Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

ELSEVIER International Journal of Approximate Reasoning 35 (2004) 291–305
www.elsevier.com/locate/ijar

Financial markets analysis by using a probabilistic fuzzy modelling approach

Jan van den Berg, Uzay Kaymak *,
Willem-Max van den Bergh

Faculty of Economics, Erasmus University Rotterdam, Room H9-11, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands

Received 1 February 2003; accepted 1 August 2003

Abstract

For successful trading in financial markets, it is important to develop financial models where one can identify different states of the market for modifying one's actions. In this paper, we propose to use probabilistic fuzzy systems for this purpose. We concentrate on Takagi–Sugeno (TS) probabilistic fuzzy systems that combine interpretability of fuzzy systems with the statistical properties of probabilistic systems. We start by recapitulating the general architecture of TS probabilistic fuzzy rule-based systems and summarize the corresponding reasoning schemes. We mention how probabilities can be estimated from a given data set and how a probability distribution can be approximated using a fuzzy histogram technique. We apply our methodology to financial time series analysis and demonstrate how a probabilistic TS fuzzy system can be identified, assuming that a linguistic term set is given. We illustrate the interpretability of such a system by inspecting the rule bases of our induced models.

Keywords: Probabilistic fuzzy systems; Fuzzy reasoning; Fuzzy rule base; Data-driven design; Time series analysis

* Corresponding author. Tel.: +31-10-4081350; fax: +31-10-4089167.
E-mail addresses: jvandenberg@few.eur.nl (J. van den Berg), u.kaymak@ieee.org (U. Kaymak), vandenbergh@few.eur.nl (W.-M. van den Bergh).

0888-613X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2003.08.007

1. Introduction

Complex systems such as financial markets are characterized by changing process dynamics, which manifest themselves in various ways like regime shifts and volatility variations. In the specific case of financial markets, it is important to recognize the 'state-of-the-market', so that the market participants' decisions (e.g. trading decisions) can be adapted to the prevailing market conditions in order to safeguard success in the markets. Consequently, many financial models try to capture the changes in the market conditions. An example of such a model is the so-called Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) model [1], which assumes that the current volatility of market returns is dependent on a weighted sum of the long-run volatility, of estimated recent volatilities and of observed recent market returns. More details are given below.

The GARCH model is an example of a probabilistic model. This type of models is almost always used in the world of finance. Probabilistic models deal with probabilistic uncertainty regarding the market developments (e.g. return series). There are, however, often other types of uncertainty present, such as fuzziness in the definitions of concepts and the linguistic uncertainty, which are related to the perception of market participants. These other types of uncertainty should best be modelled explicitly by using paradigms other than probabilistic modelling. Financial models should then ideally combine different paradigms in order to deal with different types of uncertainty. The advantages of such an approach are two-fold. First, explicit modelling of different types of uncertainty separates quantities that are conceptually different. This improves the interpretability of the models, since conceptually different quantities are treated separately. Second, the adaptability of the models can be improved, since different types of information can be dealt with during the model induction process.

In this paper, we propose to use probabilistic fuzzy systems as a general tool for modelling of financial problems, and, more particularly, for the analysis of financial time series. We examine an artificially generated noisy time series following a GARCH process, and a real-world high frequency returns series. Fuzzy systems (FSs) are widely applied in fields like classification, decision support, process simulation, and control [2,3]. Financial and marketing applications have also been reported regularly [4,5]. Original applications of FSs have concentrated on their design from expert knowledge [6,7]. In the past decade, however, data-driven techniques for designing FSs have gained much attention, partly due to the availability of large amounts of data from modern sensory, measurement and computer systems. One important advantage of fuzzy inference systems is their linguistic interpretability, whereby the results from the data-driven approach can be combined with or compared to the knowledge available from experts. When applying FSs, one usually focusses on this aspect by modelling fuzziness and linguistic vagueness using membership functions. However, one has often ignored the probabilistic uncertainty, which is often also present. *Probabilistic fuzzy systems* (PFSs) combine both types of uncertainty in order to provide "the best of the two worlds".

The advantage of PFSs over probabilistic systems is that they combine *interpretability* of fuzzy systems with the *statistical properties* of probabilistic systems. We consider PFSs where the rules describe a stochastic mapping from the antecedent space to the consequent space [8–11]. These PFSs can be considered as a generalization of deterministic rule-based fuzzy systems. In this paper, we concentrate on probabilistic Takagi–Sugeno fuzzy systems and their design from data. We demonstrate how these systems can be applied to financial time series modelling and illustrate how the resulting model can be analyzed and interpreted assuming that a linguistic term set concerning the financial returns is given. The emphasis is put on how fuzzy and probabilistic uncertainties can be simultaneously dealt with. The issue of optimal design is not considered.

The rest of the paper is structured as follows. In Section 2, we recapitulate the general architecture of TS probabilistic fuzzy rule-based systems and summarize the corresponding reasoning schemes. In Section 3, we mention relevant results from the theory of mathematical statistics on fuzzy sets in order to be able to estimate *probabilities on fuzzy sets*. We also illustrate how a probability distribution can be approximated by a fuzzy histogram. In Section 4, we apply the proposed methodology for financial time series analysis and demonstrate how a probabilistic TS fuzzy system can be identified, assuming that a linguistic term set is given. We illustrate the interpretability of such a system by inspecting the rule bases of our models. Finally, the conclusions and a short discussion are given in Section 5.

