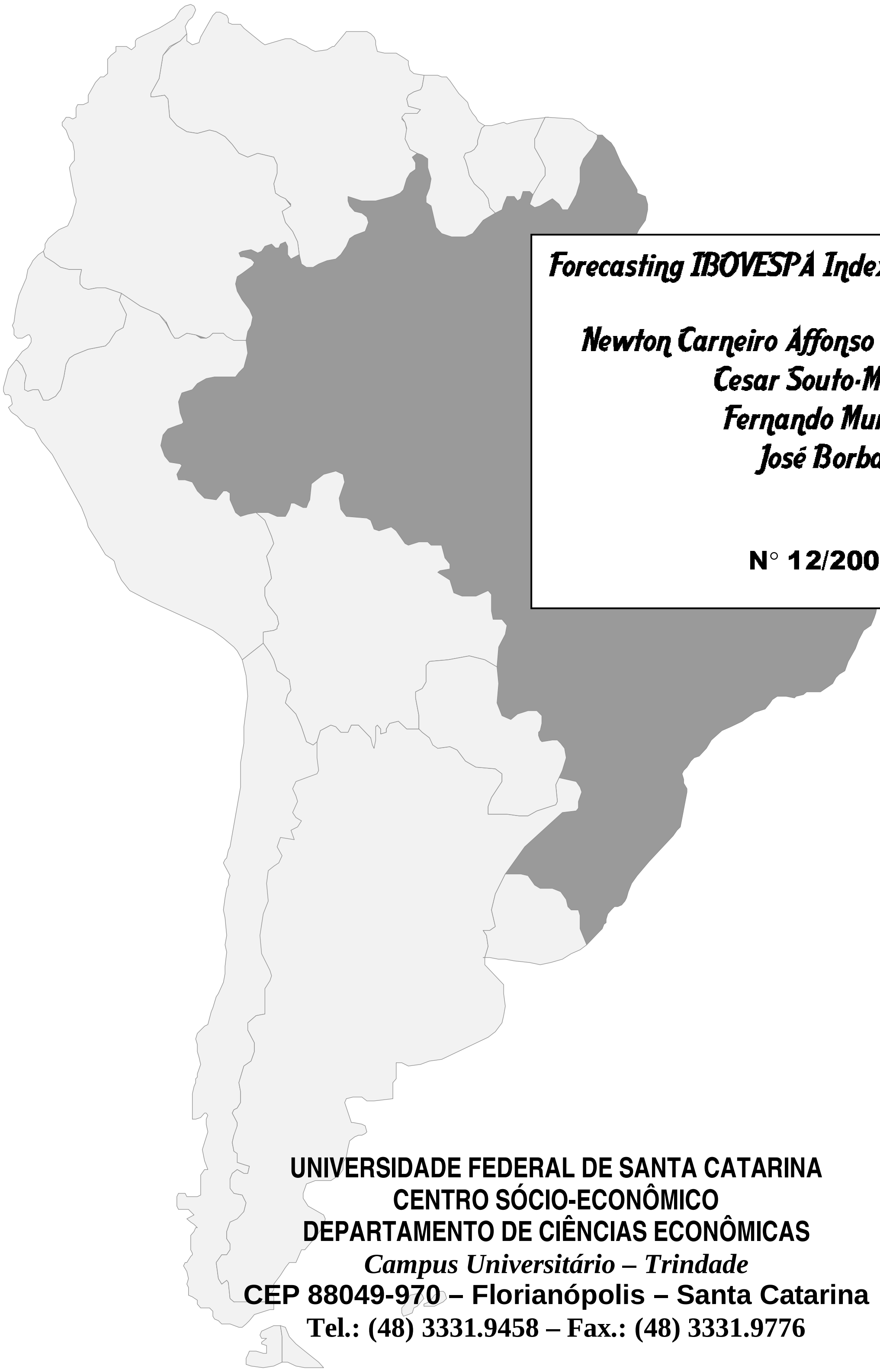
TEXTO PARA DISCUSSÃO



Forecasting IBOVESPA Index with Fuzzy Logic

Newton Carneiro Affonso da Costa Júnior
Cesar Souto-Maior
Fernando Murcia
José Borba

N° 12/2006

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SÓCIO-ECONÔMICO
DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS
Campus Universitário – Trindade
CEP 88049-970 – Florianópolis – Santa Catarina
Tel.: (48) 3331.9458 – Fax.: (48) 3331.9776

Forecasting IBOVESPA Index with Fuzzy Logic

Newton da Costa Jr.*
Cesar Souto-Maior
Fernando Murcia
José Borba

Abstract

This paper presents a new application of fuzzy logic: forecasting the direction/sign of the change in price levels. In order to test this approach, it was applied to forecast the direction of the movements of the Sao Paulo Stock Exchange Index (IBOVESPA). The study period extends over 2,000 daily observations, from January/1997 to February/2005. The first half of the observations was used for parameters estimation, while the second half was used for forecasting. Despite the model produced a linguistic output, it was possible to delineate a statistically significant investment strategy, which outperformed a buy-and-hold one. For futures applications this proposal could be enhanced through the use of other economic and non economic informations, including intuition, to help investment decisions and to produce even better forecasts.

