FX Trading using Logistic Regression Analysis and Naive Bayes Model

Kimihisa KAWABATA[1] and Hitoshi TAKATA[2]

[1]Graduate School of Economics and Business Administration, Kyushu Sangyo University JAPAN
[2]Graduate School of Science and Engineering, Kagoshima University, JAPAN

1 Introduction

In FX trading, judgment whether it trades with the next term is very difficult. The existing technical analysis depends on the past statistical data. Especially, in the index of an oscillator system, there are much wrong indexes such as divergence, falsehood, time lag and so forth and judgment is very difficult. The propriety (probability p) of whether to trade with the next term will be judged according to the posterior probability of Bayes using the basic formula of Bayes in this paper.
The binomial distribution is adapted as likelihood of whether trade is carried out or not and a beta distribution is adapted as a prior distribution. Then posterior probability distributions also turn into a beta distribution, since the natural conjugate distribution is a beta distribution. However, about the initial value of prior probability, the output value (probability p) of logistic regression analysis is used not using subjective probability (prior probability).
Moreover, economic news are incorporated from the web site for forecasting calculation, and the expected value of the probability of the propriety of a trade was modified. As a result, performance has been improved greatly.

2 FX Trading and Decision-making

2.1 The Problem of a Technical Analysis and Difficulty of a Trade

A technical analysis tends to be available to forecast next trading subjectively incorporating the price movement of FX currency pair which is essentially non-stationary stochastic process by the elementary statistical technique. Bollinger band, RSI, MACD, etc. are the techniques of being popular among a trader. However, as we know, "divergence", "falsehood", "time-lag", and so forth exist in these indexes largely, and a trade goes fail in many cases. This is the big fault of a technical analysis. One of these examples in USD/JPY are shown in Fig.1

Fig.1 RSI and Divergence

Source: Gaitame Dot Com, Chart & Tool
https://trade.gaitame.com/members/index.asp 2011.09.30.

2.2 The Possibility of a Trade and Bayes Statistics

However, in this paper, we deal with only an entry in order to understand easily. The technique of "Bayes statistics" is introduced as such a mathematical model. A beta distribution is used for calculating the prior probability of Bayes. Hence, the probability (p) of the possibility of a trade shall follow a beta distribution. Probability density function of a beta distribution here is written in the next equation.

$$\pi(p)=\frac{1}{B(\alpha,\beta)}p^{\alpha-1}\cdot(1-p)^{\beta-1} \;;0<p<1,\ \alpha>0,\ \beta>0 \qquad (1)$$

where $B(\alpha,\beta)=\int_0^1 p^{\alpha-1}\cdot(1-p)^{\beta-1}dp$: beta function

p: the trading probability

Here, beta function can also write as follows using a gamma function.

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

where $\Gamma(\alpha)=\int_0^\infty e^{-t}\cdot t^{\alpha-1}dt$:gamma function (2)

As to likelihood, we adapt the binomial distribution so as to carry out the trade or not. The probability density function of binomial distribution is expressed with the next equation (3).

$$f(n,p)=\binom{n}{r}p^r(1-p)^{n-r} \qquad (3)$$

where r: trade days (day which traded)

If a beta distribution is adopted as a prior distribution to the likelihood obtained from the data (D) which follows a binomial distribution, posterior distribution will turn into a beta distribution which equal to a prior distribution. Because, if we adapt the natural conjugate distribution as prior distribution (beta distribution) against a likelihood (binomial distribution), it is known that posterior distribution will also turn into the same beta distribution as a prior distribution.
When this relation is used, the basic formula of Bayes is expressed with the next equation.

$$\pi(P_{n+1}:D_n)=\frac{f(D_n:p_n)\cdot\pi(p_n)}{f(D_n)}$$

$$=\frac{f(D_n:p_n)\cdot\pi(p_n)}{\int f(D_n:p_n)\pi(p_n)dp_n}$$

where p_{n+1} : trading probability at time n+1

D_n : the actual result data of the trade at time n

A denominator is a very complicated equation and it is calculated by the approximated calculation here. However, this is a constant, and since a trade is made decisions using natural conjugate distribution, even if a denominator is removed, it does not lose generality.
Consequently, the following is realized.

$$\pi(p_{n+1}:D_n)\propto f(D_n:p_n)\cdot\pi(p_n) \qquad (4)$$

Eq. (4) is the basic formula of Bayes and posterior probability of the possibility of a trade is simply acquired as a product of the likelihood and prior probability. However, like common knowledge, prior probability is subjectivity probability and cannot eliminate subjectivity like selection of the technique of a technical analysis. Then, with this paper, this prior probability is obtained from p; 0<p<1 which was got from the output result of a logit model. p_n denotes probability of the last term (the n-th term) of the analytical period of a logit model. For details, the next Section 2.3 describes. In addition, the reason using a beta distribution is as follows. The form of a beta distribution is various according to parameter α and β, and it is similar in part to the form of the exchange rate chart of currency pair USD/JPY which takes various form.

