市場急変に対応する予測モデルの提案

Forecast Model Corresponding to Sudden Market Change

古幡征史 [1]* 水田孝信 [2] 曹 治平 [2]
Masabumi Furuhata[1] Takanobu Mizuta[2] Jihei So[2]

[1] 北陸先端科学技術大学院大学
[1] Japan Advanced Institute of Science and Technoloy
[2] スパークスアセットマネジメント株式会社
[2] SPARX Asset Management Co., Ltd.

Abstract: In financial markets, sudden unexpected changes occur frequently. We propose a new forecast method based on paired evaluators consisting of the stable evaluator and the reactive evaluator that is good at detecting and adapting to the consecutive market changes. We conduct a back-testing using financial data in US. The experimental results show that our method is effective and robust even against the late-2000s recessions.

1 Introduction

To cope with risks in volatile financial markets, portfolio theory has been used as a standard tool for more than thirty years. Modern portfolio theory is based on capital asset pricing model (CAPM) established by Sharpe [17], Lintner [14], and Mossin [15]. A main characteristic is an emphasis on a price discovery process rather than pricing itself. In the CAPM, a theoretically appropriate required rate of return of an asset is obtained according to a consideration of the expected return of the market, the expected return of a theoretical risk-free asset and non-diversifiable risk. Hence, the non-diversifiable risk is used as a single *factor* to compare the excess returns of a portfolio with the excess returns of the entire market that entails the set of optimal equities for a portfolio. More recently, Fama and French [6] propose two risk factors, *value* and *size*, and Carhart [4] proposes a factor, *momentum* that are widely accepted to reduce some exceptional cases of the CAPM:

- momentum: historical price increase for 12 months,
- value: book-to-market ratio,
- size: size of a firm (market capitalization).

Even though several factors have been proposed to predict future market movements, a persistent factor has not found yet. Hence, a key issue for investors based on factors is to select the best factor which suddenly and significantly changes over time.

In data mining and machine learning, several methods have been proposed to deal with changes over time in unforeseen ways known as *concept drift*. In this paper, we view quick changes of financial markets as concept drift problems and propose a solution for these problems. A main difficulty to deal with concept drift is the greater number of observations does not simply lead to the increase of forecast accuracy unlike phenomena governed by laws of nature.

Researches dealing with concept drift are extensive such as determination of window size [13, 18], change detections [2, 3, 8, 9], and adaptive ensembles [11]. Our research is most closely related to the determination of window size for a prediction. In this domain, there are two main streams to cope with concept drifts, dynamically changing the window size [13, 18] or using two fixed window sizes [1, 16]. In the former stream, as soon as they observe a new data, they investigate consistencies with the histories. Once they suspect an occurrence of concept drift, they adjust their window sizes. In the latter stream, they use paired classifiers to control two types of window size. A common point in the researches in both streams is that they adjust window size for classification problems.

In contrast, we propose a new forecast method that runs a set of base forecast having different window

*連絡先：北陸先端科学技術大学院大学　知識科学研究科
E-mail: k439bk439b@gmail.com

sizes of reference histories to generate forecast values. Among the base forecasts, the most accurate base forecast on back-testing is selected by an evaluator. Hence, the best window size in the historical data is selected by our method instead of adjusting window size like existing work above. In order to cope with concept drifts, we use two types of evaluators, a stable evaluator and a reactive evaluator, similarly to the existing researches of the paired classifiers [1, 16]. While the stable evaluator is used as a default evaluator which is supposed to be appropriate for versatile situations, the reactive evaluator is sensitive to changes. If performances of the reactive evaluator exceeds ones of the stable evaluator, our method switches to use a base forecast selected by the reactive evaluator. With respect to decisions for switches, we use learning algorithm according to the histories of performances. A main characteristic of our proposing method is robustness against consecutive occurrences of concept drifts. We examine a back-testing using actual financial data in US in order to demonstrate how our proposing method performs compared with other existing methods.

For a prediction, we avoid an investor ' s intuition to select a factor to evaluate a performance of our forecasting method purely. In addition, we do not rely on the external data such as macro economic statistics in order to be independent from noises contained in the external data.

