Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

A Self-Impact Analysis by Artificial Market Simulation

Kiyoshi Izumi
Digital Human Research Center, AIST
Tokyo, Japan
Email: kiyoshi@ni.aist.go.jp

Hiroki Matsui
Information Technology Research Institute, AIST
Ibaraki, Japan
Email: hiroki.matsui@aist.go.jp

Yutaka Matsuo
Information Technology Research Institute, AIST
Tokyo, Japan
Email: y.mastuo@aist.go.jp

***Abstract*— We constructed an evaluation system of the self-impact in a financial market using an artificial market and text-mining technology. Economic trends were first extracted from text data circulating in the real world. Then, the trends were inputted into the market simulation. Our simulation revealed that an operation by intervention could reduce over 70% of rate fluctuation in 1995. By the simulation results, the system was able to help for its user to find the exchange policy which can stabilize the yen-dollar rate.**

I. Introduction

When determining behavior in a financial market, there are two difficult problems. One is understanding the present economic trend and another is evaluating appropriately the influence of own behavior on a market. Behavior of a big player like the decision-maker of a central bank may affect other market participants' behavior, and change a trend of a market. The pattern of such feedbacks changes with the economic conditions of the time. Therefore, evaluation of self-impact and presumption of an economic trend are closely related. By integration of an artificial market and text-mining technology, we propose the approach of performing them simultaneously.

In recent years, artificial market studies have obtained some success in market analysis. Especially they could reveal new mechanisms of market phenomena such as financial bubbles that previous models could not explain well. And they were used for tests of existing economic theories such as the efficient market hypothesis. Many of them were however too simple and abstract to be a useful model of actual markets. One major purpose of multi-agent simulation of social phenomena is to build models of actual social phenomena on computers, and to use them for supporting actual action. For that purpose, a model must reflect the real world so that its simulation results are reliable.

II. Framework of SEMAS

This paper proposes a new approach SEMAS (Socially Embedded Multi Agent based Simulation). This approach integrates an artificial market simulation and text mining, and takes real information into the simulation. And it can support its user's decision in an actual market. Furthermore, the user's action can affect the real world and the model can reflect the changes of economic trends. In the SEMAS approach, an artificial market simulation is embedded into such a feedback loop.

The SEMAS approach consists of the two stages in Fig. 1.

1) *Extraction of the information from the real world*: First, economic trends are extracted from various text data such as news articles and webs sites using text-mining technique.
2) *Support of users' action to the real world*: Next, the extracted trend is inputted into a multi-agent model of a foreign exchange market, and its computer simulation is carried out. Based on the simulation results, our proposed system can support a user in determining his/her behavior in the market.

"Socially embedded" in SEMAS means that this approach can treat the simulation with real information and the evaluation of feedback of user's action to the real world. Based on this approach, A social simulation can be embedded into the feedback loop with the extraction of social information, the action selection by a simulation, and the action's effects to the real world.

In this paper, the purpose of SEMAS is to find the best action rule which stabilizes yen-dollar rate in 1995 using data from 1992 to 1994, for decision makers of an exchange policy.

III. Extraction of the information from the real world

The procedure of the economic trend analysis as the first stage of SEMAS is shown in Fig. 2. A feature vector is first calculated from text data related to this week's economic conditions (Fig. 2a-d), and economic trends are estimated by a statistical classification method C4.5 using the feature vector (Fig. 2e).

The original text data are market reports which JCIF (Japan Center for International Finance, `http://www.jcif.or.jp/`) distributed every week to its members such as professional dealers. Each document reports economic news relevant to trends of financial markets of each week and is written in Japanese of about 1,000 characters. In this paper we used the JCIF documents from 1992 to 1994 (156 weeks).

