese market has improved as compared to the original Lyle-Wang model. Moreover, according to the empirical results of Japanese stock market, we find that the capital costs of high-contribution group among Nikkei 225 index universe shows lower than the others especially during the Bank of Japan purchased stock index ETF in large quantities.

[1] Corresponding author. School of Interdisciplinary Mathematical Science, Meiji University. Email: inui@meiji.ac.jp. This work was supported by JSPS KAKENHI Grant Number 16K03755.

1. Background

While Japanese economy and industry probably have high capacity for innovation, they have been left in downturn for a long time of those decades. To create long-term sustainable corporate value moving away from short-term oriented decision, "Ito report" (Ministry of Economy, Trade and Industry (2014)) describes "high-quality discussion" is required for both of investors and companies. In order to achieve the objectives by "high-quality discussion", reliable prediction methods of expected earnings and cost of capital, which are consistent the general requirements of the capital market should be required, however, they are not easily achieved.

Well-known financial models such as CAPM, Fama-French model have been applied to determine the capital cost of individual companies so far, but there are some critics such as the accuracy of the estimates by the models are not sufficient and also they are not suitable as a basis for the discussion about the future as regards they are estimated by using past data. On the other hand, implied returns estimated by Residual Income Valuation model (hereafter RIV mode), firstly presented by Ohlson (1995), represent investors' consensus that must be reflected in the stock prices observed simultaneously. Since the accounting data is also hired for RIV model, it meets superior in suitability in the context of forward looking cost of capital. There are many researched on the RIV model for estimating capital costs, in recent years, Easton(2004) presented new estimate method to separate the grow rate of earnings, Lyle and Wang (2015) assumed the cost of capital to fluctuate around their sustainable mean level and derived simple regression model.

In this study, we focus on Lyle-Wang model and propose improved method to estimate expected returns and capital costs referring the original paper Lyle and Wang (2015) which presented the case study on U.S. stock market and Ono (2015) which correspond to Japanese market.

2. Lyle – Wang model

In this section, we review the Lyle-Wang model.

At first we present total stock return and return on equity (ROE) of i-th firm, supposing clean-surplus condition has satisfied ($B_{i,t+1} = B_{i,t} + X_{i,t+1} - D_{i,t+1}$) as follows.

$$R_{i,t+1} = \frac{S_{i,t+1} + D_{i,t+1}}{S_{i,t}} \tag{1}$$

$$ROE_{i,t+1} = \frac{B_{i,t} + X_{i,t+1}}{B_{i,t}} = \frac{B_{i,t+1} + D_{i,t+1}}{B_{i,t}} \tag{2}$$

where $S_{i,t}$ is stock price, $D_{i,t}$ is dividend paid for unit stock, $B_{i,t}$ is net worth per stock and $X_{i,t}$ is net earnings per stock of i-th firm at time t. With these results, book-to-market ratio becomes as follows.

$$\frac{B_{i,t}}{S_{i,t}} = \frac{B_{i,t+1} + D_{i,t+t}}{ROE_{i,t+1}} \quad \frac{R_{i,t+1}}{S_{i,t+1} + D_{i,t+1}} \tag{3}$$

We take logarithm of total return, ROE and book-to-market and define as follows, $r_{i,\mathrm{t}} = \log R_{i,t}$, $roe_{i,\mathrm{t}} = \log ROE_{i,t}$, $bm_{\mathrm{t}} = \log \frac{B_{i,t}}{S_{i,t}}$, and substitute them into (3) , then we get:

$$bm_{i,\mathrm{t}} = r_{i,t+1} - roe_{i,t+1} + bm_{i,t+1} + \log(1 + e^{d_{i,t+1} - b_{i,t+1}}) - \log(1 + e^{d_{i,t+1} - s_{i,t+1}}) \tag{4}$$

where $d_{i,\mathrm{t}} = \log D_{i,t}$, $b_{i,t} = \log B_{i,t}$, $s_{i,\mathrm{t}} = \log S_{i,t}$.

Now we consider y as approximate of the long-run medium value between logarithm of dividend-to-book value ratio: $\log \frac{D_{i,t+1}}{B_{i,t+1}}$ $(= d_{i,t+1} - b_{i,t+1})$ and logarithm of dividend return: $\log \frac{D_{i,t+1}}{S_{i,t+1}}$ $(= d_{i,t+1} - s_{i,t+1})$, and we can get Taylor expansion on y as approximation of those ratio.

$$\log(1 + \mathrm{e}^{d_{i,t+1} - b_{i,t+1}}) = \log(1 + \mathrm{e}^{y}) + \frac{e^{y}}{1 + e^{y}} (\mathrm{e}^{d_{i,t+1} - b_{i,t+1}} - y) \tag{5}$$

$$\log(1 + \mathrm{e}^{d_{i,t+1} - s_{i,t+1}}) = \log(1 + \mathrm{e}^{y}) + \frac{e^{y}}{1 + e^{y}} (\mathrm{e}^{d_{i,t+1} - s_{i,t+1}} - y) \tag{6}$$

Substitute (5) and (6) into (4), then we get next equation.

$$bm_{i,\mathrm{t}} = r_{i,t+1} - roe_{i,t+1} + \frac{1}{1 + e^{y}} bmi_{,\mathrm{t+1}} \tag{7}$$

For example, once we assume dividend-book value ratio and dividend return become 4% and 2% each, then $\log \frac{D_{i,t+1}}{B_{i,t+1}} = \log 0.04 \approx -3.2$, $\log \frac{D_{i,t+1}}{S_{i,t+1}} = \log 0.02 \approx -3.9$ so that the medium value can be regard as $\mathrm{y} = -3.5$ naturally. This implies that $\frac{1}{1+e^{y}} \approx 0.97 < 1$ and we can decide value of y according to the actual conditions. We can redefine $\frac{1}{1+e^{y}} = k$ and substitute it into (7), and then we have:

$$bm_{i,t} = r_{i,\mathrm{t+1}} - roe_{i,t+1} + k \ bm_{i,t+1} = \sum_{j=1}^{T} k^{j-1} \left(r_{i,t+j} - roe_{i,t+j}\right) + k^{T} bm_{i,t+T} \tag{8}$$

Because k satisfies $0 < k < 1$, $k^{\mathrm{T}} \to 0$ $(T \to \infty)$ and we can conclude the expected value of (8) become:

$$bm_{i,t} = \sum_{j=1}^{\infty} k^{j-1} \left(E_t\left[\ r_{i,t+j}\right] - E_t\left[roe_{i,t+j}\right]\right) \tag{9}$$