Keywords: Fuzzy logic, Forecasting, IBOVESPA Index

1. Introduction

Since the 1980's the literature related to financial time series has produced important studies that have questioned the (weak) efficient market and the random walk hypotheses (e.g., see Lo and Mackinley, 1988; Poterba and Summers, 1988; Fama and French, 1988). These authors claim that considerable evidence exists that stock returns are to some extent predictable. They show that there are strong evidences of conditional heteroskedasticity in many financial time series, meaning that these time series returns are not independent and identically distributed as established by the random walk model.

These facts have provoked theoretical and practical interest in nonlinear financial time series models based on techniques like auto-regressive moving-average (ARMA family), generalized auto-regressive conditional heteroskedasticity (GARCH family), and more recently, methods based on artificial intelligence like artificial neural networks (ANN), genetic algorithms (GA), and fuzzy logic.

* Corresponding author. This work was partly supported by Capes Foundation, Brazil, while Newton da Costa, Jr. was a visiting scholar at Universitat Pompeu Fabra, Spain.

Basically, predictability of financial time series can be seen in two different ways: Forecasting by level estimation models and forecasting by classification models. The first one relies on accurate prediction of the price level of stocks, indices, and others financial series instruments. The degree of accuracy and the acceptability of the forecast are measured by its deviation from the actual observations, minimizing forecasting errors. Examples of this approach are given by Teixeira and Rodrigues (1997), Rodrigues, Torra and Félix (2005), Ribeiro and Silva (2005), who compared many artificial intelligence based methods with random walk and other linear models, concluding that ANN methods are among the top performance ones. Some methods commonly used for forecasting the level of financial time series are adaptative exponential smoothing, multivariate transfer function, Kalman filter, and multilayered feedforward neural network.

The second way is guided by forecasts on the direction/sign of the change in price level. This approach is defended by some authors (e.g., see Wu and Zhang, 1997; Aggarwal and Demaskey, 1997; Tsaih, Hsu and Lai, 1998; Leung, Daouk and Chen, 2000; Leung and Daouk, 2003; Kim and Chun, 2005), who claim that trading driven by a certain forecast with small forecast error may not be as profitable as trading based on an accurate prediction of the direction/sign of the movement. Some techniques commonly used as classification models are discriminant analysis, logit and probit models, and probabilistic neural network.

In line with the latter view, this study suggests the application of fuzzy logic in order to forecast the direction of the change in price level. Differently from Huarng and Yu (2005) and Yu (2005), which also use fuzzy logic, but in a different context, the proposed model does not return an exact output, but a probabilistic output of linguistic variables. This linguistic output can be used with other economic and non economic information, inclusive intuition, to help investment decisions.

It is worth noting that our approach does not intend to refute or to be compared with other parametric and nonparametric models, but to propose a new solution, based on fuzzy logic concepts.

The paper is organized as follows. In the next section it is given a brief summary of fuzzy logic and its concepts. Section 3 presents the construction of the proposed model. Section 4 presents the final results and Section 5 discusses and concludes the work.

2. Fuzzy Logic

In binary logic, which was initially formulated by the Greek philosopher Aristotle (384 – 322 A.C.), a proposition is either true or false. This type of logic assumes that the states of nature are well defined events. However, in most contexts like in the accounting and business areas, the states of nature are vague, and transitions between ‘what is’ and ‘what is not’ are not very well defined.

Zadeh (1965) published the first paper about fuzzy logic called “Fuzzy Sets”. The model was developed to convert subjective values into objective values. A fuzzy set does not have precise and limited boundaries; the difference between belonging and not belonging does not exist, but a degree of pertinence. According to Zebda (1998) fuzzy sets theory is not a decision theory but rather a calculus (a modeling language) where vague humanistic events can be treated in a systematic manner.

The main objective of fuzzy logic is to provide concepts to perform approximate reasoning. Fuzzy logic assumes a degree of pertinence within the 0 to 1 range, which permits the fuzzy set element to be partially true or partially false.

Bojadziev and Bojadziev (1997, p. 9) define a fuzzy set as:

$$\mathrm{A}' = \{(x, \mu a(x)) \mid x \in \mathrm{A}, \mu a(x) \in [0,1]\},$$

where $\mu a(x)$ is called function of pertinence and specifies the grade in which each element x in A belongs to the fuzzy set A'.

According to von Altrock (1997), the theory of fuzzy sets is a generalization that covers the classical sets where $\mu a(x)$ = 0 or $\mu a(x)$ = 1. In other words, the classical sets are special cases of the fuzzy sets. Table 1 and Figure 1 show the differences between classical and fuzzy sets.