1-4244-0705-2/07/$20.00 ©2007 IEEE

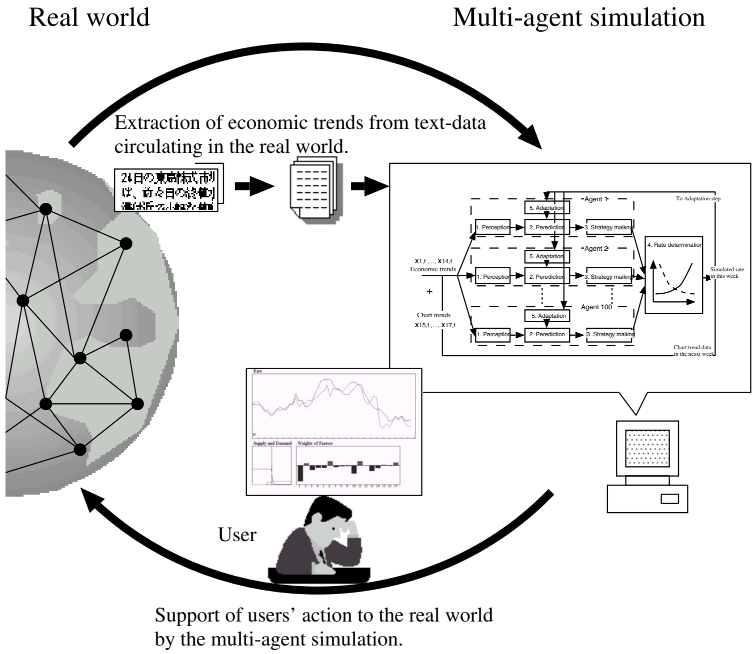


Fig. 1. Framework of SEMAS

A. *Extraction of each week's feature*

As preprocessing of the extraction, morphological analysis, filtering, and classification are carried out. Each sentence is first divided into morphemes, the smallest meaningful units, using Japanese language morphological analysis system ChaSen [1] (Fig. 2b). Each morpheme has information about its basic form and part-of-speech from the morphological analysis. Next, each morpheme is checked against a list of keywords as shown in Table I. The list consists of pairs of keywords and morphemes, which relate to economic trends. There are 183 keywords containing nouns like "yen", verbs like "rise", and adjectives or adverbs like "large". Morphemes which do not appear in the list are deleted. And morphemes with the same meaning are classified into the same keyword according to the list[1] (Fig. 2c).

After the preprocessing, a feature vector of each week is calculated using tf-idf (term frequency inverse document frequency) values based on frequencies of appearance of the 143 keywords[2]. A tf-idf value of keyword k in week t is as follows.

$$\mathrm{tfidf}_{k,t} = \mathrm{tf}_{k,t} \times \log\left(\frac{N}{\mathrm{df}_k}\right), \tag{1}$$

where $\mathrm{tf}_{k,t}$ is (a frequency of keyword k in week t) / (a total of frequencies of the 143 keywords in week t). $N(=156)$ is a number of all weeks, and df_k is a number of the weeks when the keyword k appeared at least once. About a keyword that appears many times only in this document, its tf-idf value becomes large. A feature vector of week t is as follows.

$$\mathbf{D}_t = (\mathrm{tfidf}_{1,t}, \mathrm{tfidf}_{2,t}, \cdots, \mathrm{tfidf}_{183,t}). \tag{2}$$

[1] Our filtering method is based on [2], [3].

[2] See [4] as another example of economy news analysis with tf-idf values.

TABLE I

EXAMPLE OF A KEYWORD LIST

Keywords	Morpheme 1,	Morpheme 2,	···
yen	en (yen) !noun-suffix-numerative	– – –	
rise	joushou, (go up) –	kyuushin, (jump) –	···
large	ookii, (large) –	tairyou, (massive) –	···

The part-of-speech condition "!noun-suffix-numerative" means that a morpheme is not a numerative noun suffix like "134-*yen*".

B. *Estimation of economic trends*

As shown in Fig. 2e, the feature vector $\mathbf{D}_t$ is inputted into decision trees, and economic trends of week t are estimated with respect to 14 kinds of categories in Table II. These categories follow the classification in many books which explains financial markets. The decision trees were determined by J4.8. J4.8 is implementation of C4.5 algorithm in data-mining software WEKA [5], [6]. using the feature vectors and training data. The training data were coding values with seven levels $\{0, \pm1, \pm2, \pm3\}$ that we classified each document into

(a) Original text data

Document A (04 Oct. 1993 - 08 Oct. 1993)

Sentence a

「ドル買いが優勢で円は一時 134 円 38 銭まで下落した。」

"Dorugaigayuuseideenhaichiji134en38senmadegerakushita."

(The dollar buying was superior and yen fell to 134 yen 38 sen temporarily.)