Furthermore, we rewrite $E_t\left[\ r_{i,t+j}\right] = \mu_{i,\mathrm{t+j}}$, and $E_t\left[roe_{i,t+j}\right] = \eta_{i,\mathrm{t+j}}$ and those expected value converge to the same level μ_i (this is correspond to *sustainable ROE*) with according to Auto Regressive manner:

$$\mu_{i,\mathrm{t}+1} = \mu_i + \kappa_i(\mu_{i,\mathrm{t}} - \mu_i) + \xi_{i,t+1} \quad , \qquad \eta_{i,\mathrm{t}+1} = \mu_i + \omega_i(\eta_{i,\mathrm{t}} - \mu_i) + \varepsilon_{i,t+1} \quad (10)$$

Substitute (10) to (9), and then we can get the following regression model.

$$E_t[r_{i,t+1}] = \beta_{i,0} + \beta_{i,1} bm_{i,t} + \beta_{i,2} roe_{i,t} \quad (11)$$

where the parameters satisfy:

$$\beta_{i,0} = \mu_i\left(1 - \omega_i \frac{\alpha_{i,1}}{\alpha_{i,2}}\right), \quad \beta_{\mathrm{i},1} = \frac{1}{\alpha_{i,2}}\ , \quad \beta_{\mathrm{i},2} = \omega_i \frac{\alpha_{i,1}}{\alpha_{i,2}}, \quad \alpha_{\mathrm{i},1} = \frac{1}{1-k\ \omega_i}, \quad \alpha_{\mathrm{i},2} = \frac{1}{1-k\kappa_i} \quad (12)$$

(11) is a regression model for expected return explained by $bm_{i,t}$ and $roe_{i,t}$ can be estimated with historical data, and we can derive the key parameters such as mean level of long-run capital cost μ_i, and mean-reverting speed κ_i and $\omega_i$as a solution set from simultaneous equations presented as (12).

3. Revisit the estimation procedure of Lyle-Wang model

(1) Focusing on being conditional expectation

Lyle and Wang (2015) and Ono (2015) presented empirical study on U.S. and Japanese market respectively with the same original Lyle-Wang model. When we reconsider the procedure of those studies with extreme caution, we can identify some points for improvement in the procedure.

Not only the dependent variable of the basic regression model shown as (11), but also AR(1) assumptions on expected value of total return and ROE shown as (10) are the formulae with respect to conditional expectations at time t. If we explicitly rewritten (10) as conditional expectation with superscripts on the variables, then, the equations become as follows.

$$\mu_{i,t+1}^t = \mu_i^t + \kappa_i(\mu_t^t - \mu_i^t) + \xi_{i,t+1} \quad , \qquad \eta_{i,t+1}^t = \mu_i^t + \omega_i(\eta_{i,t}^t - \mu_i^t) + \varepsilon_{i,t+1} \quad (13)$$

κ_i and ω_i are assumed to be constant within an industry group, then, (11) becomes,

$$E_t[r_{i,t+1}] = \beta_{i,0}^t + \beta_{i,1}\ bm_{i,t} + \beta_{i,2}\ roe_{i,t} \quad (14)$$

where $\beta_{i,0}^t = \mu_i^t\left(1 - \omega_i \frac{\alpha_{i,1}}{\alpha_{i,2}}\right)$, besides $\beta_{i,1}$ and $\beta_{i,2}$ follow (12). It is obvious that new regression model (14) becomes consistent with original Lyle-Wang model when the sustainable ROE (μ_i^t) meets constant level.

(2) Consistency to panel data analysis

In the previous study, Lyle and Wang (2015) and Ono (2015), assumed AR(1) parameters be identical across all stocks in a same industry group and performed regression analysis by industry groups:

$$r_{i,t+1} = \beta_{\ 0} + \beta_1 \ \ bm_{i,t} + \beta_2 \ \ roe_{i,t} + \psi_{i,t+1} \ , \qquad i = 1, \dots, N, \qquad t = t, \dots T \tag{15}$$

($\psi_{i,t+1}$ is error term) But as we mentioned previous section, we should check whether $\beta_{\ 0}$ can be constant over t or not before estimation, moreover we should select alternative model taking into account time-dependence possibility.

According to empirical knowledge, expected values of sustainable ROE (μ_i^t) implied in stock prices seem fluctuate in response to the economic conditions. For example, it is reasonable aspect that downturn in Japanese stock market from the end of 2015 was mainly caused by uncertainty due to downturn in Chinese economy and/or drops in crude oil prices. That is, the decline in stock price is explained by increasing risk premium as compensation for the increasing uncertainty in short-term perspective, while may also be explained by deteriorating economy accompanying with loss of business earnings in the longer view.

In this study, we propose following modified model consistent to (14). We add dummy variables D_t^τ to absorb the difference between $\beta_{\ 0}$ that reflect conditional expected sustainable ROEs observed in each time. Other parameters are assumed identical across stocks in the same industry group.

$$r_{i,t+1} = \beta_{\ 0} + \sum_{\tau=t}^{T} b_t D_t^\tau + \beta_1 \ \ bm_{\ i,t} + \beta_2 \ \ roe_{i,t} + \psi_{i,t+1} \quad , i = 1, \dots, N, \qquad t = t, \dots T \tag{16}$$

where

$$\beta_0 + \textstyle\sum_{\tau=t}^{T} b_t D_t^\tau = \mu^t \left(1 - \omega \frac{\alpha_1}{\alpha_2}\right), \quad \beta_1 = \frac{1}{\alpha_2} \ , \ \ \beta_2 = \omega \frac{\alpha_1}{\alpha_2}, \ \ \alpha_1 = \frac{1}{1-k \ \ \omega}, \ \ \alpha_2 = \frac{1}{1-k \ \ \kappa} \tag{17}$$

This regression model can be regard as *one-dimension fixed effect model* in panel data analysis.

In equation (16) , $\beta_{\ 0}$ is assumed identical for all stocks in the same industry, however, we can take into account for easing it to accept differences between individual stocks even in the same industry. For example, it is important to consider bland-value and/or intangible asset for corporate valuation, however those bland and intangible cannot figure out in financial statement accounts, instead they can be observed as a part of cost of capital implied in stock prices. Beyond that credit risk and/or liquidity risk also reasonably increase capital cost, so we propose another model that contain additional cross-section dummy variables D_i^τ :

$$r_{i,t+1} = \beta_{\ 0} + \sum_{\tau=t}^{T} b_t D_t^\tau + \sum_{i=1}^{N} c_i D_t^i + \beta_1 bm_{\ i,t} + \beta_2 roe_{i,t} + \psi_{i,t+1}, i = 1, \dots, N, t = t, \dots, T \tag{18}$$

This is consistent to *two-dimension fixed effect model* in the context of panel data analysis.