The rest of this paper is organized as follows. In Section 2, we specify the research problems using actual market data. In Section 3, we detail our proposing forecasting method. Some key characteristics of our method are shown with some examples in Section 4. In Section 5, we examine a back-testing and compare the performances of our method with ones of other representative approaches. In Section 6, we concludes this paper.

2 Fund Operation in Volatile Financial Market

In this section, we detail a factor selection problem in financial markets using monthly historical factor data in US equities market which can be obtained from [7]. We focus on three factors, momentum, value, and size as described in the previous section. For a comprehension of the effectiveness of factors, we calculate factor spread return between the top 10 % and the bottom 10 %. Among these factors, we investigate a method to predict the most effective factor in each month. In Figure 1, we illustrate cumulative monthly 10 percentile spread return on these factors in US equity markets from 1964 to 2009 classified by decades. If a cumulative spread return is increased constantly and sharply, this factor is considered to be effective. In addition, a constant sharp decrease is also an effective which works for contrarians. Considering these points, we observe that momentum has been well-performed until 1990s; however it fluctuates heavily in 2000s. Size factor is constantly decreased in 1960s, 1970s and 2000s. This negative sign is a desired phenomenon, since smaller size companies are expected to grow faster as proposed in Fama and French [6]. Throughout the years, effective factors change over time quickly and sharply. In the figure, we illustrate a case to invest equally on these factors that is average investment, $(momentum + value - size)/3$, as a benchmark purpose. While the average investment does not make a huge loss entirely, it looses some opportunities to gain greater profits. We are interested in developing a forecasting method that gains greater profits without increasing risks of huge losses.

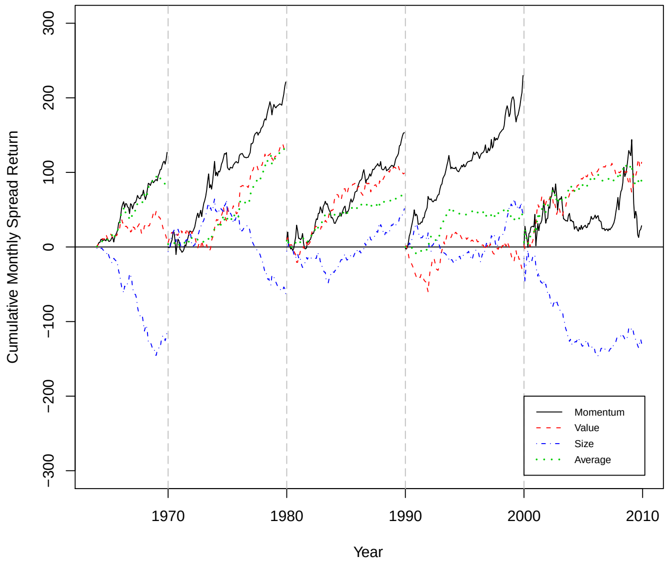

図 1: Cumulative Monthly Spread Return of Factor Investments Classified by Decades

In contrast to this static approach, our interest is to develop a forecasting method that predicts the most effective factor, which may change over time. As we are able to comprehend from the historical data in the figure, the most effective factors are not constant over time. Difficulties with respect to the prediction of the most effective factors are mainly caused by the following two reasons. First, the price discovery process is not consistent over time for each factor. Some factors may be rapidly effective, but others may take time. Second, an effective factor can be either forward or contrary. Hence, forecasting methods are necessary to predict both the largest absolute value of factors and their signs. Under such a dynamically changing environment, key questions are how to identify quick market changes and how to adapt to these changes appropriately. Notice that these changes are not unique. In some cases, the most effective factor is suddenly swapped by another factor. In other cases, the swap is gradually occurred. An important aspect is that the change types are not fixed. Hence, a challenge is to develop a forecast model which is robust against some different types of change types. In the following section, we propose our forecasting method considering these aspects.

3 Proposing Method

In this section, we detail our proposing forecasting method, paired evaluators method (PEM), which is adaptive to market changes.