Sentence b

「その後も日銀総裁が · · ·」

"Sonogomonichiginsousaiga· · ·"

(Also after that, the Governor of the Bank of Japan · · ·.)

Sentence c

⋮

⇓

(b) Morphological analysis

Sentence a'

「/ドル/買い/が/優勢/で/円/は/一時/134/円/38/銭/まで/下落/した/./」

"/Doru/gai/ga/yuusei/de/en/ha/ichiji/134/yen/38/sen/made/geraku/shita/./"

(dollar)(buying)( )(superior)( )(yen)( )(temporarily)(134)(yen)(38)(sen)(to)(fell)( )

Sentence b'

「/その/後/も/日銀/総裁/が/· · ·」

"/Sono/go/mo/nichigin/sousai/ga/· · ·"

(that)(after)(also)(Bank of Japan)(Governor)( )

Sentence c'

⋮

⇓

(c) Filtering and classification

Sentence a"

"/dollar/buy/ -/strong/ -/yen/ -/temporary/-/ -/ -/ -/ -/ down/ -/ -/ "

Sentence b"

"/-/-/-/central_bank/-/-/· · ·"

Sentence c"

⋮

⇓

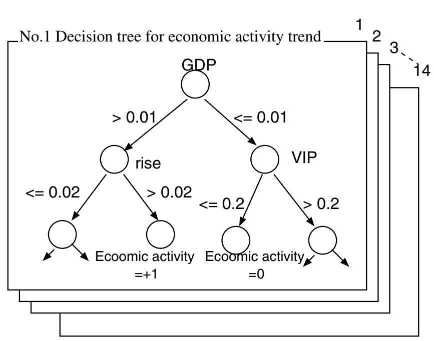
(d) Calculation of word frequencies (tf-idf value)

Document A (04 Oct. 1993 - 08 Oct. 1993): Sentence a,b,c,· · ·

{dollar 0.041978}, {strong 0.018933}, {yen 0.081418}, · · ·

⇓

(e) Estimation of economic trends using decision trees about 14 economic factors.



➡

Document A (04 Oct. 1993 - 08 Oct. 1993)

1. Economic acitivity	+1
2. Price	0
3. Interst rate	+3
4. Money supply	0
5. Trade balance	-2
6. Employment	-1
7. Personal comsumption	0
8. Intervention	+1
9. VIP announcement	-3
10. European trend	0
11. Goods market	+1
12. Political condition	+2
13. Stock market	0
14. Bond market	+1

Fig. 2. Procedure of text-mining

with respect to the 14 categories. A positive (negative) value of the coding data means a trend for stronger (weaker) yen.

We tested the decision trees acquired by J4.8 using percentages of correctly estimated documents. As shown in Table II, the decision trees correctly estimated economic trends of 92.2% of the 156 documents in estimation of the training data, and 71.9 % in 10-fold cross-validation test[3]. The percentages reached high levels enough for practical use, although there were differences among the categories. The multi-agent model in the SEMAS approach corresponds to the real world by using the estimated economic trends.

A part of decision tree about the Price category is shown in Fig. 3 as an example of the actually obtained decision tree. When the tf-idf value of keyword *price* was low, the decision tree interprets that the situation of the prices category of the week especially had no change (the value of the price category is zero). When the tf-idf values of keywords *price* and*rapid* (e.g. "Consumer price index rapidly changed $\cdots$") is high or those of *price* and *attitude* is high (e.g. "The attitude about prices $\cdots$"), the decision tree places non-zero value on the price trend of the week. Thus, it can be thought that the obtained decision trees make reasonable judgments.

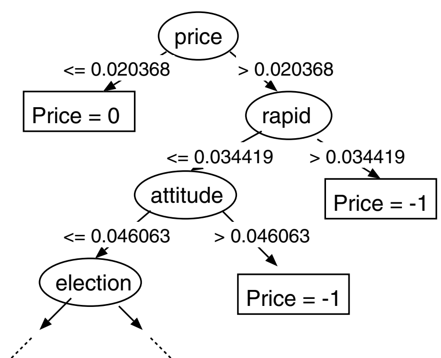


Fig. 3. A decision tree about Price trend

IV. MULTI-AGENT MODEL OF FINANCIAL MARKET

As the second stage of SEMAS, a multi-agent simulation is performed using the economic trends extracted from the real world in the first stage. This section describes a multi-agent model of a foreign exchange-rate market, AGEDASI TOF (A GEnetic-algorithmic Double Auction Simulation in TOkyo Foreign exchange market). AGEDASI TOF is an artificial market with 100 agents, as illustrated in Fig. 4. Each agent is a virtual dealer which has dollar and yen assets and changes positions in the currencies for the purpose of making profits. Each week of the model consists of the following five steps.