Along with panel data analysis, we have other alternatives *one-dimension random effect model* and *two-dimension random effect model* instead of (16) and (18) each. In the fixed effect model, the dummy variables assumed to be correlated to explanatory variables ($bm_{i,t}$ and $roe_{i,t}$), on the other hand, random effect models are assumed that the dummy variables are not correlated to the explanatory variables and estimated regarding the random effect make a part of error term by generalized least square. In the following section, we select best fitting model among the candidates by statistical test and compare the forecasting ability with the original Lyle-Wang model.

4. Empirical test

(1) Data

Data set are downloaded from QUICK Astra manager, from 1981 to 2015 yearly, available at the end of August basis, among Tokyo stock exchange 1st and 2nd section. The reason why we fixed the basis day at the end of August is that many Japanese companies closed their financial settlements at the end of March, and it took about a few months until disclosing recently, 3 or 4 months in the past years, so August (5 months margin) can be enough to avoid front running.

In every year, we exclude data rows which total return or book-to-market or ROE cannot available. New listing or delisting is taking into account if it has more than 10 years listing term length. To avoid outlier, explanatory variables in the panel dataset are winsorized to be within 1 percentile and 99 percentile point values.

(2) Model selection

At first we select the best model by using whole data (from 1981 to 2015) from following candidates:

① Pool : Pooled regression model（Lyle-Wang model）
② FixOneT : One-dimension fixed effect model (time series)
③ FixTwo : Two-dimension fixed effect model
④ RanOneT : One-dimension random effect model（time series）
⑤ RanTwo : Two-dimension random effect model

Different types of statistical test are applied to different combination of the models. At first, F-test is performed with pooled regression model and fixed effect model. Pooled regression model is consistent to fixed effect model that restricted all dummy variables to be identical. That is, F-test is performed on the null hypothesis that supports the constraint, and if rejected, the alternative fixed effect model is supported. If the null hypothesis cannot be rejected, however, pooled regression model cannot be supported because of asymmetric nature of F-test.

Next we apply Breusch-Pagan (BP) test to pooled regression model and random effect model. This test assumes null hypothesis that support pooled regression model and apply χ^2-test to Lagrange Multiplier statistics. BP test support random effect model when the null hypothesis has rejected, and pooled regression model is supported when the null cannot be rejected.

Finally we hire Hausman test to select fixed effect model or random effect model. This test focuses on m-statistics defined as variance ratio of error terms and applied χ^2-test to it. Hausman test support fixed effect model when the null hypothesis has rejected, on the other hand, random effect model is supported when the null cannot be rejected.

We use those tests in combination and select the best one from the candidates. The details of those tests can refer to Kitamura (2005).

(3) Model selection results

Table 2-a, 2-b show results of F-test and BP-test. Take a look at the results of F-test, the null hypotheses that support Pooled have rejected and FixOneT or FixTwo have supported for all industry groups. BP-test, the null hypotheses that support Pooled have also rejected and supported RanOneT or RanTwo for all industry groups. Additionally FixOneT vs FixTwo are tested by F-test and the results are not monotonic depending on industry groups. Nevertheless, pooled regression model which is identical to original Lyle-Wang model have rejected and alternative models have supported in any case.

Next table 2-c shows the results of Hausman test that compare fixed effect model and random effect model. In case of RanOneT vs FixOneT, almost the all industry groups cannot rejected null hypotheses and support RanOneT, except "1.food"group. This would be significant findings that the intertemporal gaps on expected ROE can be occurred randomly (not related to explanatory variables) and should not be considered to be fixed for certain time frame.[2]

In case of RanTwo vs FixTwo, 13 groups among 15 reject null hypotheses and support FixTwo except "2.Energy" and "17.Realestate". We can conclude in conjunction with the fact that RanOneT has supported above, the cross sectional gaps between individual stocks can exist significantly rather than randomly. This is not coincide to the assumption of Lyle and Wang (2015). However, fixed effect dummy variables which are

[2] As shorter the period of dataset, there is a tendency that FixOneT become relatively significant to RanOneT.

estimated by individual stocks are not statistically significant in all stocks. In fact, as we see in later, the accuracy of forecast on expected returns becomes significantly lower than Pool and RanOneT. Therefore, if summarize the model selection results, we can conclude that RanOneT should be recommended as the best model.

Next we present the estimated parameters κ, ω and μ which are determined by solving simultaneous equations (12) for Pool and RanOneT in the table 3-a, b. κ and ω are mean reverting speed parameter of AR(1), should be restricted 0 or more and less than 1. According previous study, we determine lower/upper limit of those parameters at -0.9999/0.9999, and aggregated the parameters with replacing the exceeded parameters to the limits. Take a look at table 3-a,b, not only the Pool (Lyle-Wang model) but also RanOneT, negative value of long-run cost of capital (μ) can be found often in the table, therefore both model seems insufficient in terms of reliability as capital cost model. For the sake of improvement, we should be required for further reconsider about (a) constant parameter assumption for explanatory variables by each industry, (b) review AR(1) assumption on log book-to-market ratio and log ROE, (c) additional proxy variables for risk factors such as credit risk and liquidity risk.

(4) Forecasting ability of expected return

To evaluate the reliability of the RanOneT model we attempt to comparative evaluation test for predictability of estimated expected return. For comparison, we choose Pool (Lyle-Wang model), RanOneT and FixTwo. While previous studies conducted evaluation tests by changing the length of the forecast period, however in this study, we evaluate predictability by simple correlations between estimated one year prediction by the models and actual return, and evaluate the magnitude and stability by quantile portfolio simulation.

More specifically, as shown in figure 1, each model is estimated by past 15 years data from reference time point and the predicted retunes are obtained by substituting book-to-market ratio and ROE available at reference point into the estimated model, and calculate cross sectional correlation between estimated returns and realized returns. This procedure is applied rolling manner from 1996 to 2014 for every industry. The results are shown in table 4, the statistics such as average, standard deviation, t-value and positive correlation probability across all period and industry. FixTwo model have two types of dummy variables, cross sectional and time series, on the other hand, RanOneT model has selected as the best fit model, therefore the time series dummy variables of FixTwo should be assumed to be zero when we use for future return forecast, whereas cross sectional fixed effect dummy variables are still hired.[3]

From table 4, RanOneT newly proposed in this study showed higher correlation than Pool model, that is, the improvement of estimation accuracy is observed. On the other hand, FixTwo does not improve the estimation accuracy. FixTwo model contains insignificant dummy coefficients and those coefficients possibly not stable, then the predictive accuracy might become poor than the other models. Therefore, we exclude the fixed effects dummy variables from FixTwo model and tried to re-evaluate the predictive accuracy. Consequently, we can confirm this model have improved almost no less than RanOneT.

(5) Quantile Portfolio simulation

We performed quantile portfolio simulation based on each model forecasts to evaluate its stability and reliability. In order to avoid industry bias, we create 10-quantile portfolios within each industry group and aggregate every 10 portfolios across the industry groups. Assuming equal weighted investing, every portfolio returns are calculated as simple average of the constituent stocks return. In addition, we assume perfect market conditions that do not take into account for trading cost, tax and market impact etc.