3.1 Preliminaries

Let $\mathcal{T} = \{-T, \dots, -1, 0, 1, \dots, T\}$ be the set of discrete time and $t \in \mathcal{T}$ be a certain time. We call $t = 0$ as the current time. We denote $\mathcal{H} = \{t \in \mathcal{T} : t < 0\}$ as the set of historical periods. Let $X_t \in \Re^z$ be a vector in z-dimensional feature space observed at time $t \in \mathcal{T}$ and $y_t \in \Re$ be its corresponding label to be predicted. We refer to X_t as an instance, a pair (X_t, y_t) as a labeled instance, instances $(X_{-T}, \dots, X_{-1})$ as historical data X^H, and an instance X_0 as a target instance. As time is incremented, the number of historical data is increased and the current time is shifted. Notice that a target instance X_0 is not observed until time is incremented.

3.2 An Overview of Our Proposing Forecasting Method

In this section, we present an overview of our forecasting method. We use several different base forecasts and select a base forecast which is expected to be the best forecast according to the past experiences. We refer a way to select a base forecast as an evaluator. A key characteristic of our method is an evaluation of performances of evaluators.

Let f be a base forecast. Let $\mathcal{F}$ be a set of base forecast. Let us denote f^i as the i-th forecast among $\mathcal{F}$ and we also denote $\mathcal{I}$ as the set of forecasts. We denote $\delta_t^i = y_t - f^i(X_t^H)$ as a forecast error at time t which is obtained at time $t+1$.

Our proposing forecast method proposes the optimal base forecast $f^{\tilde{i}}$ among a set of base forecasts $\mathcal{F}$ without severe parameter tunings. As depicted in Figure 2, this method consists of four parts: (1) pre-processing, (2) base forecast proposal according to evaluator, (3) evaluator selection, and (4) forecasting. In part (1), past forecast errors δ for all base forecast are calculated, respectively.

In part (2), we use paired evaluators that are a stable evaluator and a reactive evaluator. Each evaluator proposes the expected best base forecast based on different weights for evaluations. Let $\mathbf{w}$ be a weight vector of forecast errors. Let $\tilde{i}$ be the $\tilde{i}$-th base forecast which is expected to be the best base forecast. Multiplying a weight vector to forecast errors, we are able to compare performances of base forecasts and we are also able to obtain the estimated best base forecast as follows:

$$\tilde{i} = \arg\min_{i \in \mathcal{I}} \sum_{t \in \mathcal{H}} w_t \delta_t^i \tag{1}$$

where w_t is an element of the weight vector. According to Equation (1), a base forecast that minimizes the weighted errors is estimated as the best base forecast. Here, let us denote g as an evaluator that selects the optimal base forecast based on Equation (1).

If a weighting vector has heavier weights for more recent errors, it prefers short-term forecast accuracy which is a reactive evaluator g^R. Contrary, the flat weight vector prefers long-term forecast accuracy which is a stable evaluator g^S. Long-term accuracy is preferred in general. However, right after a concept drift, the long history of forecast errors may not tell a proper forecast. According to these errors and evaluators, a base forecast is proposed as shown in Equation (1).