Figure 1, Part A, uses classical logic. If the IBOVESPA index variation is positive, the variation is considered “up” and if it is negative, the variation is considered “down”. There is a rough transition between “up” and “down”. Infinitesimal negative values are classified as “down” and infinitesimal positive values are classified as “up”. However, the market could classify both values as the same information and a system based on classical logic could present not a good performance. Figure 1, part B, shows how fuzzy logic can be used to give a smooth transition between “up” and “down” variations.

3. Methodology

Figure 2 illustrates the proposed model, based on the software FuzzyTech®. This model is divided in three main parts: the fuzzification of input variables, inference rules, and defuzzification of output variables. Here, fuzzification (defuzzification) is understood by the conversion of a numeric (linguistic) value into a linguistic (numeric) one.

However, our conceptual model was divided into four main parts:

- Choice of variables for the forecasting model and definition of the data sets for training and testing;
- Fuzzification;
- Inference rules;
- Output variables.

Each of these parts is described in the next sub-sections.

3.1 Choice of variables to the forecasting model and definition of the data sets for training and testing

The percentage variations of IBOVESPA index in the three days before the day to be forecasted were chosen as input variables. This choice is based on O´Connor, Remus and Griggs (1997). They show that individuals present different tendencies and behavior for up and down series. In that way, people could make decisions based on recent information. As an example, if a stock price increases during many consecutive days, it could provoke a selling behavior that would drive the stock price down, despite of a positive macroeconomic scenario. On other hand, if the stock price decreases during many consecutive days, it could provoke a buying behavior and a tendency to increase the stock price, despite of a negative macroeconomic scenario.

A total of 2,000 daily data sets were collected, with the percentage variation of the forecasted day plus the three previous days.[1] The data were collected from the site http://www.investshop.com.br. The total period extended from January 8, 1997 to February 2, 2005.

In order to calculate day *d* percentage variation, the following formula was used:

$$Var_d = \frac{(V_d - V_{d-1})}{V_{d-1}},$$

where Var_d is the numeric daily variation or return of the IBOVESPA index in day *d*, V_d and V_{d-1} are the IBOVESPA index values in days *d* and *d-1*.

The first 1,000 daily data sets were used for in-sample estimation, covering the period from January 8, 1997 to January 22, 2001. The second 1,000 daily data sets were used for out-of-sample evaluation, covering the period from January 23, 2001 to February 2, 2005.

3.2 Fuzzification

All numerical input variables must be converted to linguistic variables. In this work, the input linguistic variables adopted were “up” and “down”. To accomplish this task, pertinence functions were developed. Figure 3 shows these pertinence functions for the linguistic variable d_1, related with the variation of the day *d-1*, the day before.

3.3 Inference Rules

After the fuzzification of all input values, the next step involves the establishment of inference rules. These rules represent the manner in which humans make decisions, inferring from linguistics premises. For this part of the proposed model, 24 inference rules were created with the help of the software FuzzyTech®. These rules are logical statements, and to each rule it can be assigned a value from zero to one, called Degree of Support (DoS), that depends on the characteristics of the based sample. When a rule is assigned with a DoS equal to zero (one), the rule is considered insignificant (significant). The DoS also allows for values between zero and one for partial significant rule. Below is an example of one of the used rules

[1] Also, it was used the percentage variation of the previous 4 and 5 days with similar results.

in our model.

If (d_3 = “up” **and** d_2 = “up” **and** d_1 = “up”)
Then *d* = “up”, with a DoS (Degree of Support) of .44.

In order to establish the inference rules, from the 1,000 training sets, 557 characteristics sets were selected. These characteristics sets had the values of *d_1* associated to 100% “up” or 100% “down”, the values of *d_2* associated to 100% “up” or 100% “down” and the values of *d_3* associated to 100% “up” or 100% “down”. Table 2 shows these sets distribution.

For a preliminary verification of the model it was selected, from the 1,000 test sets, 422 characteristics sets. Same as before, these sets had the values of *d_1* associated to 100% “up” or 100% “down”, the values of *d_2* associated to 100% “up” or 100% “down” and the values of *d_3* associated to 100% “up” or 100% “down”. Table 3 shows these sets distribution.

Figure 4 presents the probabilistic distribution of the training and testing data when the input is the set “*d_3* = down; *d_2* = up; *d_1* = down”. The up and down probabilities are very similar. In this situation, the investment manager has no additional information to help his decision.

Figure 5 presents the probability distribution of the training and testing data when the input is the set “*d_3* = up; *d_2* = down; *d_1* = up”. In this case, the up probability is superior to the down probability. In this situation, the investment manager could use this additional information to help his decision.

Based on the probability distribution, the inference rules were developed using FuzzyTech®. Table 4 presents the 24 rules used.

3.4 Output variables

The output of the model will be denominated *d*, which represents the variation of the IBOVESPA index in the forecasting day. The output of the linguistic values adopted was: down, flat, and up. In this work the defuzzification was not necessary. The linguistic values were used as output variables, corresponding to the probability of the index showing a down, flat, or up behavior.