We construct 10-quantile portfolios at the end of August in every years from 1996 to 2014 (19 years), according to moving-window method (see Fig.1). The summarized average returns and "1th - 10th" spread return are found in Table 5. According to the results, 1st and 10th quantile portfolio returns indicate significantly different levels. The rank correlations reach almost 0.9, and average spread returns show about 9% or more than 12%. Especially, RanOneT shows the highest spread return 12.47%, exceeds Pool (original Lyle-Wang model) by 2.63%, however, we pay attention to information ratio (IR) which is stability measure of performance get some of loss. Pool at 1.24 is more stable than RanOneT at 1.16.

On the other hand, it is remarkable that FixTwo (no-fix) which is excluded fixed-effect dummy variables shows spread return at 12.37% almost the same level to FixOneT.
We also graphically illustrated yearly return and cumulative return of those models in Figure 3 to help intuitive understanding. For example, Pool shows negative return only once in 1997, while RanOneT shows two times in 1997 and 1998 and FixTwo (no-fix) shows three times. Cumulative return of RanOneT goes 10.7 times larger from 1996 to

[3] We include all listed stocks meet the conditions assuming their fixed effect dummy coefficients to be 0, even if they are not included in the estimation universe of the model estimation period.

2015, which increases at a rate greater than that of FixTwo (no-fix) goes 5.6 times.

As described above, we can conclude that RanOneT shows the best performance for return forecast among tested models.

(6) What the cross-sectional fixed effect represents in FixTwo model

In the prediction test of two-dimension fixed effect model (FixTwo), we predicted future total returns with omitting the cross-sectional fixed dummy variable which absorb the disparity between individual stocks, as a result, it was confirmed that the accuracy of the prediction was improved. Considering again for what the fixed effect represent in the model, it can be regarded as averaged historical return of individual stocks in the estimation period after controlling two explanatory variables.

As it has been pointed out in many previous studies, if a certain term past return and future return is related in negatively correlated (return reversal effect), the cross sectional fixed effect can be negatively correlated to future return. Therefore, the predictive power of the FixTwo model which contains cross sectional dummy, will cancel predictive ability, has decreased reasonably.

In fact, we can find the results of quantile portfolio simulation by using only the cross sectional fixed effect of individual stocks as forecast return ("Fix Effect" in Table 5), the rank correlation is -0.893, and the spread-return is -5.84% are the consistent results. This tendency, thought it depends on time, was very significant until middle of 2000 (see Figure 3-d).

5. Empirical test for Nikkei 225 constituents

As a measure of monetary policy, the Bandk of Japan (BOJ) began mass purchase of stock index ETFs from December 2010. According to estimates based on published data, BOJ has purchased about 1 trillion yen annually from 2011, more than half of which is said to be the Nikkei 225 ETF.

Since the Nikkei 225 index is an index calculated by the so-called Dow method, it is recognized that the several stocks with insufficient liquidity have significant contribution to the index price. Therefore, when purchased in large quantities through ETFs, can possibly causes that the prices of those high-contribution stocks will be higher than that of the fair values.

Therefore, by using the model obtained in this research, we estimated the capital cost of the group with high-contribution and verify them to be decreasing after 2011.

Specifically, we calculate the capital cost of individual stocks by RanOneT model, arrange them in order of contribution of Nikkei 225 Average constituents, then check the yearly change by calculating the difference between the average value of the top 15 stocks and the other stocks. The results are as shown in the Fig 4. In 2009 and 2010, it seems that the Lehman shock seems to be affected greatly, but when comparing the levels before 2008 and after 2011 when the Bank of Japan made a large volume purchase of ETF, we could confirm the decreasing trend. In other words, it is conjectured that there is a possibility that a large volume purchase of ETF may make stocks with a large contribution of the Nikkei 225 average relatively priced expensive.

6. Conclusion

This study examines both of modeling and estimation methods presented Lyle and Wang (2015), and propose a few new models consistently estimated to panel data analysis, and tested their prediction power relative to the original model using Japanese stock data.

In Lyle and Wang (2015), key parameters of the model are determined as solution set of simultaneous equations, in this procedure it is assumed that long-term cost of capital (sustainable ROE) is constant over all individual stocks within an industry group.

As our results, one-dimensional random-effect model is selected by statistical tests, and we confirm the model to improve predictive power of expected stock returns than the original Lyle-Wang model (based on pool regression model). Moreover, according to the fact that the random-effect model is selected as a proper model to actual data, the expected value of sustainable ROE which is assumed to be constant in the estimation period and estimated as intercept in the original model should be assumed to include a random fluctuation in itself.

Further, since the two-dimensional fixed-effect model which contains both of cross section and time series effects is more appropriate than pooled regression model, sustainable ROE can be reasonably allowed to be different from each individual stocks.

This indicate that further extension of the model will be achieved to find some appropriate additional variables to absorb cross sectional difference in risk premium structure. We continue to incorporate credit risk, liquidity risk and return-reversal proxy variable etc. to expanded model to verify the suggested improvement policy.

References

Easton, P. (2004), "PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital", The Accounting Review, Vol. 79, pp.73-96.

Lyle, M. R. and C.Y. Wang (2015), "The cross section of expected holding period returns and their dynamics: A present value approach", Journal of Financial Economics, 116(2015), 505-525.

Ministry of Economi, Trade and Industry (2014)"Ito report: Competitiveness and Incentives to sustainable growth - desirable relationship building between companies and investors – ", http://www.meti.go.jp/press/2014.html (Japanese)

Shinichiro Ono (2015) "Capital cost estimation by using book-market ratio and ROE: Verification of Lyle-Wang model by Japanese stock data", Oita University, September 2015