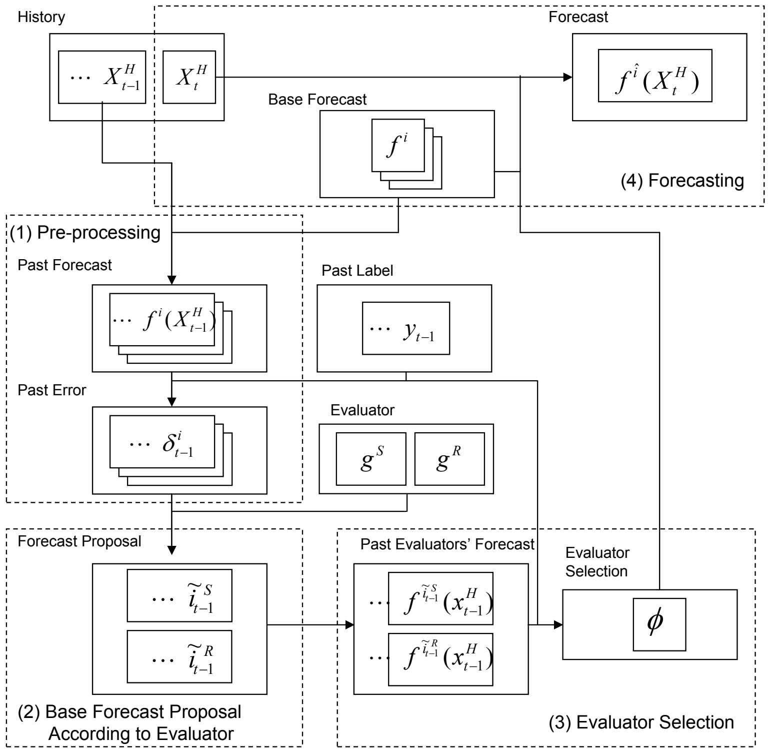

図 2: Structure of Paired Evaluator Method

Let $\tilde{i}^S_t$ be the expected best base forecast according to the stable evaluator which is $\tilde{i}^S_t = g^S(\boldsymbol{\delta}^H_t)$, where $\boldsymbol{\delta} = \{\delta^i\}_{i \in \mathcal{I}}$. Similarly, the expected best base forecast according to the reactive evaluator is $\tilde{i}^R_t = g^R(\boldsymbol{\delta}^H_t)$.

Since the proposed base forecast depends on the setting of the evaluator, a central issue is how to select evaluators which is part (3) of our method. With respect to the selection, there are some key ideas behind our method. First, the stable evaluator works well under versatile situations. It is used as a default evaluator. Second, the stable evaluator may not work well right after a poor performance. If the reactive evaluator has performed better than the stable evaluator at the similar cases in the past, our method switches to use the reactive evaluator. Finally, if past experiences are inconsistent, recent experiences have a greater importance for decision makings. Considering these aspect, we select an evaluator based on a learning-algorithm. Let Φ be an evaluation function of evaluators that assigns a degree of superior evaluator on a performance of the stable evaluator. Based on this evaluation function, we obtain the expected best base forecast $\hat{i}$ such that:

$$\hat{i}_t = \begin{cases} \tilde{i}^S_t, & \text{if } \Phi_t(\delta_{t-1}^{\tilde{i}^S_{t-1}}) \geq \theta; \\ \tilde{i}^R_t, & \text{otherwise.} \end{cases} \tag{2}$$

where θ is a threshold parameter. We detail how the evaluation function is updated according to the past experiences in the following section.

Once an evaluator is selected in part (3), part (4) is directly induced and we obtain the best performing base forecast $f^{\hat{i}}$ and its forecast value $f^{\hat{i}}(X^H_t)$. In the following section, we detail part (3) of our proposing forecasting method.

3.3 Update Rules of the Evaluation Function

The evaluation function of evaluators is updated according to a learning-based approach that consists of three types of update rules: (i) initialization, (ii) a performance of an evaluator exceeded a performance of another evaluator, and (iii) no differences on performances between two evaluators. Regarding to the first update rule, we set $\Phi(\delta) = 0$ for all δ. The second rule is for cases where one evaluator performs better than another. Once actual value y_{t-1} is realized at time t, we obtain forecast errors of the stable evaluator and the reactive evaluator, $\delta_{t-1}^{\tilde{i}_{t-1}^S}$ and $\delta_{t-1}^{\tilde{i}_{t-1}^R}$, respectively. If the stable evaluator performs better than the reactive evaluator, i.e., $\delta_{t-1}^{\tilde{i}_{t-1}^S} < \delta_{t-1}^{\tilde{i}_{t-1}^R}$, $\Phi_t(\delta) := \Phi_t(\delta) + \lambda$ for all $\delta \leq \delta_{t-1}^{\tilde{i}_{t-1}^S}$, where $\lambda > 0$ is an update coefficient. Contrary, if the stable evaluator performs worse than the reactive evaluator, i.e., $\delta_{t-1}^{\tilde{i}_{t-1}^S} > \delta_{t-1}^{\tilde{i}_{t-1}^R}$, $\Phi_t(\delta) := \Phi_t(\delta) - \lambda$ for all $\delta \geq \delta_{t-1}^{\tilde{i}_{t-1}^S}$. The third rule is for cases where both evaluators perform equal. In such cases, the effects of past experiences are reduced by a reducing coefficient $0 \leq \alpha \leq 1$ as follows, $\Phi_t(\delta) := \alpha\Phi_t(\delta)$ for all δ.

