

# Estimation of flexible fuzzy GARCH models for conditional density estimation<sup>☆</sup>

Rui Jorge Almeida<sup>a,\*</sup>, Nalan Baştürk<sup>a</sup>, Uzay Kaymak<sup>b</sup>, João Miguel da Costa Sousa<sup>c</sup>

<sup>a</sup>*Erasmus University Rotterdam, Erasmus School of Economics, Econometric Institute,  
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*

<sup>b</sup>*School of Industrial Engineering, Eindhoven University of Technology, P.O. Box 513,  
5600 MB Eindhoven, The Netherlands*

<sup>c</sup>*Technical University of Lisbon, Instituto Superior Técnico, Dept. of Mechanical  
Engineering, CIS/IDMEC – LAETA, Av. Rovisco Pais, 1049-001 Lisbon, Portugal*

---

## Abstract

In this work we introduce a new flexible fuzzy GARCH model for conditional density estimation. The model combines two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. The probabilistic uncertainty is modeled through a GARCH model while the fuzziness or linguistic vagueness is present in the antecedent and combination of the rule base system. The fuzzy GARCH model under study allows for a linguistic interpretation of the gradual changes in the output density, providing a simple understanding of the process. Such a system can capture different properties of data, such as fat tails, skewness and multimodality in one single model. This type of models can be useful in many fields such as macroeconomic analysis, quantitative finance and risk management. The relation to existing similar models is discussed, while the properties, interpretation and estimation of the proposed model are provided. The model performance is illustrated in simulated time series data exhibiting complex behavior and a real data application of volatility forecasting for the S&P 500 daily returns series.

---

<sup>☆</sup>This work is partially supported by the European Science Foundation through COST Action IC0702 and by the Netherlands Organisation for Scientific Research (NWO) grant number 400-09-340.

\*Corresponding author

*Email address:* rjalmeida@ese.eur.nl (Rui Jorge Almeida)

*Keywords:* Conditional density estimation, Fuzzy GARCH models, Volatility forecasting, Linguistic descriptions.

*JEL classification:* C14, C22, G32.

---

## 1. Introduction

The conditional density of a random variable is an estimate of the probability distribution of the current value of that variable, given its past values or other variables. Conditional density estimation has an important role in many fields such as quantitative finance and risk management for two main reasons. First, most financial return series appear to be uncorrelated over time, but to be dependent through their higher moments such as the conditional variance [1]. Traditional models aiming at point forecasts cannot capture such dependency and the need to estimate the full conditional density arises. Second, investors are not only interested in the expected return from an asset but also in the risk involved in the asset. This risk factor can be calculated using statistical quantiles of the estimated returns distribution, such as Value-at-Risk or Expected Shortfall [2], and it cannot be assessed from models providing point forecasts.

Estimating an accurate model for the distribution of financial returns is not a simple task since financial time-series typically possess non-trivial statistical properties, such as fat tails, asymmetric distributions and changing variation over time. For this reason, several methods are proposed to estimate the density of returns, conditional on past information, or other macroeconomic variables. A popular approach where volatility, and hence the return distribution, changes dynamically is the Generalized Autoregressive Heteroskedasticity (GARCH) model [1]. In this model, the variation in returns is explained by past returns and past variations in returns. Extended GARCH models are proposed in the literature to capture different aspects of data behavior, such as the GJR-GARCH [3] models to capture skewness and Student- $t$  GARCH models to capture fat tails [4].

The existence of different types of GARCH models led to the introduction of models which can encompass different GARCH specifications and different return distribution properties. In terms of purely probabilistic models, smooth transition GARCH models [5] and regime-switching GARCH models [6] are proposed. Despite the generality of these model structures, estimation of these models is not trivial [7] and it is impossible to apply

standard maximum likelihood estimation, due to the recursive structure of conditional volatility [8, 9]. Artificial neural networks [10] and fuzzy systems have also been combined with GARCH models in different forms. In [11, 12, 13, 14, 15, 16, 17], fuzzy GARCH models are presented in the form of fuzzy rule base systems, where each rule corresponds to an individual GARCH model. Different types of GARCH models were also combined using adaptive neuro-fuzzy inference systems [18, 19, 20], and rough-set based neuro-fuzzy systems [21], although in these cases the models are used to approximate either the return series [19, 20] or realized volatility [18, 21], which is the sum of squared intra-daily (e.g. 5 minutes data) returns. The class of models and objective functions used to estimate realized volatility are different from the models used to estimate conditional distribution of financial returns, since realized volatility is a point estimation, while return volatility considers the estimation of the whole conditional distribution of financial returns <sup>1</sup>. A GARCH model with fuzzy coefficients is presented in [28] and [29], where the error term is modelled using a set of fuzzy rules. These models combine fuzzy systems with a statistical model. A similar approach that combines different types of uncertainty is the probabilistic fuzzy system [30, 30], also used to estimate conditional volatility of returns [31, 32].