Yukinobu Kitamura (2005) "Panel data analysis",Iwanami Shoten

Table 1 Summary statistics of dataset

No.	Name	Number of stocks			Median of PBR			Median of ROE		
		1981/8	2000/8	2014/8	1981/8	2000/8	2014/8	1981/8	2000/8	2014/8
1	Food	74	102	97	1.56	1.06	1.04	9.71%	4.88%	5.54%
2	Energy	16	17	13	2.02	0.91	0.74	15.26%	2.63%	7.08%
3	Construction	175	212	213	1.31	0.58	0.89	9.10%	2.13%	6.89%
4	Materials/Chemical	167	184	199	1.75	0.86	0.81	8.54%	2.58%	5.41%
5	Drug	30	36	32	2.34	1.90	1.53	9.89%	5.84%	8.24%
6	Automobile/Transport	68	86	79	1.69	0.83	0.96	9.43%	2.17%	10.70%
7	Steel/Non-ferrous	79	75	61	2.19	0.86	0.86	10.56%	0.83%	4.77%
8	Machine	103	137	155	2.06	1.08	1.03	11.38%	1.74%	7.24%
9	Electric/Precision	151	187	188	2.65	1.57	1.12	12.09%	2.87%	6.87%
10	IT/Communication	48	106	193	2.00	1.04	1.23	8.88%	3.84%	6.77%
11	Electricity/Gas	15	19	20	1.39	1.09	0.90	18.61%	5.49%	4.81%
12	Transportation/Logistics	69	80	90	2.07	1.20	0.88	7.63%	1.94%	6.04%
13	Trading co.	60	113	175	1.64	0.82	0.72	9.47%	4.25%	6.77%
14	Retail	33	113	159	1.67	1.01	1.12	9.98%	3.51%	5.26%
15	Bank	72	84	74	1.25	0.99	0.54	6.79%	3.20%	5.48%
16	Finance(ex. Bank)	6	57	36	2.85	1.25	1.05	11.28%	8.27%	8.53%
17	Real Estate	14	26	36	1.77	1.01	1.10	7.37%	5.00%	5.08%
	Total/Average	1,180	1,634	1,820	1.89	1.06	0.97	10.35%	3.60%	6.56%

The universe contains First and Second section of Tokyo Stock Exchange excluded if listed period less than 12 months. Also negative book-to-market value, missing ROE records are omitted..

Table 2-a Model selection : F-test and Breusch-Pang test

Industry	Model	Null Model	Test	DF-model	DF-null	F-value m-value	P-value
1.Food	FixOneT	Pool	F test	33	1,949	54.1	0.000
	FixTwo	Pool	F test	91	1,891	20.8	0.000
	FixTwo	FixOneT	F test	58	1,891	1.4	0.020
	RanOneT	Pool	BP test	1	–	5583.5	0.000
	RanTwo	Pool	BP test	2	–	822.0	0.000
2.Energy	FixOneT	Pool	F test	33	164	8.1	0.000
	FixTwo	Pool	F test	38	159	7.2	0.000
	FixTwo	FixOneT	F test	5	159	1.1	0.374
	RanOneT	Pool	BP test	1	–	27.3	0.000
	RanTwo	Pool	BP test	2	–	8.5	0.004
3.Construction	FixOneT	Pool	F test	33	4,750	164.5	0.000
	FixTwo	Pool	F test	176	4,607	31.9	0.000
	FixTwo	FixOneT	F test	143	4,607	1.1	0.128
	RanOneT	Pool	BP test	1	–	60276.2	0.000
	RanTwo	Pool	BP test	2	–	1250.8	0.000
4.Material Chemical	FixOneT	Pool	F test	33	4,621	139.7	0.000
	FixTwo	Pool	F test	172	4,482	28.3	0.000
	FixTwo	FixOneT	F test	139	4,482	1.4	0.002
	RanOneT	Pool	BP test	1	–	33026.5	0.000
	RanTwo	Pool	BP test	2	–	276.8	0.000
5.Drug	FixOneT	Pool	F test	33	663	15.0	0.000
	FixTwo	Pool	F test	53	643	9.9	0.000
	FixTwo	FixOneT	F test	20	643	1.3	0.204
	RanOneT	Pool	BP test	1	–	478.5	0.000
	RanTwo	Pool	BP test	2	–	54.7	0.000
6.Automobile Transport	FixOneT	Pool	F test	33	1,896	53.7	0.000
	FixTwo	Pool	F test	90	1,839	20.5	0.000
	FixTwo	FixOneT	F test	57	1,839	1.1	0.210
	RanOneT	Pool	BP test	1	–	4225.6	0.000
	RanTwo	Pool	BP test	2	–	278.1	0.000
7.Steel Non-ferrous	FixOneT	Pool	F test	33	1,690	56.0	0.000
	FixTwo	Pool	F test	85	1,638	22.5	0.000
	FixTwo	FixOneT	F test	52	1,638	1.1	0.332
	RanOneT	Pool	BP test	1	–	5264.5	0.000
	RanTwo	Pool	BP test	2	–	12.0	0.001
8. Machine	FixOneT	Pool	F test	33	2,979	99.4	0.000
	FixTwo	Pool	F test	124	2,888	27.8	0.000
	FixTwo	FixOneT	F test	91	2,888	1.4	0.014
	RanOneT	Pool	BP test	1	–	18066.8	0.000
	RanTwo	Pool	BP test	2	–	128.6	0.000
9.Electric Precision	FixOneT	Pool	F test	33	3,440	98.4	0.000
	FixTwo	Pool	F test	137	3,336	24.8	0.000
	FixTwo	FixOneT	F test	104	3,336	1.3	0.045
	RanOneT	Pool	BP test	1	–	19878.2	0.000
	RanTwo	Pool	BP test	2	–	115.0	0.000
10. IT Communication	FixOneT	Pool	F test	33	1,539	40.1	0.000
	FixTwo	Pool	F test	80	1,492	17.8	0.000
	FixTwo	FixOneT	F test	47	1,492	1.6	0.007
	RanOneT	Pool	BP test	1	–	3278.5	0.000
	RanTwo	Pool	BP test	2	–	21.8	0.000

High-lights correspond to unable to reject Null Model.

Table 2-b Model selection : F-test and Breusch-Pang test

Industry	Model	Null Model	Test	DF-model	DF-null	F-value m-value	P-value
11.Electricity Gas	FixOneT	Pool	F test	33	504	31.6	0.000
	FixTwo	Pool	F test	48	489	22.6	0.000
	FixTwo	FixOneT	F test	15	489	1.5	0.093
	RanOneT	Pool	BP test	1	–	1467.1	0.000
	RanTwo	Pool	BP test	2	–	638.8	0.000
12.Transportation Logistics	FixOneT	Pool	F test	33	2,214	83.8	0.000
	FixTwo	Pool	F test	100	2,147	28.7	0.000
	FixTwo	FixOneT	F test	67	2,147	1.2	0.093
	RanOneT	Pool	BP test	1	–	11363.8	0.000
	RanTwo	Pool	BP test	2	–	202.3	0.000
13. Trading co.	FixOneT	Pool	F test	33	1,730	73.6	0.000
	FixTwo	Pool	F test	85	1,678	30.0	0.000
	FixTwo	FixOneT	F test	52	1,678	1.5	0.011
	RanOneT	Pool	BP test	1	–	8564.0	0.000
	RanTwo	Pool	BP test	2	–	49.3	0.000
14. Retail	FixOneT	Pool	F test	33	583	17.8	0.000
	FixTwo	Pool	F test	51	565	12.3	0.000
	FixTwo	FixOneT	F test	18	565	1.6	0.062
	RanOneT	Pool	BP test	1	–	879.5	0.000
	RanTwo	Pool	BP test	2	–	11.8	0.001
17. Real Estate	FixOneT	Pool	F test	33	202	23.2	0.000
	FixTwo	Pool	F test	39	196	19.9	0.000
	FixTwo	FixOneT	F test	6	196	1.1	0.346
	RanOneT	Pool	BP test	1	–	407.1	0.000
	RanTwo	Pool	BP test	2	–	409.1	0.000

High-lights correspond to unable to reject Null Model.