2.3 A Naive Bayes Model and Logit Transformation

Prior probability $\pi(p_n)$ was calculated mechanically as the function of p_t using logit transformation value z_t which is computed as follows.

$$z_t=logit(p_t)=ln\left(\frac{p_t}{1-p_t}\right)=\alpha_0+\alpha_1x_{1t}+\alpha_2x_{2t}+\cdots+\alpha_px_{pt} \;;t=1,2,\cdots,n \qquad (5)$$

Namely, parameter $\alpha_0, \alpha_1, \cdots, \alpha_p$ of eq. (5) will be computed by using the maximum likelihood method. Subsequently, next equation is obtained by solving eq.(5) about p_t.

$$p_t=\frac{1}{1+e^{-z_t}} \;;t=1,2,\cdots,n \qquad (6)$$

$$E[p_n]\equiv p_n \qquad (7)$$

Combination of a logit model and a naive Bayes model is described in Section 3.2 in detail. The objective variable of a logit model $z_t = logit(p_t)$ is composed of 2-value variable (say, 1: trade is carried out 0: trade does not be carried out). Although various variables could be considered as explanatory variable, finally the following variables x_{1t}, x_{2t}, ・・・, x_{nt} were chosen. x_{8t}, x_{9t}, x_{10t}, and x_{11t} are dummy variables of the economic news at time t . The discriminant hitting ratio (η) of whether it trades or not to trade is required from a discriminant crossing table. The criteria of discriminant is decided politically (for example, 65% or more).

2.4 Precondition of Decision-Making

Although the style of a trade has scalping, one day trade, swing trade, middle-term trade, and long-term trade first, swing trade is dealt with in this paper. However, data is a closing price of USD/JPY. An analytical period is carried out from April 1, 2011 to August 31, 2011. Subsequently, let a forecasting period be 12 business days from September 1, 2011 to September 16, 2011. The objective variable y of a logit model is calculated by the rule of a trade. It will be referred to as "up-trend" if three exchange rate charts (4-leg chart) of USD/JPY continue by positive line. Conversely, it will be referred to as "down-trend" if three exchange rate charts (4-leg chart) of USD/JPY continue by the negative line. When it is not any, it will be referred “0” as “un-change”. If the market is either “up-trend” or “down-trend”, then “1” will be given. Conversely, the market is “un-change”, then “0” will be given. It corresponds to the meaning that “1” is “it trades” and “0” is “it does not trade”. Economic news has received from the web site of http://forexpros/jp/economic-calendar 2011.10.3 .

Subsequently, although we consider a swing trade, a swap point shall not be considered here. Moreover, neither a commission nor a tax shall also be considered. The parameters a and b of beta distribution of posterior distribution become $\alpha+r, \beta+n-r$ respectively by the relation that the "natural conjugate distribution" is beta distribution against likelihood. Then, using not the MCMC method (Markov Chain Monte Carlo method) but "a natural conjugate distribution" , the expected value of a posterior distribution (beta distribution) can be simply computed by the next equation.

$$E[p] = \frac{a}{a+b} \qquad (8)$$

Calculation of statistic can be simply performed by avoiding the complicated calculation including the integration operation of posterior distribution, and using "natural conjugate distribution".

3 Trading of USD/JPY

3.1 The Recent Trend of USD/JPY

Let’s look at the recent trend of USD/JPY (Daily) in the period on April 1, 2011 to August 31, 2011. Analysis period shows five phases divided by the peak and the bottom.

3.2 Combination of Logit Model and a Naive Bayes Model

A naive Bayes model is used for forecasting of the possibility of FX trading. The basic formula of a naive Bayes model was expressed by eq. (4). Though, here, the right-hand side $\pi(p_n)$ is subjective probability, we defined this the last probability (p_n) of n-th term of a logit model. Hence, the next equation is assumed.

$$\Pi(p_n) \stackrel{\text{def}}{=} p_n \text{ : the last output value of a logit model} \qquad (9)$$

The discriminant equation for which we are requiring can be written such as eq. (10).

$$logit(p) = -0.06e + 0.067SPA/J - 0.212Gold - 0.280BondA10y - 0.163BondA5y + 0.070H\text{-}L$$
$$-0.41e1 - 0.324D1 + 0.0108D2 - 0.007D3 + 0.034D4$$

Maximum Likelihood method $\eta=67.6\%$ *AIC=150.56* *periods:2011.4.1-2011.8.31* (10)

The discriminant hitting ratio (η) is computed as following.