A. Step 1: Perception

Each agent first receives coding data x_i of the 14 economic trends and 3 chart trends. The data of the 14 economics trends are produced from the feature vector of text data $\mathbf{D}_t$ and the decision tree functions $F_i(\cdot)$.

$$x_{i,t} = F_i(\mathbf{D}_t), \quad i = 1, \cdots, 14. \tag{3}$$

The 3 chart trend data are a *short-term trend* (a change of the exchange rate in the last week), a *change of short-term trend*, and a *long-term trend* (change through five weeks). These data are coded into 7 levels from -3 to $+3$ like the 14 economic trend data, by normalization with their standard deviations.

B. Step 2: Prediction

Each agent assigns its own weights, $w_i = \{\ 0, \pm0.1, \pm0.5, \pm1.0, \pm3.0\}$ to the 17 trend data and uses the weights to predict the rate fluctuation for the next week. The mean of its forecast, M, is the weighted average of x_i as follows.

$$M = trunc\left(\sum_i w_i x_i\right) \times 0.02, \tag{4}$$

where $trunc(\cdot)$ is a truncation function and 0.02 is a scaling coefficient. It was calculated from a ratio of the average of $\Sigma w_i x_i$ and rate changes.

The variance of its forecasts, V, is calculated from the difference between the stronger-yen factors and the weaker-yen factors, as follows

$$V = \frac{1}{\sqrt{\left(\sum_{w_i x_i > 0} w_i x_i\right)^2 - \left(\sum_{w_j x_j < 0} w_j x_j\right)^2}} \tag{5}$$

The variance is inversely proportional to the coherence of forecast factors. Hence, the larger the variance, the lower the confidence of the forecast.

C. Step 3: Strategy Making

Each agent has dollar assets and yen assets. On the basis of its prediction, it determines its trading strategy (to buy or sell dollars) in order to maximize its utility function $U(q)$ of its position of dollar assets q as follows.

$$U(q) = M \times q - \frac{1}{2} V \times q^2 \tag{6}$$

The first term of equation 6 is the expected return and the second term is the risk (variance) of the position. Therefore, each agent tries to increase returns and reduce risk. The optimal amount of the dollar assets q^* is given by.

$$q^* = \frac{M}{V} \tag{7}$$

If each agent's current position q_t is lower (higher) than its optimal position q^*, it will want to buy (sell) dollars. Then it places an order to buy (sell) when the rate is lower (higher) than its forecast M.

[3]In this test, 9/10 of data is used for learning and 1/10 is for the test.

Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)

TABLE II

EVALUATION OF DECISION TREES (NUMBER OF DOCUMENTS N= 156).

	Categories	% of correct classification Training data	% of correct classification 10-fold cross-validation
1	Economic activity	88.9	62.9
2	Price	94.1	75.9
3	Interest rate	81.1	57.1
4	Money supply	97.4	97.4
5	Trade balance	96.7	85.0
6	Employment	90.9	66.2
7	Personal consumption	91.5	90.9
8	Intervention	95.4	66.8
9	VIP announcement	87.6	45.4
10	European trend	88.3	41.5
11	Goods market	99.3	99.3
12	Political condition	92.2	69.4
13	Stock market	93.5	75.3
14	Bond market	94.1	74.0
	Average	92.2	71.9