2. Probabilistic fuzzy systems

For the scope of this paper, we concentrate on zero-order Takagi–Sugeno PFSs, although extensions to other types of fuzzy systems are also possible. The heart of a zero-order Takagi–Sugeno probabilistic fuzzy system consists of a *probabilistic fuzzy rule-base* which is made up of a set of probabilistic fuzzy rules, together with an appropriate *inference mechanism* for reasoning. The probabilistic fuzzy rules have the general form [10]:

$$\begin{aligned} \text{Rule } R_q: \quad \text{If } \mathbf{x} \text{ is } A_q \text{ then} \quad & \underline{y} = y_{q1} \text{ with } \Pr(y_{q1}|A_q) \text{ and} \\ & \underline{y} = y_{q2} \text{ with } \Pr(y_{q2}|A_q) \text{ and} \\ & \vdots \\ & \underline{y} = y_{qN} \text{ with } \Pr(y_{qN}|A_q), \end{aligned} \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_M) \in X$ is an M-dimensional input vector, A_q is an antecedent linguistic value defined by a fuzzy membership function $\mu_q(\mathbf{x})$, $\underline{y}$ is the stochastic consequent variable being equal to one of the values $y_{q1}, y_{q2}, \ldots, y_{qN}$. The selection of this consequent value is done proportionally to the conditional probabilities $\Pr(y_{q1}|A_q), \ldots, \Pr(y_{qN}|A_q)$, with $\forall j : \Pr(y_{qj}|A_q) = Pr(\underline{y} = y_{qj}|\mathbf{x}$ is $A_q)$.

In this paper, we use fuzzy rules (1), where the consequent values y_{qj} are the same for all rules. Mathematically expressed, we assume that

$$\forall j, q, q' : y_{qj} = y_{q'j} = y_j. \tag{2}$$

Hence, each rule describes a probabilistic mapping from a fuzzy antecedent to the same set of crisp consequents. The rules differ in the probabilistic mapping that they describe. This assumption is not restrictive if the consequents are chosen such that they can be used to characterize the whole system output (or equivalently consequent) space.

The reasoning in probabilistic systems essentially performs an interpolation as in many fuzzy systems. The following paragraphs summarize two reasoning schemes.

2.1. Probabilistic fuzzy reasoning I

In this scheme, we begin by estimating the conditional probabilities $\Pr(y_j|\mathbf{x})$ for arbitrary $\mathbf{x}$ and then calculate the regression hyperplane y on $\mathbf{x}$. First, the conditional probabilities $\Pr(y_j|\mathbf{x})$ are calculated by using a weighted sum of conditional probabilities $\Pr(y_j|A_q)$

$$\Pr(y_j|\mathbf{x}) = \sum_{q=1}^{Q} \phi_q \Pr(y_j|A_q) = \frac{\sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x}) \Pr(y_j|A_q)}{\sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x})} \tag{3}$$

with $\phi_q = \Pr(A_q)\mu_q(\mathbf{x}) / \sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x})$. The weight factors ϕ_q take into account both the membership $\mu_q(\mathbf{x})$ to the fuzzy antecedent A_q (often termed the 'firing rate' [3]) and the probability of the fuzzy event A_q. Note that (3) actually implements a *stochastic mapping* $X \to Y$: for each arbitrary input vector $\mathbf{x}$, the conditional probability distribution $\Pr(y_j|\mathbf{x})$, $(j = 1, 2, \ldots)$ is given by (3).

In practice, one often wants to know the *expected behavior* as described by a regression curve, i.e. the regression hyperplane of y on X. This is defined as the location of the mathematical expectations $E(y|\mathbf{x})$ [12], and it can be calculated according to

$$y = E(\underline{y}|\mathbf{x}) = \sum_{j=1}^{N} y_j \Pr(y_j|\mathbf{x}). \tag{4}$$

Note that (4) reduces (3) to a crisp mapping.

2.2. Probabilistic fuzzy reasoning II

It might be of interest to know—for each qth rule—the mathematical expectation of the output variable $\underline{y}$ given fuzzy antecedent A_q. This expectation $E(\underline{y}|A_q)$ can be estimated according to

$$E(\underline{y}|A_q) = \sum_{j=1}^{N} y_j \Pr(y_j|A_q). \quad (5)$$

Using probabilistic fuzzy reasoning, we can also estimate y as a function of $\mathbf{x}$ by calculating the weighted sum of expectations $E(\underline{y}|A_q)$, $q = 1, 2, \ldots, Q$ conform

$$y = \sum_{q=1}^{Q} \phi_q E(\underline{y}|A_q) = \frac{\sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x})E(\underline{y}|A_q)}{\sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x})} \quad (6)$$

with $\phi_q = \Pr(A_q)\mu_q(\mathbf{x}) / \sum_{q=1}^{Q} \Pr(A_q)\mu_q(\mathbf{x})$. Hence, (6) calculates the expected output of the probabilistic fuzzy system by combining the expected output of all rules. Again, the weight factors ϕ_q take into account both the firing rate $\mu_q(\mathbf{x})$ and the probability $\Pr(A_q)$ of the fuzzy event A_q. Note that (6) involves an interpolation procedure, just like (3). Note also that Eqs. (4) and (6) describe the same hyperplane [10], so probabilistic fuzzy reasoning schemes I and II end up with the same deterministic mapping.