4 Paired Evaluators Method and Drift Types

In this section, we show some simple examples to show how our proposing forecasting method, paired evaluators method (PEM), deals with typical drift types, sudden drift, incremental drift, gradual drift, and recurring contexts. Notice that a single window size approach has a problem to deal with these drift types. While smaller window sizes tend to fit for sudden drift and incremental drift, they are too sensitive for gradual drift and recurring contexts. It is significant to switch to an appropriate window size that corresponds to an observed drift type. In order to switch the window size, PEM uses two types of evaluators, the stable evaluator and the reactive evaluator, as we have shown in the previous section. This approach, particularly, works for gradual drift and recurring contexts as we show in some examples.

We prepare the set of artificial data that characterizes typical drift types. There are two time series that are either $100 \pm \epsilon$ or $20 \pm \epsilon$ where ϵ is randomly drawn from uniform distribution ranging $[-5, 5]$. A key task is to predict a series that will be the greater value in the following time. Regarding to PEM, we use three base forecasts, 3-month average, 6-month average and 12-month average; we set an update coefficient as 1, a decreasing coefficient as 0, a threshold as 0. As for parameter of evaluators, we use 12 months equitable weights $w = \{1,1,1,1,1,1,1,1,1,1,1,1\}$ for a stable evaluator and 3 months equitable weights $w = \{1,1,1\}$ for a reactive evaluator.

At first, we show examples of sudden drift and incremental drift in Figure 3. The top graphs show

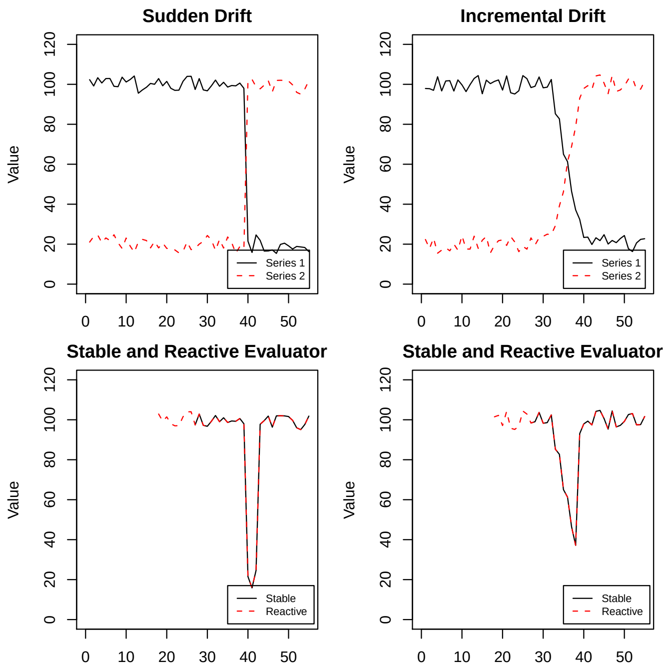

図 3: Sudden Drift and Incremental Drift

series 1 and 2 of the respective drift patterns. For both cases, series 1 is to be predicted at the beginning, series 2 swaps at certain time, and series 2 is to be predicted after the swap. While sudden drift occurs at time 39, incremental drift starts changing at time 33, takes over observed at time 37, and ends changing at time 39. The bottom graphs show performances of paired evaluators. For both drift patterns, both evaluators have exactly the same performances. They require the minimum time, three time periods, to correspond to the drifts. Differences between the stable evaluator and the reactive evaluator are observed if drifts occur more frequently.

Next, we show examples of gradual drift and recurring contexts in Figure 4. The top graphs show series 1 and 2. For both cases, series 1 is to be predicted at the beginning and sudden changes occur frequently. While occurrence of swaps becomes more frequent over time in gradual drift, swaps occur cyclically and randomly in recurring contexts. In such cases, some inconsistencies of past experiences occur between the stable evaluator and the reactive evaluator as we show in the middle of the graphs. In some cases, the past experience works for a good forecast. If the past experiences work good, PEM tends to follow such experiences. In the bottom graphs, we show difference of performances of PEM and the stable evaluator. PEM tends to exceed the performance of the