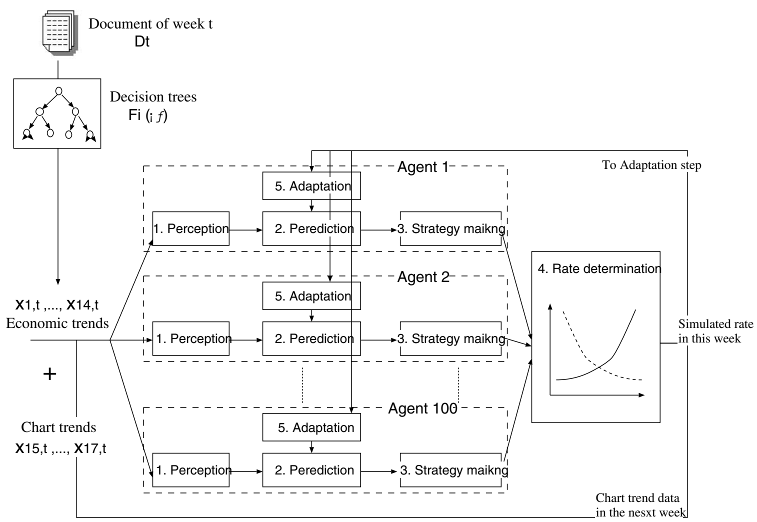


Fig. 4. Framework of multi-agent model

D. Step 4: Rate Determination

The market-clearing rate of the artificial market is the equilibrium rate where the quantity of demand and the quantity of supply are equal. Agents who submit orders to buy at rates above the market-clearing rate and agents who submit orders to sell at rates below the market-clearing rate can exchange their assets for the purpose of optimizing assets. Assets of other agents remain the same as the previous week.

E. Step 5: Adaptation

Each agent improves its prediction method by referring to the prediction methods of other agents. The change of the prediction methods in the market is described by the following three operations of simple GA [7]. An individual in GA is a string of all 17 weights of one agent $\{w_i\}$ and a fitness value of each individual is calculated using a forecast error, a difference between the mean of its forecast M and the market-clearing rate.

Selection A certain ratio G of agents replace their prediction methods by others' prediction methods with higher fitness on the probability proportional to the fitness values. The selection operator is interpreted economically as the propagation of successful opinions about forecast factors.

Crossover A pair of agents sometimes exchanges parts of their weights. The crossover operator works like the agent's communication with other agents.

Mutation Each agent has a small probability of changing each weight. This mutation operator works like the independent change of each agent's prediction method.

The crossover and mutation operators produce new combinations of weights. The crossover (mutation) occurred at a certain probability *pcross* (*pmut*). We used the following parameter sets: $pcross = 0.3, pmut = 0.003, G = 0.8$. These values were determined from forecast tests [8]. After the *Adaptation* Step, our model proceeds to the next week's *Perception* Step.

V. Support of users' action to the real world

Our artificial market model can support its users such as central bank staffs, policy makers, and authorities in deciding exchange rate policies. In this section our model is used for users to estimate the impact of users' own action to the market. The users try to find the best IF-THEN rule of their action for the stabilization of yen-dollar rate in 1995 using data from 1992 to 1994. The condition part of the rule consists of economic trends and/or chart trends. The action part of the rule is how to operate three *control factors* (interest rates, intervention, and announcement) during 1995.

First, based on simulation results using data from 1992 to 1994, the users select several *condition factors* from the 17 factors. The condition factors are the factors to which many agents attached large weights. The users will operate the control factors according to the trends of the condition factors. Second, seven candidates of the users' action were prepared according to the kind of the control factors to operate. These candidates were estimated by standard deviations of rate changes of the simulation in 1994, and the best action option was suggested.

A. Selection of condition factors

First, we trained the 100 agents in the artificial market using sample data from 1992 to 1994 by the following procedure.

1) *Initialization*: The initial population included a hundred agents whose weights were generated randomly and the position of each agent is square.
2) *Training*: We trained our model using data of the 17 factors and the actual rates from 1992 to 1994. During the training we skipped the rate determination step, and in the adaptation step, the fitness in GA was the cumulative value of differences between the forecast means of each agent and the actual rates.

As a result, *bond market*, *money supply*, *European trend*, *rate change in one week*, and *change of rate change* factors got the biggest weights from the agents on the average of the 100 simulation runs. That is, the artificial market was sensitive to these five condition factors in this period. Therefore, if there is a large change of the five factors' trends, the artificial market tends to become unstable.

B. Decision of the best action option

Second, seven candidates of the users' action (a)-(g) were prepared according to the kind of control factor to operate (Fig. 5). Each exchange policy options was input into the artificial market simulation and results were compared. We generated 100 simulation paths for each option by repeating the following procedure a hundred times.

1) *Initialization*: This is the same as mentioned above.