3. Mathematical statistics on fuzzy sets

In this section, we describe how the probabilities in Section 2 can be computed from data. Furthermore, we discuss the approximation of probability density functions using so-called fuzzy histograms. The analysis is based on Zadeh's definition of the probability of a fuzzy event [13].

3.1. Probability estimation

Given a set of S samples $\mathbf{x}_s$, $(s = 1, \ldots, S)$ in a 'well-defined' [8] sample space X, the probability $\Pr(A_c)$ describing the probability of the 'fuzzy event' '$\mathbf{x}$ is A_c', can be estimated according to

$$\Pr(A_c) \approx \tilde{f}_{A_c} = \frac{f_{A_c}}{S} = \frac{1}{S} \sum_{\mathbf{x}_s} \mu_{A_c}(\mathbf{x}_s) = \hat{\mu}_{A_c}. \quad (7)$$

Here, $\tilde{f}_{A_c}$ denotes the relative frequency and f_{A_c} the absolute frequency of the fuzzy sample values $\mu_{A_c}(\mathbf{x}_s)$ for fuzzy class A_c. In addition, conditional probabilities on fuzzy sets can be assessed according to

$$\Pr(A_c\,|\,A_b) = \frac{\Pr(A_c \cap A_b)}{\Pr(A_b)} \approx \frac{\sum_{\mathbf{x}_s} \mu_{A_b}(\mathbf{x}_s)\mu_{A_c}(\mathbf{x}_s)}{\sum_{\mathbf{x}_s} \mu_{A_b}(\mathbf{x}_s)}. \tag{8}$$

In Section 2, we mentioned expressions of type $\Pr(y_j\,|\,A_q)$ describing the probability of a crisp event $\underline{y} = y_j$, given the occurrence of fuzzy event $\mathbf{x}$ is A_q. In general, the data regarding a system will have values other than y_j. Then, one needs to define crisp classes Y_j (e.g. by discretizing Y), each of which is represented by a crisp value y_j. In that case, the conditional probability $\Pr(y_j\,|\,A_q)$ can be calculated from a training set of data pairs $(\mathbf{x}_s, y_s)$, $s = 1, \ldots, S$, by using a modified version of (8)

$$\Pr(y_j\,|\,A_q) = \Pr(Y_j\,|\,A_q) \approx \frac{\sum_{(\mathbf{x}_s, y_s)} \chi_j(y_s)\mu_{A_q}(\mathbf{x}_s)}{\sum_{\mathbf{x}_s} \mu_{A_q}(\mathbf{x}_s)}, \tag{9}$$

where $\chi_j(y)$ is defined as

$$\chi_j(y) = \begin{cases} 1 & \text{if } y \in Y_j, \\ 0 & \text{if } y \notin Y_j. \end{cases} \tag{10}$$

3.2. Fuzzy histograms

The technique for estimating a probability density function (pdf) using (crisp) histograms is well-known. By appropriately partitioning the domain of sample space X in a set of Q disjunct classes C_q, each "column" $f_q(\mathbf{x})$, $(q = 1, 2, \ldots, Q)$ of the histogram is defined by the functions

$$f_q(\mathbf{x}) = \begin{cases} \dfrac{\Pr(C_q)}{c_q} & \text{if} \mathbf{x} \in C_q, \\ 0 & \text{if} \mathbf{x} \notin C_q, \end{cases} \tag{11}$$

where the probability $\Pr(C_q)$ is estimated in the usual way (using the relative frequency of samples $\mathbf{x}_s \in C_q$) and where the scaling scalar c_q equals the size of class C_q (which in the one-dimensional case, equals the length of the interval C_q). The pdf $f(\mathbf{x})$ is approximated by a summation of the functions $f_q(\mathbf{x})$ according to

$$f(\mathbf{x}) \approx f_{\text{app}}(\mathbf{x}) = \sum_q f_q(\mathbf{x}). \tag{12}$$