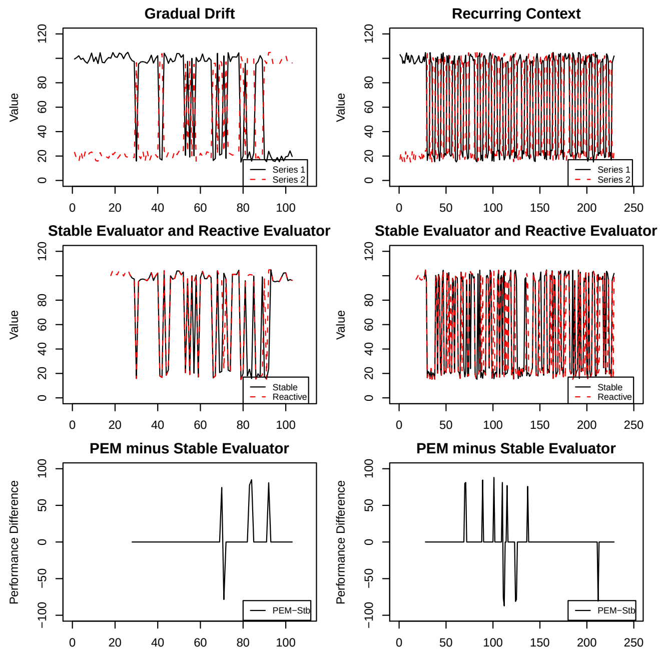

図 4: Gradual Drift and Recurring Contexts

stable evaluator if one evaluator works better than another in consecutive times.

In reality, it is difficult to know or identify drift patterns before occurrences. In the following section, we use actual financial data to show performances of PEM considering such difficulties.

5 Experimental Results

In this section, we show performances of our proposing forecasting method based on a back-testing using Fama-French financial data which is described in Section 2. For a benchmark purpose, we compare with two representative forecasting methods with dynamic forecast window size, *Competing Windows Algorithm (CWA)* and *FLORA*.

CWA adjusts the size of forecast windows according to characteristics of historical data proposed by Lazarescu et al. [13]. If a new observation is consistent over time, this algorithm uses larger windows size in order to increase forecast accuracy with an expectation of no concept drift occurrence. Otherwise, it uses a smaller window size to deal with concept drifts. In order not to focus too much on a particular window size, it uses three types of windows, small-medium-large, that are dynamically changing. For a forecasting, the most accurate forecast is used among the forecasts generated from the three windows. FLORA is a representative learning system that deals with recurring contexts with its dynamically changing window size proposed by Widmer and Kubat [18]. While every step FLORA observes a new data, it searches relevant historical data and classifies the searched data into positive data, negative data and both type data. Based on these data, FLORA generate forecast. According to results, FLORA updates its source of concepts: either addition of a new concept into the system or discard of the old concept.

First, we describe some configurations of parameters used for PEM and two well-known dynamic window size methods, CWA and FLORA. Then, we show experimental results.

With respect to PEM, we use 6 basic forecasts, 3-month top mode, 6-month, 9-month, 12-month, 18-month, 24-month average. 3-month top mode selects the most frequent factor that performs the best among three months. We use smoothed data for averages in which highly effective points exceeding 1.5 standard deviations are reduced. As for parameter of evaluators, we use 12-month equitable weights $w = \{1,1,1,1,1,1,1,1,1,1,1,1\}$ for a stable evaluator and 5-month decreasing weights $w = \{0.03125, 0.0625, 0.125, 0.25, 0.5\}$. With respect to coefficients, we set an update coefficient as 1, a decreasing coefficient as 0.005, a threshold as 0.

Regarding to CWA, there are three types of windows size, small-medium-large, with default window sizes, 3, 6 and 12-month, respectively. If a distance between the latest instance and the historical instances is within a consistent coefficient 3.88, which is 2 standard deviation of the sum of the absolute instances within the first 12 months, the medium window size is enlarged up to 11-month and the large window size is the double size of the medium window size. If consistencies are persistent more than 12 months, the large window corresponds to the size of persistence. Based on averages of three window sizes, the best factor is calculated. The type of the window size is selected according to the performance of the previous month.