In previous work, we introduced key ideas for a new flexible fuzzy GARCH model for conditional density estimation [33], but a formal description and analysis of this type of systems still needs to be made. The model combines two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. The properties, estimation issues and interpretation, which were not considered in [33], are studied in detail in this paper. Furthermore, an overview of the differences with previous fuzzy GARCH models is given. The presented model is more general than the previously studied fuzzy rule base GARCH models and can capture different properties of data, such as fat tails, skewness and multimodality in one single model. The conditional distribution of the data can vary smoothly over time in mean

---

<sup>1</sup>Forecasting return volatility can be seen as equivalent to forecasting realized volatility [22, 23, 24], under the assumptions that high-quality intraday return data are available (noisy intraday data has a detrimental impact on forecast accuracy [25]) and the (log) return series is a continuous semimartingale process [26]. For the case of discrete time process, such as GARCH, the assumption of homoskedasticity (all random variables have the same finite variance) must hold [27]. If these assumptions are not fulfilled, the point estimation of return volatility will be biased.

and variance, where the smooth changes are related to linguistic descriptors. Previous fuzzy GARCH models [11, 12, 13, 14, 15, 16, 17] only allowed for unimodal and symmetric distributions, hence this type of systems could only model fat tail distributions, not skewed or multimodal distributions.

An interpretation of the proposed fuzzy GARCH models, from both statistical and fuzzy linguistic points of view is provided in this paper. The proposed fuzzy GARCH model provides a linguistic interpretation of the gradual changes in return density, providing a simple understanding of the underlying changes. The output of the proposed fuzzy GARCH model is similar to the output of regime-switching and smooth transition GARCH models since the obtained return distribution can have a nonstandard functional form. An advantage of the proposed model is the tractable form of the likelihood function, which in turn does not suffer from the estimation issues reported in pure probabilistic flexible GARCH models [5, 6].

The performance and estimation issues of the proposed model are examined using simulated data and a real data application on S&P 500 return series. It is shown that the proposed model captures the conditional volatility of the data in all examples considered. The proposed model is suitable for analysis of the returns distribution. The main focus in analyzing the returns distribution is not to consider a single model and the parameters, e.g. to draw policy conclusions, but rather to estimate the expected gains and losses from investing in an asset and to use the latest information in the market for investment decisions. The reason for the proposed model to successfully capture such interesting values is two fold. First, the flexible functional form allows to approximate a nonstandard returns density. Second, possibly complex effects of current market information on future returns is explained using simple linguistic descriptors and with a well studied GARCH-type rule base system.

The outline of the paper is as follows. Section 2 gives an overview of previously studied probabilistic and fuzzy GARCH models. Special attention is given to inconsistencies in the explanation of estimation in existing fuzzy GARCH models. The proposed new fuzzy GARCH model is presented in Section 3 and compared to other fuzzy and probabilistic GARCH models. An interpretation of the model is given, from the point of view of its probabilistic output as well as from a linguistic perspective. We show that all model parameters can be estimated using a maximum likelihood approach, in which the objective function includes the whole output density. Examples of this estimation are given in Section 4, where we also show that the proposed

model can successfully capture existing fuzzy GARCH models. In Section 5 we present a real world application of the new fuzzy GARCH model for conditional density estimation, and finally, Section 6 concludes the paper.

## 2. GARCH Model and extensions

GARCH models are used to capture the time varying behaviour of variance. These models relate the unobserved volatility/variance of data to the past variance and past observations. Hence, the conditional density of the data is a normal distribution, but the occurrence of positive or negative extreme data values depends on the past observations together with past volatility. The standard GARCH( $p, q$ ) model for  $t = 1, \dots, T$  observations is defined as:

$$y_t = \sqrt{h_t} \epsilon_t \sim NID(0, h_t) \quad (1a)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \quad (1b)$$