The output variable *d* can be represented as a vector of dimension 3x1 as shown below:

$$d = \begin{bmatrix} d1 \\ d2 \\ d3 \end{bmatrix},$$

where $d1$ is the linguistic component "down" of output *d*, $d2$ is the linguistic component "flat" of output *d*, and $d3$ is the linguistic component "up" of output *d*,

Figure 6 shows the output graphic to the output *d* for March 27, 2001. According to the model, the probabilities of down, flat, and up are 38%, 14%, and 48%, respectively.

4. Results

After the model estimation was completed, a buying and selling strategy was applied to the IBOVESPA index which served as a surrogate of an investment fund.

In days in which the linguistic component "down" is greater than a variable called ε, the money is withdrawn from the fund, and applied again in the next day. For example, if ε = .5 it means that the money will be withdrawn from the fund when the "down" probability, indicated by the fuzzy model is greater than 50%. Implicit in this strategy is the supposition that the fund is able to buy and sell stocks in the same proportion as the IBOVESPA index stocks.

Briefly, the algorithm used was:

If $(d1_i > \varepsilon)$

Then $\alpha_i = 0$

Else $\alpha_i = 1$

Where $d1_i$ is the linguistic component "down" for the day *i*, $\alpha_i = 1$ means that the money must be applied in the fund at day *i*, and $\alpha_i = 0$ means that the money must be withdrawn from fund at day *i*.

For the model simulation, the premises of a "frictionless market" were adopted. In other words, all the costs related with transactions were considered as inexistent. However, to counterbalance, the return that could be achieved by investing in the open market in days when the money is withdrawn from the fund was not taken into account.

In order to calculate the buy-and-hold strategy returns, the following formula was used:

$$\prod_{i=1}^{1000}\left(1+Var_{di}/100\right)$$

Actually, the value given by the above equation is the total return from the IBOVESPA index during the test period.

The fuzzy based strategy return is calculated by the following formula:

$$\prod_{i=1}^{1000}\left[\alpha_i \times \left(1+Var_{di}/100\right)+\left(1-\alpha_i\right)\right]$$

In this last strategy it was simulated the use of several ε values. Table 5 presents these results.

The buy-and-hold strategy shows an overall return of 41.48% on the test period, from January 23, 2001 to February 2, 2005. As can be observed, if the ε value is greater or equal to .53, the return will be equal to the fuzzy strategy return, because the fuzzy system returns no output with the “down” probability greater than .53 for the test period. Table 5 shows that the fuzzy based strategy presents returns up to 77.80% in the same period. In absolute values, the fuzzy model outperforms the passive strategy up to 36.33% and, in relative terms, up to 87.58%, with a one-tailed *t* statistic of 1.26, significant at 10% level.

Generalized Sharpe index for passive strategy was 0.0278 and Table 5 presents the generalized Sharpe index for the fuzzy strategy (with each ε value). Figure 7 shows the passive and fuzzy strategy profitability (with ε = .505) of one monetary unit investment during the 1,000 day test period.

5. Conclusion

This study presents a new approach, based on fuzzy logic to forecast the direction of movement of the IBOVESPA index using the FuzzyTech® software. The sampled period comprised 2,000 daily observations of the index, from Jan/1997 to Feb/2005. The first half of the observations is used for the estimation procedures, while the second half is used for forecasting.

The proposed model returns a non exact answer, with a probabilistic output. Despite of this imprecision, the model output could proportionate a (statistically significant) profitable buying and selling strategy which outperforms a buy-and-hold one during the test period.

However, it was not taken into account any transaction costs or taxes. However, to counterbalance, it was not taken into account any possibility of applying in the open market.

Also, the proposed model, with its probabilistic output, can be used as a support to investment decisions, as investors could have other information, secret or not, or even intuitions about political or economic tendencies.

There are many new researches possibilities that could be derived from this work. One of them is the use of derivatives in order to lower transaction costs. Another possibility is the choice of new input variables and linguistic terms. Also other artificial intelligence techniques, as ANN, could be associated with models based on fuzzy logic.

6. References

AGGARWAL, R.; DEMASKEY, A. Using derivatives in major currencies for cross-hedging currency risks in Asian emerging markets. *Journal of Futures Markets*, v. 17, p. 781-796, 1997.

BOJADZIEV, G.; BOJADZIEV, M. *Fuzzy logic for business, finance and management*. Singapore: World Scientific, 1997.

CHEN, A.; LEUNG, M. T.; DAOUK, H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index. *Computers & Operational Research*, v. 30, p. 901-923, 2003.

FAMA, E. F.; FRENCH, K. R. Permanent and temporary components of stock prices. *Journal of Political Economics*, v. 96 (2), p. 246-273, 1988.

FUZZYTECH. http://www.fuzzytech.com/

HUARNG, K.; YU, H. A type 2 fuzzy time series model for stock index forecasting. *Physica A*, v. 353, p. 445-462, 2005.