Pdfs defined on a sample space X that is fuzzily partitioned can also be estimated, this time by using a 'fuzzy histogram'. To do so, we need a generalization of the above-given crisp approach. Let X be fuzzily partitioned in a set of Q fuzzy classes A_q described by membership functions $\mu_{A_q}()$, then the (fuzzy) column $f_q(\mathbf{x})$ for fuzzy class A_q can be estimated according to

$$f_q(\mathbf{x}) = \frac{\Pr(A_q)\mu_{A_q}(\mathbf{x})}{\int_{-\infty}^{\infty} \mu_{A_q}(\mathbf{x})\,\mathrm{d}\mathbf{x}}, \tag{13}$$

with $\int_{-\infty}^{\infty} \cdot\, \mathrm{d}\mathbf{x}$ representing an M-fold integral. It is assumed that this integral exists. Eq. (13) is a generalized version of (11): the numerator in (13) describes a probability weighted with membership function $\mu_{A_q}(\mathbf{x})$, the denominator of (13) is a scaling factor representing the fuzzified size of class C_q (which in the one-dimensional continuous case, equals the fuzzy length of the interval C_q). The complete pdf $f(\mathbf{x})$ is again approximated by a summation of the functions $f_q(\mathbf{x})$

$$f(\mathbf{x}) \approx f_{\mathrm{app}}(\mathbf{x}) = \sum_q f_q(\mathbf{x}) = \sum_q \frac{\Pr(A_q)\mu_{A_q}(\mathbf{x})}{\int_{-\infty}^{\infty} \mu_{A_q}(\mathbf{x})\,\mathrm{d}\mathbf{x}}. \tag{14}$$

Due to the overlap of the fuzzy sets, in practice, fuzzy histograms approximate probability distributions better. Fig. 1 shows this phenomenon where a normal probability density function is approximated using both a crisp and a fuzzy histogram. In both cases, seven classes have been used.

Finally, we mention here that definition (14) guarantees that, like in the crisp case, the approximation $f_{\mathrm{app}}(\mathbf{x})$ is properly defined in the sense that

$$\int_{-\infty}^{\infty} f_{\mathrm{app}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1. \tag{15}$$

The proof of this observation is obtained by using (14), so that

$$\begin{aligned}\int_{-\infty}^{\infty} f_{\mathrm{app}}(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \int_{-\infty}^{\infty} \sum_q \frac{\Pr(A_q)\mu_{A_q}(\mathbf{x})}{\int_{-\infty}^{\infty} \mu_{A_q}(\mathbf{x})\,\mathrm{d}\mathbf{x}}\,\mathrm{d}\mathbf{x} \\ &= \sum_q \Pr(A_q)\frac{\int_{-\infty}^{\infty} \mu_{A_q}(\mathbf{x})\,\mathrm{d}\mathbf{x}}{\int_{-\infty}^{\infty} \mu_{A_q}(\mathbf{x})\,\mathrm{d}\mathbf{x}} = \sum_q \Pr(A_q) = 1. \end{aligned} \tag{16}$$

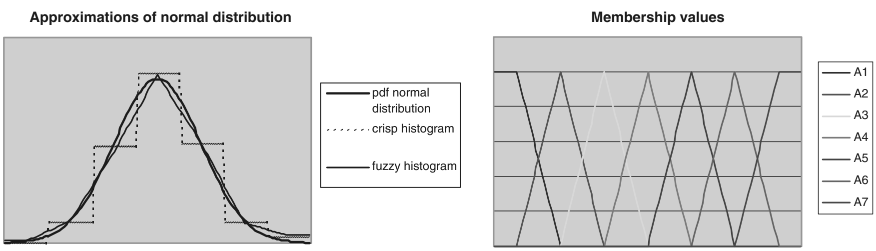

Fig. 1. (left) A fuzzy histogram better approximates a pdf than a crisp histogram, due to (right) overlapping membership functions.

4. Analysis of financial time series

In this section, we give examples of the analysis of financial time series by using probabilistic fuzzy systems. In Section 4.1 an artificial time series generated by a GARCH system is studied. It is shown that a probabilistic TS system can be used to discover some basic properties of the underlying data generating system without making extensive assumptions about the structure of this system. Afterwards, we study high frequency Dow Jones data and discuss the results of our proposed method.

4.1. GARCH modelling

GARCH models are often used in financial literature to describe the volatility behavior of asset return series [1]. Being able to infer something about the volatility of tomorrow from today's volatility has important implications for the valuation of many financial contracts, more particularly for so-called derivatives like 'futures' and 'options'. Typically, the value of such contract depends on the probability that the price $\underline{S}$ of some underlying asset attains a pre-specified level. We define the asset return $\underline{u}(t)$ at time t as the instantaneous relative price change: $\forall t : \underline{u}(t) = \ln(\underline{S}(t)/\underline{S}(t-1))$. Then $\underline{\sigma}(t)$ is the volatility at time t of the return $\underline{u}(t)$, i.e. the standard deviation over a given previous period. This *local* volatility $\underline{\sigma}(t)$ is assumed to move around the constant *global* volatility $\overline{\sigma}$, so in the long run, a GARCH model recognizes that the local volatility reverts to the overall mean value. This property is known as 'mean reversion'.

For purposes of our study, we generate data according to a GARCH(1,1) process [1], which is characterized as follows:

(1) Each return $\underline{u}(t)$ is drawn from a normal distribution with a constant mean μ and with a standard deviation equal to the local volatility $\underline{\sigma}(t)$: $\underline{u}(t) \sim N(\mu, \underline{\sigma}(t))$.
(2) Each period, the local volatility estimate is updated by using
$$\underline{\sigma}^2(t) = \gamma\overline{\sigma}^2 + \alpha\underline{u}^2(t-1) + \beta\underline{\sigma}^2(t-1). \tag{17}$$
(3) The parameter values used are in line with those found empirically in stock return series: $\overline{\sigma} = 0.03$, $\gamma = 0.02$, $\alpha = 0.2$ and $\beta = 0.78$. The series is initiated with $\sigma(0) = \overline{\sigma}$.