$$\epsilon_t \sim NID(0, 1) \quad (1c)$$

where  $y_t$  is the data with a conditional normal distribution and  $NID(\mu, \sigma^2)$  denotes the normal and independently distribution with mean  $\mu$  and variance  $\sigma^2$ . Scalars  $p$  and  $q$  are, respectively, the lag order for past returns and past conditional volatility in the GARCH model and  $(\alpha_0, \alpha_i, \beta_j)$  for  $i = 1, \dots, q$  and  $j = 1, \dots, p$  are GARCH model parameters. At each period, the *conditional volatility*,  $h_t$ , is assumed to move around the constant *unconditional volatility*  $\bar{h}$ . In the long run, the local volatility reverts to its overall mean value. This property is known as ‘mean reversion’. The residual variance is fixed to 1 since both  $h_t$  and  $\epsilon_t$  in (1) are unobserved. This model is not identified unless the residual variance is fixed. When  $\beta_j = 0, \forall j$ , the model simplifies to an ARCH( $q$ ) model [34] which relates the data variance only on its observed past values. The long run (unconditional) volatility can be written in terms of the model parameters:

$$\bar{h} = \alpha_0 / \left( 1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j \right) \quad (2)$$

Sufficient conditions for positive variance  $h_t$  at every period are:

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1, i = 1, \dots, q, j = 1, \dots, p, \quad (3)$$

where these restrictions also ensure a stationary variance process and the existence of a finite mean and variance of  $h_t$ .

For the GARCH model,  $\max(p, q)$  is the number of observations to leave out as the past information is not available fully for these observations. The actual observations to use in the model starts from:  $t^* = \max(p, q) + 1$ . Initial observations  $y_1, \dots, y_q$  can be obtained from the data or set as the unconditional mean of the data. The initial unobserved variances  $h_1, \dots, h_p$  can be set as the unconditional variance of the data. Conditional on these initial values, the likelihood of a single observation is:

$$\ell(y_t | I_{t-1}) = \ell(y_t | h_t) = \phi(y_t; 0, h_t) \quad (4)$$

for  $t \in \{t^*, \dots, T\}$ , where  $I_{t-1} = \{y_1, \dots, y_{t-1}, h_1, \dots, h_{t-1}\}$  denotes the information set at time  $t-1$ , and  $\phi(\cdot; \mu, \sigma^2)$  is the probability density function (pdf) of a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .

Using the independence assumption in (1), the likelihood of the whole sample is obtained by multiplying (4) for all  $t$ :

$$\ell(y) = \prod_{t=t^*}^T \ell(y_t | I_{t-1}) = \prod_{t=t^*}^T \ell(y_t | h_t) = \prod_{t=t^*}^T \phi(y_t; 0, h_t), \quad (5)$$

where  $y = \{y_{t^*}, \dots, y_T\}$  and the variance term  $h_t$  is obtained recursively using the equality in (1b). Note that explanatory variables can be added in (1) and the likelihood formulation in (5) is still valid, albeit with minor changes (see e.g.[35]).

In order to obtain the parameter estimates, the likelihood in (5) is often maximized using gradient search methods. Despite the simplicity of the likelihood formulation, maximizing this function can be cumbersome due to the nonlinearities in the model structure, and hence the possibility of local maxima and multiple modes [36]. A common practice is to get robust estimates using several starting values for the algorithm.

Extensions of the standard GARCH model are proposed in order to capture different dynamics of the observed series. In particular for the stock

returns series, the normality assumption in (1) is found to be restrictive. The normal conditional density of returns may fail to account for observations in the tails of the distribution and skewness in the observed series characterized by asymmetric effects of positive and negative past observations on current variance [36]. Two commonly used extensions are the Student- $t$  GARCH model [4] and the GJR GARCH model [3], which account for fat tails and asymmetric distributions, respectively. Despite several extended GARCH models, proposing a unifying one that can capture all dynamics of the observed series is often impossible.

Another extension of the models is the regime switching GARCH models. Such models are proposed since the relationship between the current return distribution and past returns' mean and variance can be complex [37], compared to the linear variance model assumed by the GARCH model [6, 38]. Introducing such a complex relationship using the Markov-switching structure ensures that the estimated volatility reacts quickly to changes in the volatility levels, and the forecast performance of the original GARCH model is improved [39]. These models propose  $K$  separate GARCH models:

$$\begin{aligned} y_t &= \sqrt{h_t} \epsilon_t \sim NID(0, h_t), \\ h_t &= h_{t,k}, \text{ if } s_t = k, \quad k = 1, \dots, K, \\ h_{t,k} &= \alpha_{0,k} + \sum_{i=1}^q \alpha_{i,k} y_{t-i}^2 + \sum_{j=1}^p \beta_{j,k} h_{t-j}, \text{ for } k = 1, \dots, K, \\ \epsilon_t &\sim NID(0, 1), \end{aligned} \quad (6)$$

where  $s_t$  denotes the realization of the state at time  $t$ , and is characterized by a Markov process. Despite the flexibility of allowing different GARCH models in different time periods, regime switching GARCH models can still be restrictive, as each observation is assumed to belong to a single regime at each period in time. The proposed fuzzy GARCH model does not have this constraint.