IF
there is a change of the condition factors' trends ($|x_{4,t} + x_{12,t} + x_{14,t} + x_{15,t} + x_{16,t}| > 0$),

THEN
(a) operate a value of the *interest rate* factor $x_{3,t}$ as follows,

$$x_{3,t} = \begin{cases} 3 & (x_{4,t} + x_{12,t} + x_{14,t} + x_{15,t} + x_{16,t} < 0) \\ -3 & (x_{4,t} + x_{12,t} + x_{14,t} + x_{15,t} + x_{16,t} > 0) \\ 0 & (x_{4,t} + x_{12,t} + x_{14,t} + x_{15,t} + x_{16,t} = 0) \end{cases} \quad (8)$$

(b) operate a value of the *intervention* factor $x_{8,t}$ factor similarly;

(c) operate a value of the *announcement* factor $x_{9,t}$ similarly;

(d) operate values of the *interest rate* and *intervention* factor $x_{3,t}$ and $x_{8,t}$ similarly;

(e) operate values of the *interest rate* and *announcement* factor $x_{3,t}$ and $x_{9,t}$ similarly;

(f) operate values of the *intervention* and *announcement* factor $x_{8,t}$ and $x_{9,t}$ similarly;

(g) operate values of *interest rate*, *intervention*, and *announcement* factor $x_{3,t}$, $x_{8,t}$, and $x_{9,t}$ similarly.

Fig. 5. Policy options

2) *Training*: This is the same as mentioned above except having used the data from 1992 to 1993 this time.
3) *Simulation*: We conducted the extrapolation simulations in 1994. In the test period, our model forecasted the rates in the rate determination step using only the 14 economic trend data. Each option was used as input data. The 3 chart trend data (a *short-term trend*, a *change of short-term trend*, and a *long-term trend*) and the fitness in GA were calculated from the simulated rates in the rate determination step. Each option is evaluated using a standard deviation of rate change in 1994 on the average of 100 simulation runs.

As a result, the action based on the option b (*intervention* operation) could reduce the rate fluctuation the most (Fig. 6). The standard deviation of the simulated rate with this operation was 0.02310 on the average of 100 runs, and it reduced about 40% of the standard deviation of the actual rate in 1994 (0.03793). The operations including intervention (the options b, d, and f) could reduce over 30% of rate fluctuation. However, the effect of operation by all the three factors (the option g) was not so large, because its effect was too large.

The above results showed that the following exchange policy option was effective for the stabilization of the yen-

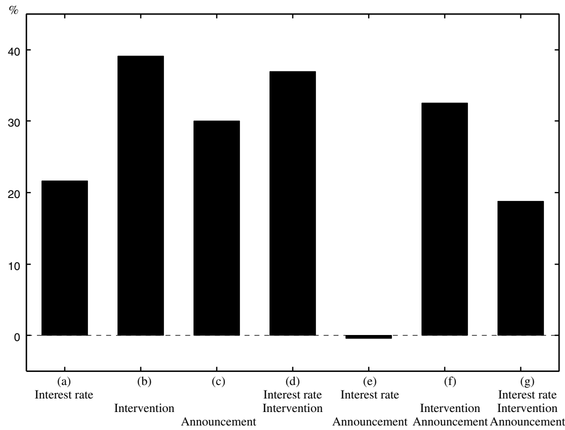


Fig. 6. The reduction rates of rate changes.

dollar rate.

IF
there is a change of the trends of *bond market*, *money supply*, *European trend*, *rate change in one week*, and *change of rate change* factors,
THEN
operate a value of the *intervention* factor against the trend.

VI. Test of the Best Option

We tested whether the best option obtained in the simulation using the data by 1994 could stabilize the rate in 1995. The agent was initialized like the above-mentioned and the training period was four years from 1992 to 1994. The extrapolation simulation in 1995 was performed 100 times using the weights acquired by training.

1) *Initialization*: This is the same as mentioned above.
2) *Training*: This is the same as mentioned above except having used the data from 1992 to 1994 this time.
3) *Simulation*: This is the same as mentioned above, but the extrapolation simulation is about rates in 1995. In the simulation in 1995, the trend data of the control factors (interest rate, intervention, and announcement) were provided in two ways; the best action option and the actual action.

A solid line in Fig. 7 is the average path of 100 simulation results with the operation by intervention. A dashed line is the average path of 100 simulation results using the trend data that are hand-coded from the JCIF document about interest rate, intervention, and announcement.