$$\eta = \frac{25+46}{105} * 100 \cong 67.6\,(\%) \qquad (11)$$

The probability of last term (business day on August 31, 2011) of a logit model is calculated with 21.23%.

$$E[p_n] \equiv p_n = \frac{1}{1+e^{-z_n}} \cong 0.2123$$

where $n=105$: the number of business days (12)

An analysis result also shows further that the trade days r become as the following value.

$$r = 44 \qquad (13)$$

Subsequently, we will roughly describe about the process after combination. A posterior distribution becomes beta distribution $\beta(\alpha+r, \beta+n-r)$. Hence, it can write as follows.

$$\beta(a,b) = \beta(\alpha+r, \beta+n-r) \qquad (14)$$

The calculation of expected probability $E[p]$ is obtained by the next equation.

$$E[p] = \frac{a}{a+b}$$

where $a=\alpha+r$

$b=\beta+n-r$ (15)

What is necessary is just to forecast "it trades" with more than 0.5 on $E[p]$. On the contrary, what is necessary is just to forecast "it does not trade" with less than 0.5 on $E[p]$. The necessary matter is just to carry out the above forecasting calculation during some periods (for example, 12 business days on September 1, 2011 to September 16, 2011) .

3.3 Forecasting of Trade of USD/JPY

We carried out the decision making of trade considering economic news in order to improve $E[p]$ to $E[p]_{rev}$. Thus, the revised $E[p]_{rev}$ was obtained re-calculating $E[p]$. The obtained forecasting result is in Fig.2.

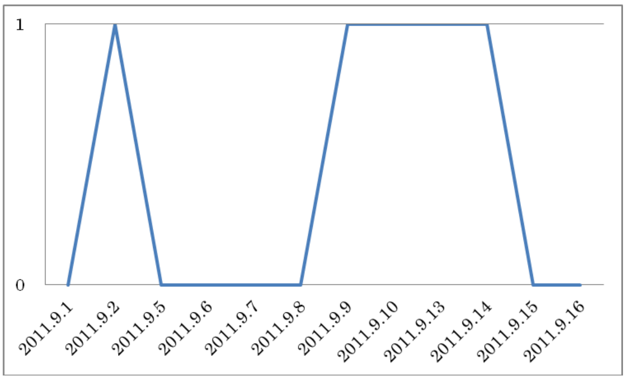

Fig.2 Forecasting Value of Trade (2011. 9.1-2011.9.16)

The performance of the "actual value" which became clear later on, and the "forecasting value" of a naive Bayes model was eight wins and four losses among 12 business days. Hence, the rate of coincidence of forecasting became 66.67%.

4 Conclusion

In this paper, forecasting of propriety of a trade of USD/JPY was treated in a fashion of mathematical science. Hence, the propriety (probability) of whether to trade with the next term was calculated by the naive Bayes model. However, an initial value of the prior probability distribution (beta distribution) of naive Bayes model was not treated subjectively but adapted the probability which is the last output value of logistic regression analysis. Hence, the following idea was introduced into the naive Bayes model. When we take the prior distribution of the probability which trades as beta distribution which is “a natural conjugate distribution “ against likelihood obtained from the data of whether to carry out the trade , which follows binomial distribution, a posterior distribution is also a beta distribution. When re-calculating of the parameters n, r, α and β and substituting them into the naive Bayes model, a posterior probability of a beta distribution could be computed sequentially (renewal of Bayes).This calculated value served as decision-making (forecasting value) of whether to trade after the next term.

However, in the process of forecasting, the value of $E[p]$ stopped at the low value, and was obliged to correction. Then, revised expected probability $E[p]_{rev}$ which incorporated the economic news announced in advance, like logistic regression analysis, was re-calculated. The result approached considerably the actual value, as became clear later on, and the performance of a trade improved to about 66.67%.

References

[1] Congdon P.: Bayesian Statistical Modelling, John Wiley & Sons, 2001.

[2] Lien, K: Day Trading of the Currency Market: Technical and Fundamental Strategies to Profit from Market Swings, John Wiley & Sons, 2006.

[3] Esumi: " Excel Tahenryo Kaiseki Ver.5.0 User's Manual," ESUMI, 2003. (in Japanese)

[4] Uchida, O. and Fukushima, R.: Reikai Tahenryo Kaiseki Gaido: Excel Adoin Sohuto wo Riyou site, Tokyo Tosho Co., Ltd., April, 2011. (in Japanese)

[5] Wakui, Y.: Dogu toshiteno Bayes Toukei, Nippon Jitsugyo Publishing Co., November, 2009. (in Japanese)

[6] Gaitame Dot Com : https://trade.gaitame.com/members/applet/main.asp, September 19, 2008. (in Japanese)