From (17) we observe that for GARCH(1,1) the local volatility is determined by the long-run volatility $\overline{\sigma}$, by the observed most recent return ($\underline{u}(t-1)$), and by the estimation of the most recent local volatility ($\underline{\sigma}(t-1)$).

In Fig. 2 we show simulation results for 1000 consecutive samples. The return series in the left graph exhibit volatility clusters that are typical for the

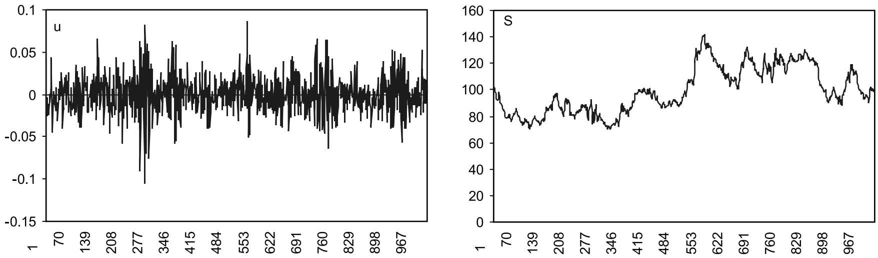

Fig. 2. (left) Return path and (right) price path from a simulated GARCH process.

process (because of the relatively high value of the autocorrelation parameter β). The right graph shows the price development that, starting with $S(0) = 100$, is calculated from the instantaneous return as $S(t) = S(t-1)e^{u(t)}$.

4.1.1. Characterizing the input space

The left panel of Fig. 3 shows a scatter plot of the product space $\underline{u}(t-1) \times \underline{u}(t)$ of the antecedents and the consequent. In the probabilistic fuzzy rule base, we consider three antecedent linguistic values A_q, defined by fuzzy membership functions $\mu_{A_q}(u), q = 1, 2, 3$, (see the right panel of Fig. 3). The corresponding linguistic values 'Low', 'Average' and 'High', respectively, describe return values in linguistic terms. Using (7), we have estimated the corresponding probabilities yielding Pr($\underline{u}(t-1)$ is 'Low') = 0.0594, Pr($\underline{u}(t-1)$ is 'Average') = 0.8722, and Pr($\underline{u}(t-1)$ is 'High') = 0.0684.

We can also approximate the pdf $f(u)$ by using a fuzzy histogram based on the fuzzy partition of $\underline{u}(t-1)$ as shown in Fig. 3. In the left panel of Fig. 4, the fuzzy histogram on the input space, computed according to (13), is presented. The calculations can be summarized as

$$f_1 = \frac{0.0594 \ * \ \mu_{A_1}(u)}{0.0625}, \quad f_2 = \frac{0.8722 \ * \ \mu_{A_2}(u)}{0.0750}, \quad f_3 = \frac{0.0684 \ * \ \mu_{A_3}(u)}{0.0625}, \tag{18}$$

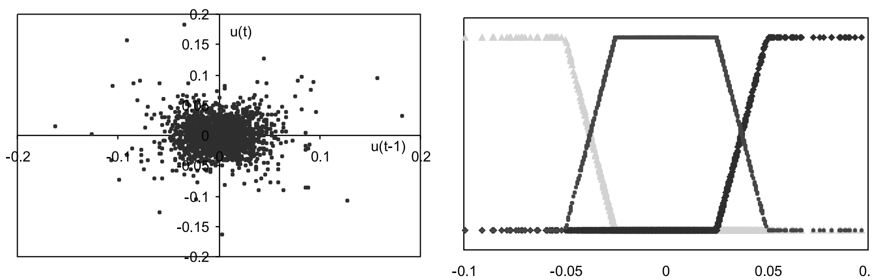

Fig. 3. (left) Scatter of $\underline{u}(t-1)$ against $\underline{u}(t)$ and (right) membership functions for $\underline{u}(t-1)$.

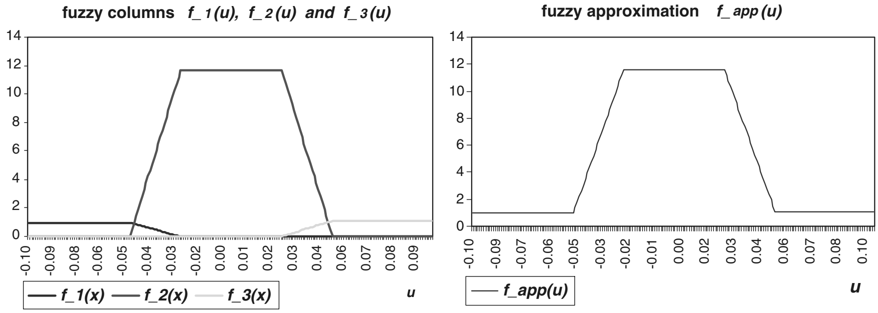

Fig. 4. (left) Fuzzy histogram and (right) fuzzy approximation $f_{\text{app}}(u)$ following (16): the area under the curve equals 1.0.

where the membership functions $\mu_{A_q}(u)$ are given by the functions in the right panel of Fig. 3. In the right panel of Fig. 4, the fuzzy approximation of the pdf $f(u)$, defined on the input space according to (14), is shown. Note that the antecedent space is partitioned very roughly in only three partitions, but the approximation is already indicative of the distribution observed in Fig. 3. This result is in line with the observations on fuzzy histograms as given in the previous section.