With respect to FLORA, we tuned the algorithm in order to deal with our test data as follows. We classify instances into 24 states that consist of the best factor and the second best factor with signs. The state patterns are to allocate the following two { the best factor with a positive sign, the best factor with a negative sign } and { the second best factor with a positive sign, the second best factor with a negative sign } in the three positions that are $2 \times 2 \times 2 \times 3$. Once FLORA observes a new instance, it looks up the same state in the past. Among the matched state, it calculates the most frequent top factor in the following month which is used for a prediction in this month. We set default search periods as 36 months. If the matched state is less than a minimum number of match 5, it grows the number of windows up to 48 months. If the accuracy is greater than 50 % and the recent result is inaccurate, reduces window size 20 % where the minimum window size is 24 months. Otherwise, keep the same window size as the previous month.

Now, we show the experimental results. We depict the cumulative performances of the three models classified by decades in Figure 5. According to the experiment, PEM performs better than the other two models in most of time. An important aspect to evaluate performances is persistency of growth. In most time, PEM continuously performs well. While 1970s, the early 1980s, the early 1990s, and the early 2000s are relatively easy periods according to the average performance, PEM is quite stable. Even though the rest of periods are not easy, it performs good due switches of evaluators effectively during this periods. Surprisingly, growth during this period is quite remarkable that includes the financial crisis of 2008-2009. Key reasons are (i) PEM quickly adapt to the drift which is the reverse of momentum, and (ii) This works consecutively.

CWA also performs well similarly to PEM. However, in the middle of 1990s, it looses its control for a while. A disadvantage of CWA is a change of window size is one even though they have three sizes of windows. Hence, it may take time to search an appropriate window size. With respect to FLORA, performances are not good dynamically changing environments such as the middle of 1990s and 2000s. The performance of FLORA is good if there are many sample data in the past. However, in our data set, this is not always true. Similarly, paired learner in [1, 16] is not implemented in this experiment, since a reactive learner is not efficient with a small number of reference periods with our data.

6 Conclusions

In this paper, we have introduced our forecasting method, paired evaluators method, which tends to improve forecast performances for consecutive concept drift patterns, such as gradual drift and recurring contexts. In our method, a set of base forecasts is used for a prediction which is selected by evaluators. We use paired evaluators, a stable evaluator and a reactive evaluator. A selection of evaluator is based on learning algorithm which learns the past performances of evaluators. By learning, paired evaluators method continuously attempts to detect an alternative evaluator to improve forecast accuracy. This approach suits for consecutive concept drift patterns.

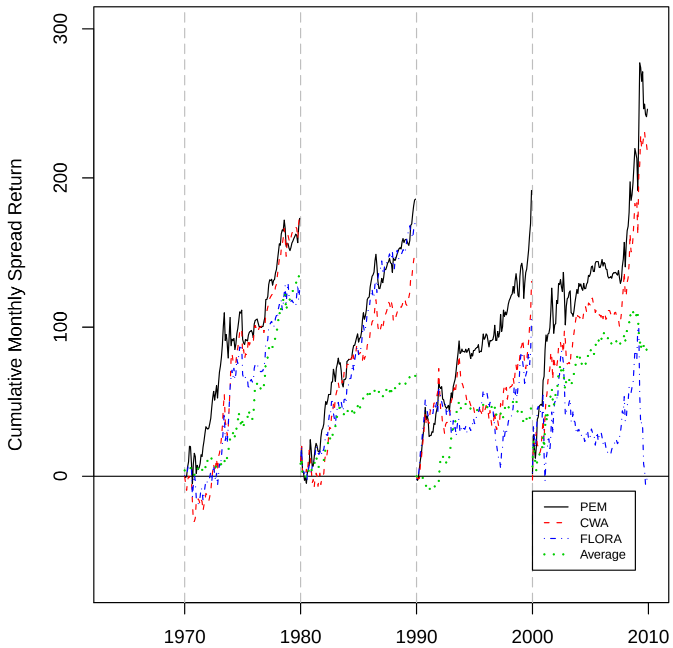

図 5: Back-testing Results of Cumulative Monthly 10 Percentile Spread Return of Portfolios Classified by Decades

We have introduced a methodology to deal with a financial investment problem, a factor selection problem, using concept drift solutions. Experimental results show that our proposing method has discovered effective factors more efficiently than the other two representative methods, CWA and FLORA, which change the forecast window size dynamically. Our method is robust against many difficult circumstances including the late-2000s recession.