Table 2-c Model selection : Hausman test

Industry	model	null model	test	DF-model	m-value	P-value
1.Food	FixOneT	RanOneT	Hasuman	2	7.29	0.026
	FixTwo	RanTwo	Hasuman	2	38.89	0.000
2.Energy	FixOneT	RanOneT	Hasuman	2	0	0.998
	FixTwo	RanTwo	Hasuman	2	2.4	0.302
3.Construction	FixOneT	RanOneT	Hasuman	2	2.6	0.273
	FixTwo	RanTwo	Hasuman	2	77.83	0.000
4.Material Chemical	FixOneT	RanOneT	Hasuman	2	0.84	0.658
	FixTwo	RanTwo	Hasuman	2	92.6	0.000
5.Drug	FixOneT	RanOneT	Hasuman	2	0.71	0.700
	FixTwo	RanTwo	Hasuman	2	13.46	0.001
6.Automobile Transport	FixOneT	RanOneT	Hasuman	2	0.38	0.829
	FixTwo	RanTwo	Hasuman	2	39.04	0.000
7.Steel Non-ferrou	FixOneT	RanOneT	Hasuman	2	1.04	0.596
	FixTwo	RanTwo	Hasuman	2	31.93	0.000
8. Machine	FixOneT	RanOneT	Hasuman	2	0.17	0.919
	FixTwo	RanTwo	Hasuman	2	53.83	0.000
9.Electric Precision	FixOneT	RanOneT	Hasuman	2	0.57	0.754
	FixTwo	RanTwo	Hasuman	2	62.19	0.000
10. IT Communication	FixOneT	RanOneT	Hasuman	2	2.45	0.294
	FixTwo	RanTwo	Hasuman	2	33.03	0.000
11.Electricity Gas	FixOneT	RanOneT	Hasuman	2	4.22	0.121
	FixTwo	RanTwo	Hasuman	2	11.6	0.003
12.Transportation Logistics	FixOneT	RanOneT	Hasuman	2	0.78	0.679
	FixTwo	RanTwo	Hasuman	2	41.44	0.000
13. Trading co.	FixOneT	RanOneT	Hasuman	2	1.59	0.451
	FixTwo	RanTwo	Hasuman	2	32.29	0.000
14. Retail	FixOneT	RanOneT	Hasuman	2	0.92	0.633
	FixTwo	RanTwo	Hasuman	2	9.99	0.007
17. Real Estate	FixOneT	RanOneT	Hasuman	2	0.18	0.914
	FixTwo	RanTwo	Hasuman	2	1.69	0.430

High-lights correspond to unable to reject Null Model.

Table 3-a Parameters estimated : Pool regression (Lyle—Wang model)

	Intercept	bm	roe	κ	ω	μ
5%	−0.0682	−0.0014	−0.3637	0.8653	−0.9999	−0.0779
25%	−0.0202	0.0538	−0.0828	0.9021	0.1046	−0.0219
Mean	0.0194	0.0920	0.1119	0.9348	0.4315	0.0208
Median	0.0118	0.0965	0.0209	0.9343	0.7071	0.0131
75%	0.0484	0.1277	0.2019	0.9785	0.9987	0.0486
95%	0.1533	0.1633	0.7751	0.9999	0.9999	0.1576
Median						
1.Food	0.0096	0.1108	0.3159	0.9195	0.7071	0.0136
2.Energy	0.0474	0.1068	−0.6788	0.9237	0.9999	0.0223
3.Construction	−0.0436	0.1204	0.0803	0.9096	0.5082	−0.0471
4.Material/Chemical	0.0003	0.1057	0.0883	0.9248	0.4541	0.0004
5.Drug	0.0783	0.1033	0.1548	0.9273	0.6726	0.0922
6.Automobile/Transport	0.0278	0.1211	−0.0284	0.9088	0.0324	0.0247
7.Steel/Non-ferrou	−0.0056	0.0958	−0.0568	0.9350	0.0317	−0.0056
8. Machine	0.0173	0.1160	−0.0596	0.9142	−0.4880	0.0164
9.Electric/Precision	0.0257	0.0997	−0.0410	0.9310	−0.3604	0.0239
10. IT/Communication	−0.0140	0.0757	−0.0735	0.9559	0.9553	−0.0118
11.Electricity/Gas	0.0133	0.1395	1.2942	0.8899	0.9332	0.0309
12.Transportation/Logistics	0.0194	0.0597	−0.0199	0.9724	0.6267	0.0215
13. Trading co.	−0.0201	0.0869	0.0009	0.9443	0.4746	−0.0223
14. Retail	−0.0178	0.0518	0.2372	0.9806	0.8163	−0.0205
17. Real Estate	0.0295	0.1142	0.0026	0.9160	0.9987	0.0261

κ, ω and μ are solutions to simultaneous equations of estimated parameters. For κ andω, we set the lower limit as -0.9999, and upper limit as 0.9999, and calculate the average with modifying exceeded values to the lower/upper limit values.

Trends in all industry average

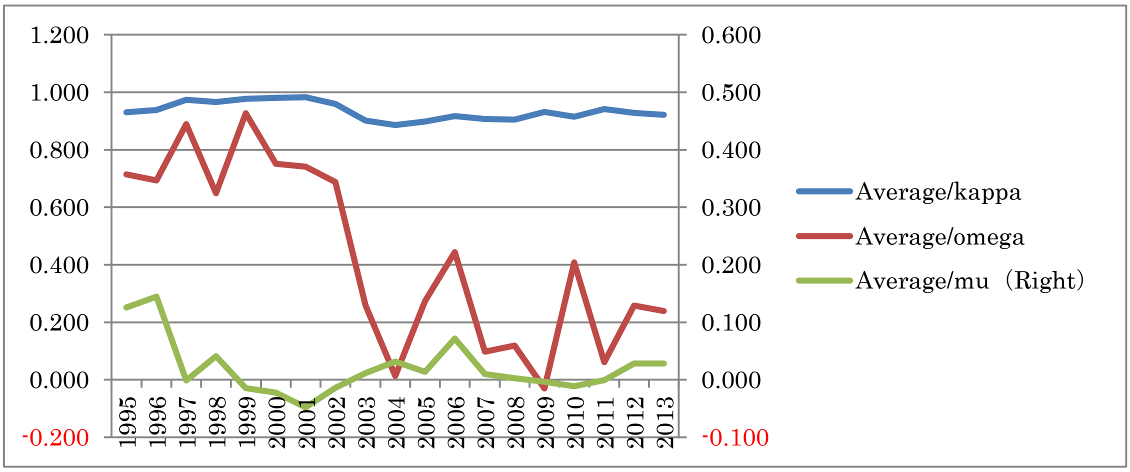

Table 3-b Parameters estimated : One-dimension Fixed effect model (FixOneT)