4.1.2. Characterizing the output space

In order to keep this illustrative example simple, we choose (without further optimization) five equidistant crisp consequent values $u_1 = -0.050$, $u_2 = -0.025$, $u_3 = 0.000$, $u_4 = 0.025$, $u_5 = 0.050$ to describe the future returns. For reasons of interpretability, we label these arithmetic values with the linguistic terms 'very low', respectively 'low', 'average', 'high', 'very high'. Then, each return value from the time series is classified according to the nearest prototype value using the Euclidian norm. By simply counting all u-values and determining the relative score, we make an estimate of the (unconditional) output probability distribution of $\Pr(u_j) = \Pr(\underline{u}(t) = u_j)$, $j = 1, 2, \ldots, 5$. The results of these calculations are shown in the (emphasized) row labelled *All* in Table 1.

4.1.3. Characterizing the probabilistic fuzzy input–output mapping

By using (9), we can also calculate $\Pr(u_j|A_q)$, $j = 1, 2, 3, 4, 5$; $q = 1, 2, 3$. It concerns probabilities like 'the probability that the future return is high given that the current return is Low'. These conditional probabilities are summarized in Table 1.

To analyze the results, we compare the probability distribution conditional on current returns to the overall probability distribution indicated by the all row of Table 1. It becomes clear that for low current returns, the probability

Table 1
Unconditional and conditional probabilities $\Pr(u_j)$ and $\Pr(u_j|A_q)$

Current return	Future return					Probability
	very low (−0.05)	low (−0.025)	average (0)	high (0.025)	very high (0.05)	
All	*0.0550*	*0.2265*	*0.4435*	*0.2140*	*0.0610*	*1.0000*
Low	0.1271	0.2084	0.2954	0.2302	0.1390	0.0594
Average	0.0437	0.2293	0.4666	0.2136	0.0468	0.8722
High	0.1374	0.2077	0.2808	0.2063	0.1679	0.0684

for very high or very low future returns is higher than the same probabilities for the overall returns. A similar conclusion can also be drawn for high current returns. For low or high current returns, the deviation of low and high future returns from the overall probability distribution is also visible, although to a lesser extent. If we attach linguistic values to the magnitude of the difference between the conditional probability and the overall probability like 'more than 5%' is 'very likely' and 'more than 2%' is 'rather likely', then the above results can thus be summarized as

> If *current* return is Low or the *current* return is High, then low or high *future* returns are rather likely, and very low or very high *future* returns are very likely.

This is a pretty good intuitive description of the GARCH process that has generated the data, where periods of high returns are correlated to periods of high volatility.

Finally, we show two additional results. First, we have plotted the regression line of $u(t)$ on $u(t-1)$ (estimated according to (4)) in the right panel of Fig. 5. As expected for this problem, we found that $u(t) \approx 0$. In the left panel of the same figure, we show the difference between the conditional probabilities $\Pr(u_j|u(t-1))$ and the unconditional probability $\Pr(u_j)$, for $j = 1, 2, 3, 4, 5$. If current returns are Average, we observe that all conditional probabilities are almost equal to the unconditional one. However, if current returns are Low or High, we observe differences in the probability distribution of the future returns $u(t)$, i.e. average future returns are less dominating under those conditions, while lower and higher future returns are more probable, indicating a high volatility regime.

4.2. Analysis of high frequency return series

In this section, we apply a similar analysis as in Section 4.1 to a real return time series. The goal is to illustrate what kind of information the probabilistic

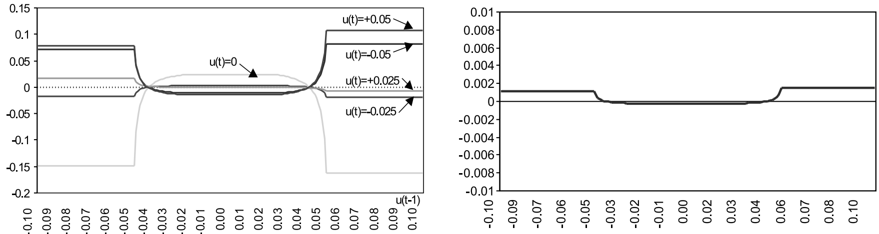

Fig. 5. (left) $\Pr(u_j\,|\,u(t-1)) - \Pr(u_j)$ and (right) regression line of $u(t)$ on $u(t-1)$.