Although the width of change in the simulation path (the dashed line) with the actual action was smaller than the actual path (the dotted line), the whole tendencies such as the direction of change were common. Compared with the simulation path with the actual action (the dashed line), in the simulation path with the best option (the solid line) was able to reduce the rate fluctuation in 1995 over 70% on the average of 100 runs. As compared with the actual path, the rate change was decreased no less than 80%. These results showed that the best action option obtained in the simulation was effective in the exchange rate stabilization in 1995.

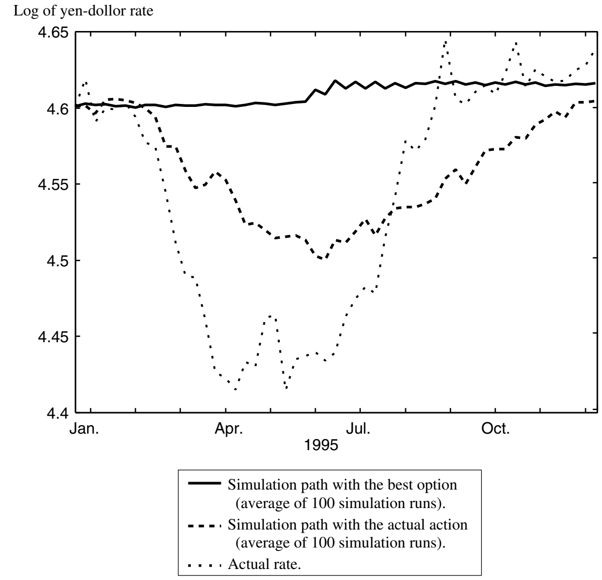


Fig. 7. Comparison between simulated paths with the best option and those with the actual action.

VII. Conclusion

The results of this paper showed the usefulness of the social simulation corresponding to the real world. If users act to the real world based on the simulation results, the action will affect the dynamics of the actual market. Then, it is extracted as the change of economic trends, and the artificial market model of SEMAS is updated. In such a feedback loop the SEMAS approach can offer a *useful* social simulation as a tool to users.

Based on the KISS (Keep It Simple and Stupid) principle, many existing social simulations studies try to eliminate the complexity of the real world as much as possible and to find the simplest model to explain social phenomena. The KISS approach has been successful to catch the essence of some social phenomena. However, since a social system is an open system, the complexity of external information may become important for model creation. Especially in social systems like financial markets, the cognitive process of participants are crucial components and those models have to include the complexity of information itself and the information process. That is, like the KIDS (Keep It Descriptive Stupid) principle [9] instead of a KISS principle, the moderate complexity is necessary for the description of the information on the real world. This research is the first step of the trial to give the moderate complexity to a social model and to embed a social simulation in the real world.

Acknowledgment

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B), no. 16300047.

References

[1] Chasen, "`http://chasen.naist.jp/`," 2006.

[2] K. Carley, J. Diesner, and M. Tsvetovat, "An integrated approach to the collection and analysis of network data," in *Proceedings of NAACSOS04*, 2004.

[3] K. Ahmad, L. Gillam, and D. Cheng, "Textual and quantitative analysis: Towards a new, e-mediated social science," in *Proc. of the 1st International Conference on e-Social Science*, 2005.

[4] Y.-W. Seo, J. A. Giampapa, and K. Sycara, "Financial news analysis for intelligent portfolio management," Carnegie Mellon University, Tech. Rep. CMU-RI-TR-04-04, 2004.

[5] Weka, "`http://www.cs.waikato.ac.nz/~ml/weka/`," 2006.

[6] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Morgan Kaufmann, June 2005. [Online]. Available: http://www.amazon.fr/exec/obidos/ASIN/0120884070/citeulike04-21

[7] D. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley Publishing Company, 1989.

[8] K. Izumi and K. Ueda, "Phase transition in a foreign exchange market: Analysis based on an artificial market approach," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 5, pp. 456–470, 2001.

[9] B. Edmonds and S. Moss, "From kiss to kids - an 'anti-simplistic' modelling approach." in *Multi-Agent and Multi-Agent-Based Simulation, Joint Workshop MABS 2004*, P. Davidsson, B. Logan, and K. Takadama, Eds. Springer, 2005, pp. 130–144.