INVESTSHOP. http://www.investshop.com/

KIM, S. H.; CHUN, S. H. Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index. *International Journal of Forecasting*, v. 14, p. 323-337, 1998.

LEUNG, M. T.; DAOUK, H.; CHEN, A. Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, v. 16, p. 173-190, 2000.

LO, A. W.; MACKINLAY, C. Stock market prices do not follow random walks: evidence from a simples specification test. *Review of Financial Studies*, v. 1 (1), p. 41-66, 1988.

O´CONNOR, M.; REMUS, W.; GRIGGS, K. Going up-going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting*, v. 16, p. 165-176, 1997.

POTERBA, J.; SUMMERS, L. Mean reversion in stock prices: evidence and implications. *Journal of Financial Economics*, v. 22 (1), p. 27-59, 1988.

RIBEIRO, T. S.; SILVA, A. L. C. Do artificial neural networks provide better forecasts?

Evidence from Latin American stock indexes. *XXIX Enampad*. Brasília, 2005

RODRÍGUES, J. V. P.; TORRA, S.; FÉLIX, J. A. STAR and ANN models: forecasting performance on the Spanish "IBEX-35" stock index. *Journal of Empirical Finance*, v. 12, p. 490-509, 2005.

TEIXEIRA, J. C.; RODRIGUES, A. J. An applied study on recursive estimation methods, neural networks and forecasting. *European Journal of Operational Research*, v. 101, p. 406-417, 1997.

TSAIH, R.; HSU, Y.; LAI, C. C. Forecasting S&P 500 stock index futures with a hybrid AI system. *Decision Support Systems*, v. 23 (2), p. 161-174, 1998.

VON ALTROCK, C. *Fuzzy logic and neurofuzzy applications in business and finance*. Upper Saddle River, New Jersey: Prentice Hall, 1997.

YU, H. A refined fuzzy time-series for forecasting. *Physica A*, v. 346, p. 657-681, 2005.

ZADEH, L. Fuzzy Sets. *Information and Control*, v.8, p.338-353, 1965.

ZEBDA, A. The problem of ambiguity and the use of fuzzy set theory in accounting: a perspective and opportunities for research. *Applications of fuzzy sets and the theory of evidence to accounting II*, v. 7, p. 20-33, London: Jai Press, 1998.

Table 1. Classical Sets x Fuzzy Sets.

Classical Sets – Binary Logic	**Fuzzy Sets – Fuzzy Logic**
Precise boundaries	Imprecise boundaries
Rough transition between belonging or not	Smooth transition between belonging or not
Represents well defined concepts	Represents imprecise and vague concepts

Figure 1. Classical Sets x Fuzzy Sets.

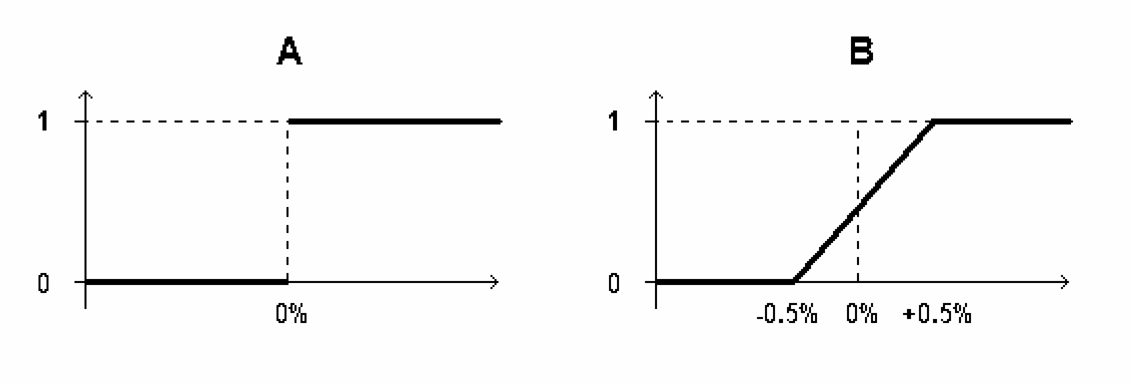
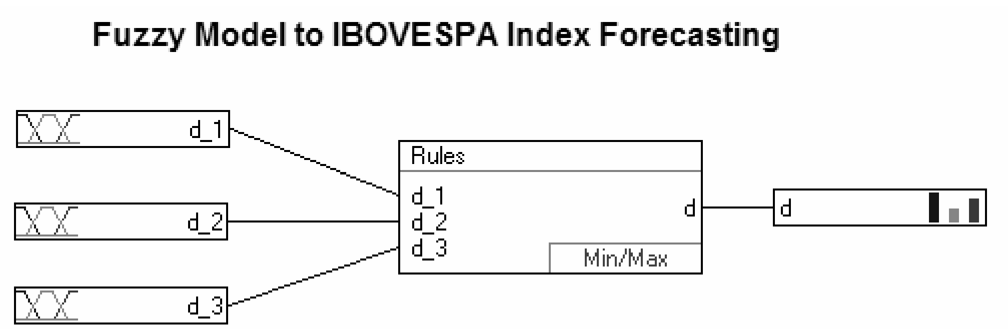
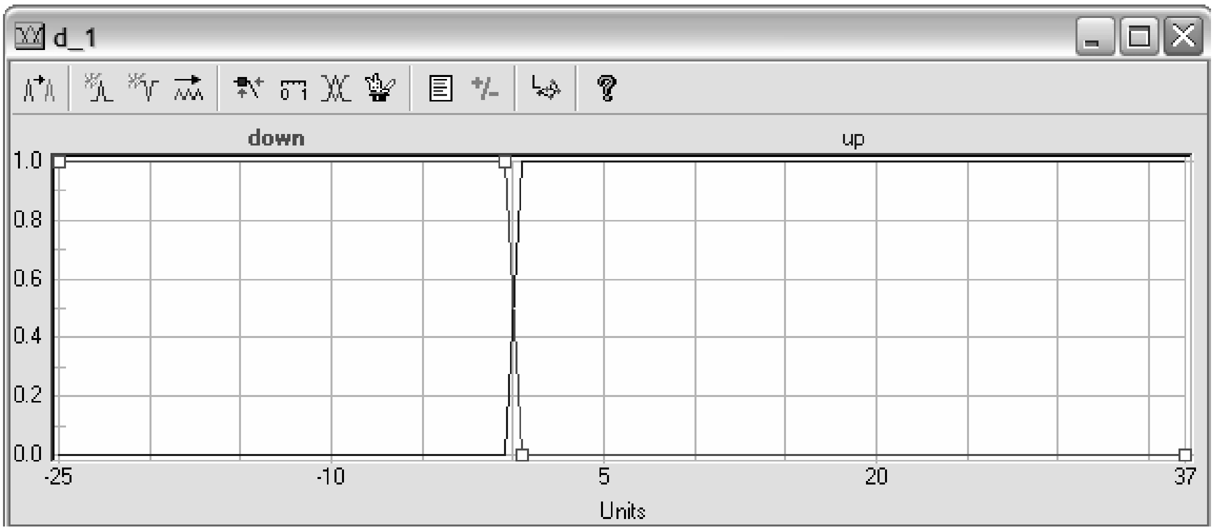