	Intercept	bm	roe	κ	ω	μ
5%	−0.0452	0.0269	−0.2931	0.9282	−0.9999	−0.0495
25%	−0.0168	0.0518	−0.0754	0.9488	0.3569	−0.0165
Mean	0.0236	0.0673	0.0275	0.9634	0.5120	0.0323
Median	0.0130	0.0687	0.0326	0.9631	0.6869	0.0133
75%	0.0497	0.0825	0.1125	0.9806	0.9999	0.0517
95%	0.1339	0.1024	0.3669	0.9999	0.9999	0.1487
Median						
1.Food	0.0103	0.0884	0.1655	0.9427	0.6660	0.0128
2.Energy	0.0382	0.0754	−0.5286	0.9562	0.9999	0.0272
3.Construction	−0.0285	0.0871	0.0640	0.9441	0.4839	−0.0321
4.Material/Chemical	0.0050	0.0627	0.0743	0.9693	0.5569	0.0055
5.Drug	0.0524	0.0746	0.2010	0.9570	0.7817	0.0588
6.Automobile/Transport	0.0290	0.0805	−0.1031	0.9509	0.9999	0.0269
7.Steel/Non-ferrou	0.0052	0.0632	0.0226	0.9688	0.5294	0.0053
8. Machine	0.0106	0.0760	0.0116	0.9556	0.3044	0.0107
9.Electric/Precision	0.0128	0.0686	−0.0060	0.9632	0.2395	0.0117
10. IT/Communication	−0.0021	0.0656	−0.0649	0.9663	0.9999	−0.0019
11.Electricity/Gas	0.0384	0.0736	0.2841	0.9580	0.8520	0.0404
12.Transportation/Logistics	0.0045	0.0406	−0.0197	0.9921	0.8001	0.0042
13. Trading co.	−0.0120	0.0671	0.0031	0.9647	0.4303	−0.0115
14. Retail	−0.0212	0.0400	0.1593	0.9927	0.7552	−0.0296
17. Real Estate	0.0083	0.0310	−0.0922	0.9999	0.9998	0.0069

κ, ω and μ are solutions to simultaneous equations of estimated parameters. For κ andω, we set the lower limit as -0.9999, and upper limit as 0.9999, and calculate the average with modifying exceeded values to the lower/upper limit values.

Trends in all industry average

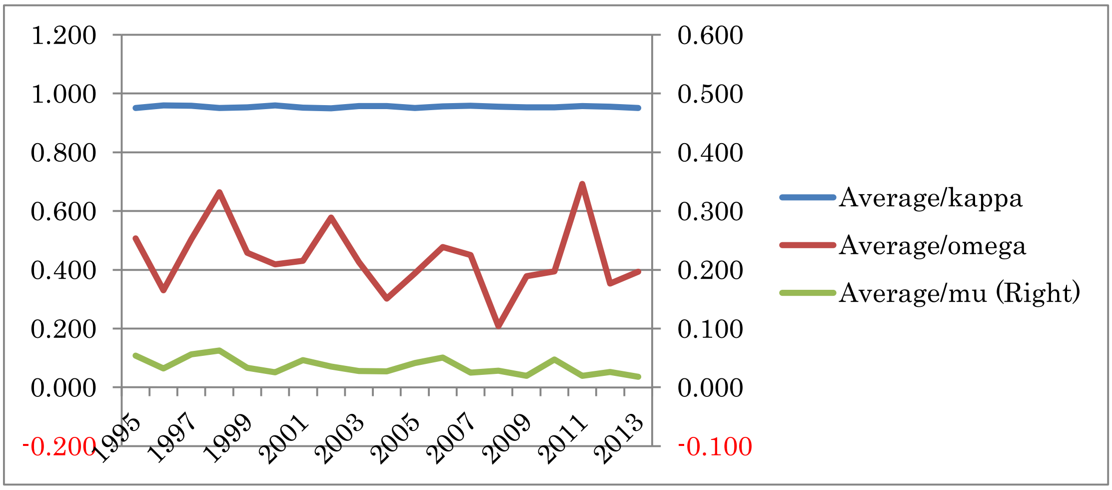

Fig. 1　Forecasting returns and portfolio simulation procedure

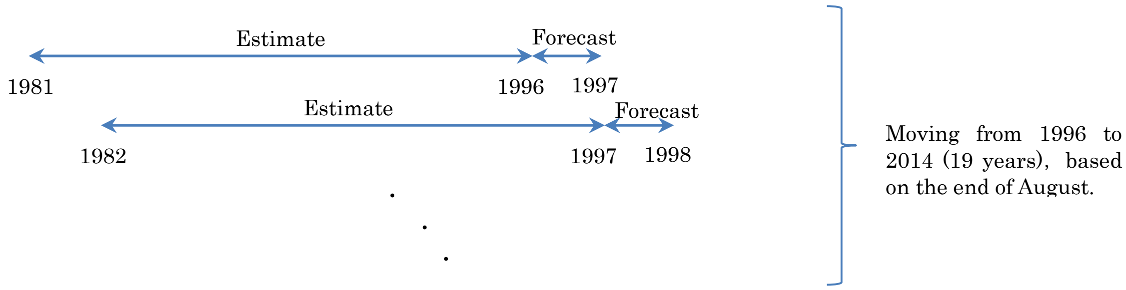

Table. 4　Correlation between predicted returns and actual returns

	Estimated by 15 years data			
	Pool	RanOneT	FixTwo	FixTwo −nofix−
Average	0.087	0.112	0.084	0.114
SD	0.227	0.223	0.201	0.222
t-value	3.40	4.45	3.71	4.55
PProb	68.8%	71.6%	69.8%	73.3%

The correlation coefficients are estimated of each industry sector for each year, and aggregated simple average and standard deviation for all periods across all industries. “PProb” means positive probability that correlation coefficient becomes a positive value.

Table 5　Quantile portfolio simulation

	Estimated by 15 years data				
	Pool	RanOneT	FixTwo	FixTwo (no-fix)	(Fix Effect)
1st	1.68%	2.52%	0.73%	2.94%	−4.33%
10th	−8.16%	−9.94%	−8.34%	−9.44%	1.51%
Rank Corr	0.890	0.890	0.803	0.913	−0.893
1th − 10th spread return					
Average	9.84%	12.47%	9.07%	12.37%	−5.84%
SD	7.94%	10.74%	11.80%	12.88%	11.61%
IR	1.24	1.16	0.77	0.96	−0.50

According to the predicted returns, we classify all stocks into 10 quantile groups within each industry group, and summarize then into 10 quantile portfolios across industry for calculate yearly actual return and spread return from 1996 to 2014. We test 5 models such as Pool: pooled regression model, RanOneT: random effect model (time series), FixTwo: two-dimensional fixed effect model, FixTwo(no-fix): FixTwo except time series and cross sectional fixed effect. “Rank Corr” means rank correlation.