TS fuzzy models can provide when the underlying process that has generated the data is not known.

The data consists of half-hourly samples of the most recent tick of the Dow Jones index of the New York Stock Exchange. The samples run from 31.12.95 19:30 to 31.12.96 19:00, in total 17567 individual data points. Of these, we have used only the samples from the opening hours of the market, which has reduced the number of data points to 3843. The half-hourly returns and the index values are depicted in Fig. 6.

We have studied the behavior of one-step ahead returns $\underline{u}(t)$ conditional on $\underline{u}(t-1)$. The scatter plot of $\underline{u}(t)$ against $\underline{u}(t-1)$ is depicted in the left panel of Fig. 7. The right panel of Fig. 7 shows the fuzzy partitioning of the antecedent space $\underline{u}(t-1)$, where the five fuzzy classes sets are placed evenly over the antecedent space, each one of which is defined by a smooth membership function. The fuzzy approximation for the distribution of the half-hourly returns is depicted in Fig. 8. Note the 'fat-tail-like' phenomenon observed in the figure. We also observe the smoothness of the approximation as resulting from the use of smooth membership functions.

For the consequent space, we use five discrete values, −0.01 ('very low'), −0.005 ('low'), 0 ('average'), 0.005 ('high') and 0.01 ('very high'). Assuming that each half-hourly return is classified to the nearest discrete value, we can compute the unconditional probability distribution $\Pr(u_j) = \Pr(\underline{u}(t) = u_j)$,

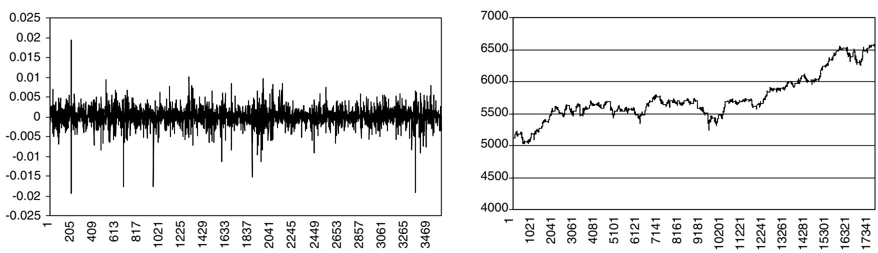

Fig. 6. (left) Half-hourly returns and (right) corresponding Dow Jones index.

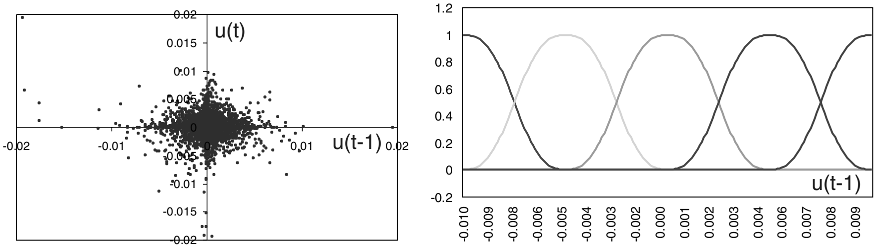

Fig. 7. (left) Scatter of $\underline{u}(t-1)$ against $\underline{u}(t)$ and (right) membership functions for $\underline{u}(t-1)$.

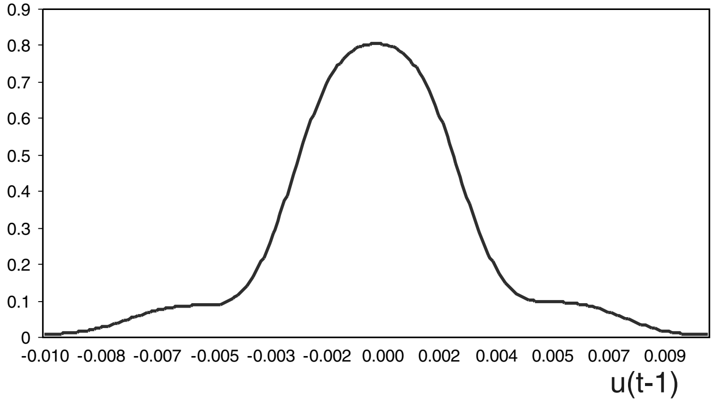

Fig. 8. Fuzzy approximation $f_{\mathrm{app}}(u)$ of the probability distribution for half-hourly returns.

$j = 1, 2, \ldots, 5$. The results of these calculations are shown in the (emphasized) row (labelled *All*) of Table 2. The conditional probability distributions are computed by using (9), and they are also shown in Table 2.

The resulting model can be interpreted by studying the deviations in the rows of Table 2 from the first row of Table 2. We observe the following. At Low current return values $\underline{u}(t-1)$, we have a somewhat higher probability of (very) low future return values $\underline{u}(t)$. This may indicate some GARCH-like behavior, although the probability of high subsequent values is not increased and the picture is different for High current return values. Far more striking is the behavior at Very Low and Very High current return values. Here, we observe a clear reversal in the sense that Very Low current values have an increased probability to be followed by very high values. For Very High values there is even a greater probability of a subsequent very low value. Such reversal behavior has been reported in literature for different sample rates, mostly for long term sampled data such as weekly, monthly or yearly data [14].