Figure 2. General View of Fuzzy Model.



Note:
d, *d_1*, *d_2*, and *d_3* are the percentage variations (in linguistic values) of IBOVESPA index in days *d*, *d-1*, *d-2* and *d-3*, respectively.

Figure 3. Pertinence functions for the linguistic variable d_1



Notes:

Horizontal axis represents numeric variable Var_{d-1} and vertical axis represents linguistic variable d_1.

Pertinence functions "down" and "up" convert numeric variable Var_{d-1} into linguistic variable d_1.

Table 2. Selected training data

d_3	d_2	d_1	sets
down	down	down	71
down	down	up	74
down	up	down	58
down	up	up	83
up	down	down	65
up	down	up	56
up	up	down	67
up	up	up	83

Table 3. Selected testing data

d_3	d_2	d_1	sets
down	down	down	44
down	down	up	63
down	up	down	48
down	up	up	61
up	down	down	50
up	down	up	50
up	up	down	50
up	up	up	56

Figure 4. Up, Down, and Flat Trend Percentiles of the Training and Testing Data when input is "d_3 = down; d_2 = up; d_1 = down"

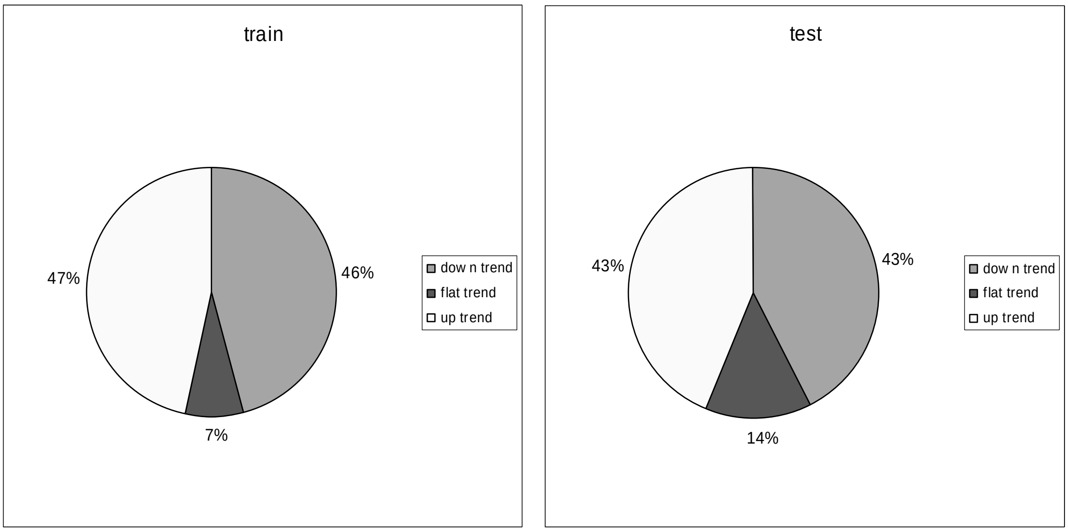


Figure 5. Up, Down, and Flat Trend Percentiles of the Training and Testing Data when input is "d_3 = up; d_2 = down; d_1 = up"