Figure 3-a　Spread return (1st-10th) : Pool (Lyle-Wang model)

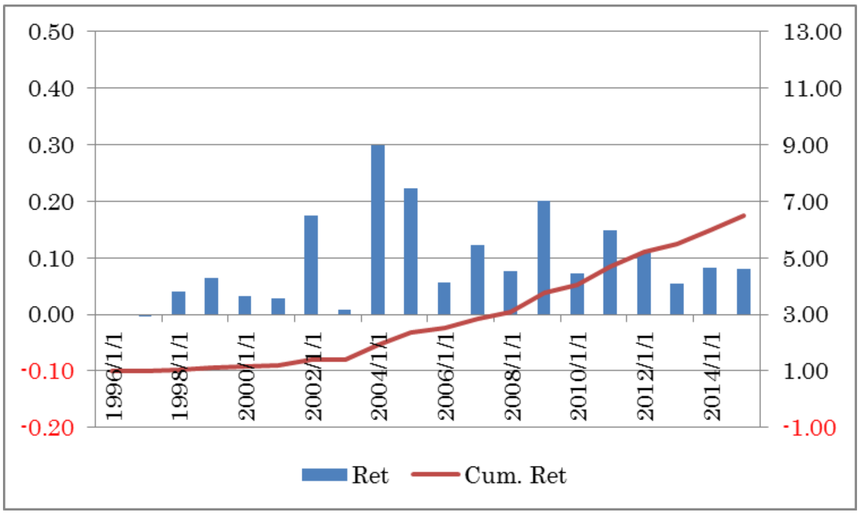

Figure 3-b　Spread return (1st-10th)：RanOneT

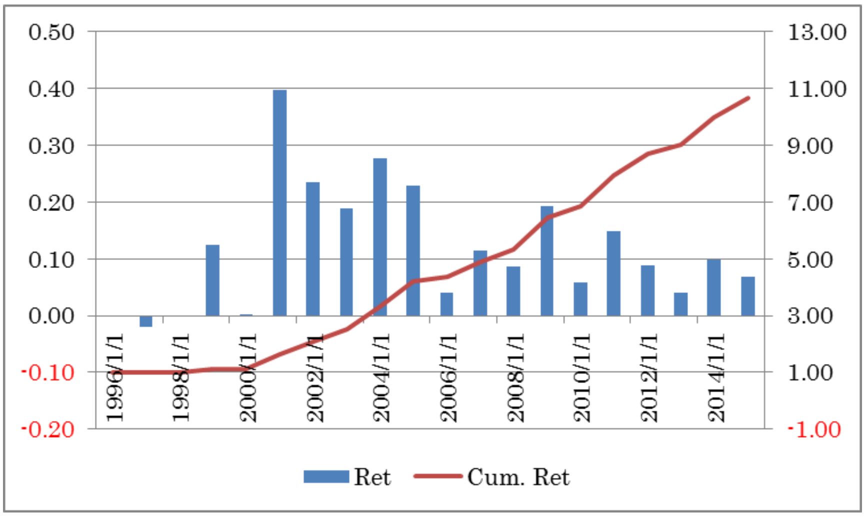

Figure 3-c　Spread return (1st-10th)：FixTwo (no-fix)

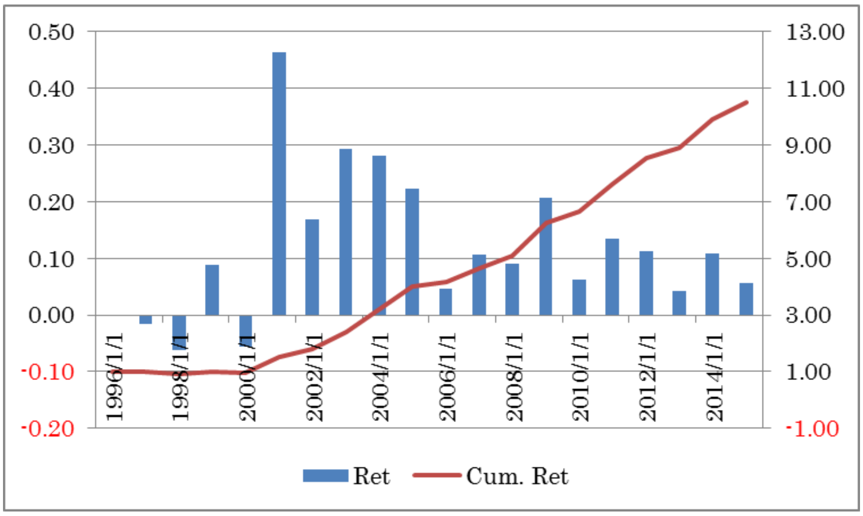

Figure　3-d　Spread　return　(1^{ST}–10^{th})　:Fix　Effect

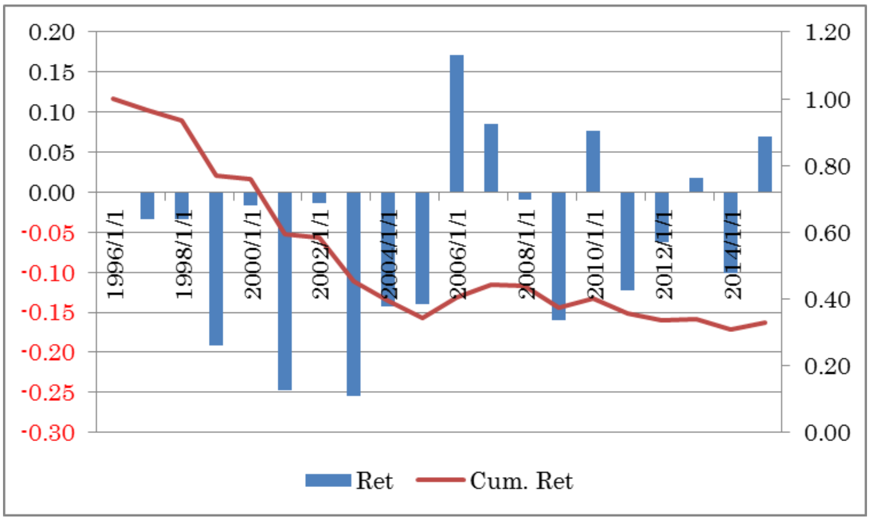

Figure 4 Difference of capital costs between two groups of Nikkei 225 stocks