Table 2
Conditional probabilities $\Pr(u_j|A_q)$ for the Dow Jones data

Current return	Future return				
	very low (−0.01)	low (−0.005)	average (0)	high (0.005)	very high (0.01)
All	*0.004*	*0.073*	*0.839*	*0.081*	*0.003*
Very low	0.000	0.056	0.683	0.220	0.041
Low	0.010	0.083	0.828	0.076	0.003
Average	0.004	0.071	0.840	0.081	0.003
High	0.001	0.079	0.846	0.074	0.000
Very high	0.062	0.127	0.774	0.037	0.000

5. Conclusions and discussion

In this paper, we have described zero-order Takagi–Sugeno (TS) probabilistic fuzzy systems, which implement a *stochastic input–output mapping*. If desired, the stochastic mapping can be converted in a *deterministic input–output mapping* describing the expected behavior. For both types of mappings, appropriate reasoning schemes are presented, which, unlike classical fuzzy systems, take the statistical properties of the data explicitly into account. Further, we have shown a technique for representing fuzzy histograms. We illustrated our theoretical observations by analyzing a simulated GARCH type of financial time series data and by analyzing high-frequency Dow Jones index data.

The findings presented in this paper constitute only a first step. Nevertheless, we already believe that the PFSs as presented in this paper will turn out to be a very fruitful paradigm for combining fuzzy and statistical uncertainty and that this framework provides tools for getting 'the best of the two worlds'. This also enhances the adaptation power of our models to different types of uncertainty present in real-world problems. Extensions of the proposed approach are under construction, such as the design of appropriate probabilistic fuzzy reasoning schemes for other types of FSs. For an example, we refer to [15], where probabilistic fuzzy reasoning schemes are introduced for probabilistic Mamdani-type of fuzzy systems having fuzzy rules with both fuzzy antecedents and fuzzy consequents. In addition, the issue of optimal design is an interesting and challenging next step. At the same time, we are working on applications, most importantly in the area of financial time series analysis.

Acknowledgements

We acknowledge the help of Olsen Associates, Zurich, who made the high frequency data series available for the example in this paper.

References

[1] J.C. Hull, Options, Futures, & Other Derivatives, fourth ed., Prentice-Hall, Upper Saddle River, 2000.
[2] M. Jamshidi, A. Titli, L. Zadeh, S. Boverie (Eds.), Applications of Fuzzy Logic, Prentice-Hall, New Jersey, 1997.
[3] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, Upper Saddle River, 1995.
[4] C. von Altrock, Fuzzy Logic and Neurofuzzy Applications in Business and Finance, Prentice-Hall, Upper Saddle River, New Jersey, 1997.
[5] C. Zapounidis, P.M. Pardalos, G. Baourakis (Eds.), Fuzzy Sets in Management, Economics and Marketing, World Scientific, Singapore, 2001.
[6] E.H. Mamdani, Applications of fuzzy algorithms for control of simple dynamic plant, in: Proceedings IEE, 121, 1974, pp. 1585–1588.
[7] E.H. Mamdani, Application of fuzzy logic to approximate reasoning using linguistic systems, IEEE Transactions on Computers 26 (12) (1977) 1182–1191.
[8] J. van den Berg, W.M. van den Bergh, U. Kaymak, Probabilistic and statistical fuzzy set foundations of competitive exception learning, in: Proceedings of the Tenth IEEE International Conference on Fuzzy Systems, Melbourne, Australia, vol. 2, 2001, pp. 1035–1038.
[9] J. van den Berg, U. Kaymak, W.-M. van den Bergh, Fuzzy classification using probability-based rule weighting, in: Proceedings of 2002 IEEE International Conference on Fuzzy Systems, Honolulu, HI, 2002, pp. 991–996.
[10] J. van den Berg, U. Kaymak, W.-M. van den Bergh, Probabilistic reasoning in fuzzy rule-based systems, in: P. Grzegorzewski, O. Hryniewicz, M.A. Gil (Eds.), Soft Methods in Probability, Statistics and Data Analysis, Advances in Soft Computing, Physica Verlag, Heidelberg, 2002, pp. 189–196.
[11] G.C. van den Eijkel, Fuzzy probabilistic learning and reasoning, Ph.D. thesis, Delft University of Technology, Delft University Press, Mekelweg 4, Delft, The Netherlands, January 1999.
[12] V. Kecman, Learning and Soft Computing, MIT Press, Cambridge, MA, 2001.
[13] L.A. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23 (1968) 421–427.
[14] W.M.F. De Bondt, R.H. Thaler, Does the stock market overreact?, Journal of Finance 40 (3) (1986) 793–807.
[15] U. Kaymak, W.M. van den Bergh, J. van den Berg, A fuzzy additive reasoning scheme for probabilistic mamdani fuzzy systems, in: Proceedings of the Twelfth IEEE International Conference on Fuzzy Systems, St. Louis, USA, 2003, pp. 331–336.