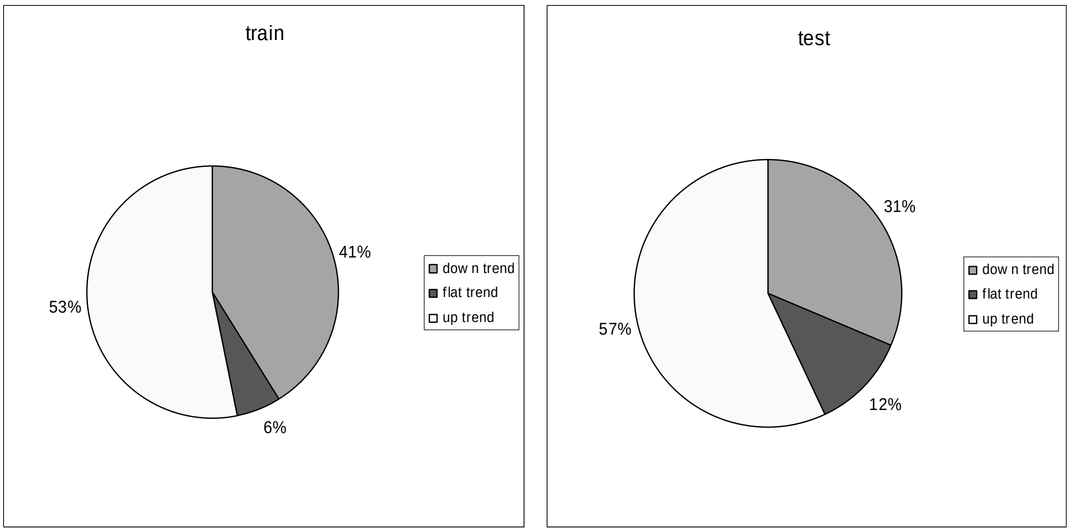
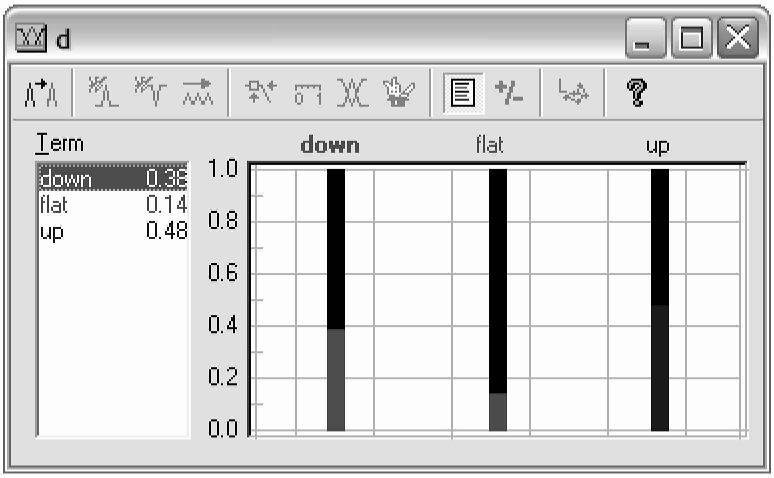


Table 4. Inference Rules

IF			THEN	
d_1	**d_2**	**d_3**	**DoS**	**d**
down	down	down	0.43	down
down	down	down	0.06	flat
down	down	down	0.51	up
down	down	up	0.45	down
down	down	up	0.04	flat
down	down	up	0.51	up
down	up	down	0.46	down
down	up	down	0.07	flat
down	up	down	0.47	up
down	up	up	0.53	down
down	up	up	0.08	flat
down	up	up	0.38	up
up	down	down	0.45	down
up	down	down	0.06	flat
up	down	down	0.49	up
up	down	up	0.41	down
up	down	up	0.06	flat
up	down	up	0.53	up
up	up	down	0.38	down
up	up	down	0.14	flat
up	up	down	0.48	up
up	up	up	0.42	down
up	up	up	0.14	flat
up	up	up	0.44	up

Figure 6. Output Linguistic Values for March 27, 2001.