In a broader sense, paired learners for online classifications, such as paired learners having two different window sizes for classifications based on naive Bayes approach [1] and Todi (two online classifiers system for learning and detecting concept drift) based on a statistic test [16], are similar to our approach. They use paired classifiers to control stability and reactivity for changes over time. While they directly set the window size on classifiers, which is fixed, in our approach, window sizes are set by respective base forecasts instead of evaluators. Hence, our approach uses multiple window sizes for a set of forecasts and a forecast having the most appropriate window size tends to be selected by evaluators. In our experiments, these paired learners for classifications are not used, since our experimental data does not have similar features in short periods.

In the financial investment problem, we have focused on a selection of the best factor. This is not a restriction in practice. In the future, we consider the method to set the most appropriate weights for the three factors.

参考文献

[1] S. Bach and M. Maloof, *Paired learners for concept drift*, In Proc. of the 8th IEEE Int. Conf. on Data Mining, pp. 23-32, IEEE 2008.

[2] A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavalda, *New ensemble methods for evolving data streams*, In KDD '09: Proc. of the 15th ACM SIGKDD int. conf. on Knowledge discovery and data mining, pp. 139-148, 2009.

[3] A. Bifet and R. Gavalda, *Learning from time-changing data with adaptive windowing*, In Proc. of SIAM Int. Conf. on Data Mining (SDM'07), pp. 443-448, SIAM, 2007.

[4] M. M. Carhart, *On persistence in mutual fund performance*, Journal of Finance, 52 (1), pp. 57-82, 1997.

[5] E. F. Fama and K. R. French, *The Cross-Section of Expected Stock Returns*, Journal of Finance, 47 (2), pp. 427-465, 1992.

[6] E. F. Fama and K. R. French, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics, 33 (1), pp. 3-56, 1993.

[7] K. R. French, *Fama/French Factors in U.S. Research Returns Data*, `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, accessed 23 July 2010.

[8] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, *Learning with drift detection*. In Advances In Artificial Intelligence, Proc. of the 17th Brazilian Symposium on Artificial Intelligence (SBIA 2004), volume 3171 of LNAI, pp. 286-295, Springer, 2004.

[9] R. Giacomini and B. Rossi, *Detecting and predicting forecast breakdowns*, Review of Economic Studies, 76 (2), pp.669-705, 2009.

[10] M. Harries and K. Horn, *Detecting concept drift in financial time series prediction using symbolic machine learning*, In Proc. of the 8th Australian joint conf. on artificial intelligence, pp-91-98, 1995.

[11] J. Kolter and M. Maloof, *Dynamic weighted majority: An ensemble method for drifting concepts* Journal of Machine Learning Research, 8, pp. 2755-2790, 2007.

[12] L. I. Kuncheva and I. Žliobaitė, *On the Window Size for Classification in Changing Environments*, Intelligent Data Analysis, 13 (6), pp. 861-872, 2009.

[13] M. M. Lazarescu, S. Venkatesh and H. H. Bui, *Using multiple windows to track concept drift*, Intelligent Data Analysis, 8 (1), pp. 29-59, 2004.

[14] J. Lintner, *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets*, Review of Economics and Statistics, 47 (1), pp. 13-37, 1965.

[15] J. Mossin, *Equilibrium in a Capital Asset Market*, Econometrica, 34 (4), pp. 768-783, 1966.

[16] K. Nisida and K. Yamauchi, *Learning and Detecting Concept Drift with Two Online Classifiers*, The 22nd Annual Conference of the Japanese Society for Artificial Intelligence, 3C2-1, 2008

[17] W. F. Sharpe, *Capital asset prices: A theory of market equilibrium under conditions of risk*, Journal of Finance, 19 (3), pp. 425-442, 1964.

[18] G. Widmer and M. Kubat, *Learning in the presence of concept drift and hidden contexts*, Machine Learning, 23 (1), pp. 69-101, 1996.

[19] I. Žliobaitė, *Learning under Concept Drift: an Overview*, Vilnius University, Faculty of Mathematics and Informatics, Technical Report, 2009.