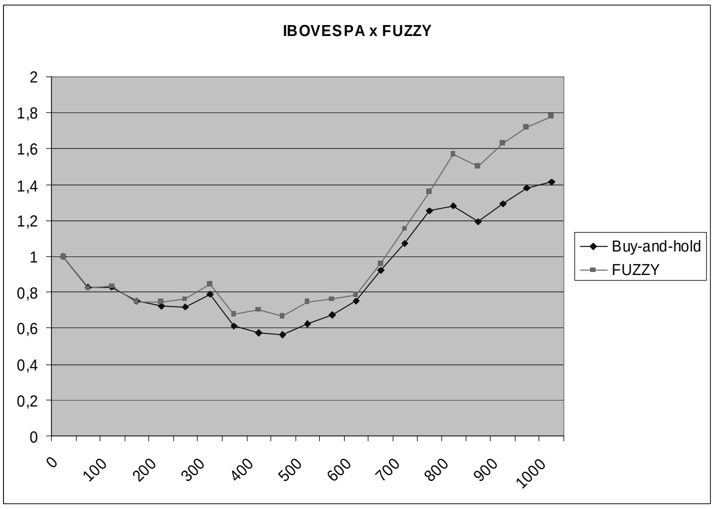
Notes:
"down", "flat" and "up" mean three different possible trends.
Vertical axis represents the probability of each trend.

Table 5. Results from the fuzzy strategy.

		(a) Fuzzy	(b) IBOVESPA	(a-b) difference	(%) difference	Sharpe index
values of ε	0.495	68.52%	41.48%	27.04%	65.19%	0.0382
	0.500	76.74%	41.48%	35.26%	85.01%	0.0408
	0.505	77.80%	41.48%	36.33% (*)	87.58%	0.0409
	0.510	71.68%	41.48%	30.20%	72.81%	0.0389
	0.515	65.63%	41.48%	24.15%	58.22%	0.0368
	0.520	67.52%	41.48%	26.04%	62.79%	0.0374
	0.525	67.23%	41.48%	25.75%	62.08%	0.0372
	0.530	41.48%	41.48%	0.00%	0.00%	0.0278

Note: (*) means statistically significant at 10%.

Figure 7. Profitability of IBOVESPA and fuzzy strategy (with ε = .505).



Notes:
Horizontal axis represents the 1,000 days of the test period;
Vertical axis represents the value of one monetary unit investment.

TEXTO PARA DISCUSSÃO

Nº 01/06 - CÁRIO, Silvio A. Ferraz e ALMEIDA, Carla Cristina Rosa de. *Indústria Automobilística Brasileira: Conjuntura Recente e Estratégias de Desenvolvimento.*

Nº 02/06 - GOULARTI FILHO, Alcides. *A Construção e a Modernização do Porto de Itajaí e Construção e modernização do Porto de São Francisco do Sul.*

Nº 03/06 - MATTEI, Lauro e NIEDERLE, Sidnei L. *O Comportamento do Mercado de Trabalho em Santa Catarina nos anos de 1990.*

Nº 04/06 - VIEIRA, Pedro; CAMERLATTO, Lairton e SANTOS, Fábio Pádua dos. *Revisitando as Origens da Indústria no Brasil: Uma Interpretação da Economia Política dos Sistemas-Mundo.*

Nº 05/06 - NICOLAU, José Antônio e CÁRIO, Silvio A. Ferraz. *Estruturas de Governança em Arranjos Produtivos Locais no Brasil: Um Estudo Empírico.*

Nº 06/06 - ALVES, João Marcos de Souza; MARTINELLI, Orlando e DEWES, Homero. A Dinâmica Inovativa no Agronegócio: A Inovação Tecnológica na Avicultura Industrial através da Análise de Patentes.

Nº 07/06 - LISBOA, Armando de M elo. *Desenterrando o Espelho. A Construção da Identidade Latino-americana.*

Nº 08/06 - SILVA, Eraldo Sérgio da; MATSUSHITA, Raul; GLERIA, Iram; FIGUEIREDO, Annibal. *Hurst exponents, power laws, and efficiency in the Brazilian foreign exchange market.*

Nº 09/06 - CAMPOS, Renato Ramos; CASSIOLATO; José Eduardo; STALLIVIERI, Fábio. *Processos de Aprendizagem e Inovação em Setores Tradicionais: Os Aranjos Produtivos Locais de Confecções no Brasil.*

Nº 10/06 - MEURER,, Roberto; MOURA, Guilherme Valle e NUNES, Maurício Simiano. *O Vencimento de Dívida Pública Cambial Influencia a Taxa de Câmbio? Um Estudo Econométrico para o Brasil no período 2003-2004.*

Nº 11/06 - WEYDMANN, Celso Leonardo, SEABRA, Fernando. *Transmissão de Preços na Cadeia de Carne Suína: Uma Aplicação para os Preços de São Paulo.*

Nº 12/06 - COSTA JÚNIOR, Newton Carneiro Affonso da; SOUTO-MAIOR, César; MURCIA, Fernando; BORBA, José. *Forecasting IBOVESPA Index with Fuzzy Logic*

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SÓCIO-ECONÔMICO
DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS
Campus Universitário – Trindade
CEP 88.049-970 – Florianópolis - Santa Catarina
Tel.: (48) 3331.9458 – Fax (48) 3331.9776