Apart from the above mentioned fully probabilistic extensions of the GARCH model, we focus our attention on fuzzy GARCH models as presented in [11, 13, 14, 16, 17]. These type of models consist of a set of if-then rules, where the antecedent of each rule are fuzzy sets and the consequents are GARCH models, consisting of  $l$ -th rules [11, 13, 14, 16, 17]:

$$R_l : \text{If } \mathbf{x} \text{ is } F_l \text{ then } h_{t,l} = \alpha_{0,l} + \sum_{i=1}^q \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^p \beta_{j,l} h_{t-j}, \quad (7)$$

where  $\mathbf{x} \in \mathbb{R}^n$  is an input vector,  $F_l : X \rightarrow [0, 1]$  is a multidimensional fuzzy

set defined on a continuous sample space  $X$ . The output of this fuzzy model is presented as

$$y_t = \sqrt{h_t} \epsilon_t, \quad (8a)$$

$$h_t = \sum_{l=1}^L g_{t,l} h_{t,l}, \quad (8b)$$

where  $g_{t,l} = u_{t,l} / \sum_{l=1}^L u_{t,l}$  are normalized membership functions with  $u_{t,l} \geq 0$  for  $l = 1, \dots, L$ ,  $\sum_{l=1}^L u_{t,l} > 0$ , and by definition  $g_{t,l} \geq 0$  and  $\sum_{l=1}^L g_{t,l} = 1$ . The inference used for the output (8a) and (8b) is similar to the inference of a Takagi-Sugeno fuzzy model [40]. Although not clear in [13, 14, 16, 17], we assume that like in [11], when  $\mathbf{x}$  is  $n$ -dimensional,  $u_{t,l}$  is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.*,

$$u_{t,l}(\mathbf{x}) = u_{F_{l_1}}(x_1) \circ \dots \circ u_{F_{l_n}}(x_n), \quad (9)$$

where  $x_n$  is the  $n$ -th component of  $\mathbf{x}$  and  $\circ$  denotes a t-norm.

In our analysis of this model, we note that the combination of  $h_{t,l}$  in (7) provides the *unobserved* conditional variance  $h_t$ . The density of output  $y_t$  is based on  $h_t$ :

$$y_t \mid h_t, x_t \sim NID(\mu, h_t). \quad (10)$$

In [11], the parameters of the model in (7) were estimated in a two step approach. First the antecedents were obtained using a fuzzy clustering heuristic, followed by the estimation of the GARCH parameters using maximum likelihood estimation. The chosen explanatory variable was the return at the previous period,  $y_{t-1}$ . For a simulated nonlinear GARCH model, good results are reported using as variance term  $h_{t,l}$  in (7) a GARCH(3,3) model or by constraint of  $\beta_j = 0$  using a GARCH(0,5) model. For the real data example [11], the conditional variance is not given by a GARCH model but it is considered to be constant over time  $h_{t,l} = h_l, \forall t$ , which gives

$$R_l : \text{If } y_{t-1} \text{ is } F_l \text{ then } h_{t,l} = h_l. \quad (11)$$

In [13, 14], the parameters of the fuzzy GARCH model are obtained using a genetic algorithm, while in [16] particle swarm optimization is used.



The objective function  $E_1$  is defined as the mean squared error between the estimated output density  $y_t^* = \sqrt{h_t} \epsilon_t$  and observation  $y_t$ , as

$$E_1 = \sum_{t=1}^T (y_t^* - y_t)^2. \quad (12)$$

To the best of our knowledge, the calculation of this objective function as a difference between a density function and a point is not possible. We further note that the same objective function is used in [15], although the rule base model is different. Furthermore, in [17] a similar objective function, based on the squared returns, is defined:

$$E_2 = \sum_{t=1}^T (y_t^{*2} - y_t^2)^2, \quad (13)$$

where the difference between the square of the implied density of returns  $y_t^{*2}$  and the point squared returns  $y_t^2$  is again not possible. We further note that using squared daily returns as a comparison to a model is not appropriate, because squared daily returns provide a poor approximation of realized daily volatility [41]. More accurate results can be obtained using the sum of squared intraday results [22, 23, 24, 25].

Despite the aforementioned issues regarding the explanation for the parameter estimation of these fuzzy GARCH models in the literature, the general idea of these models is very appealing. They possess the advantage of the linguistic interpretation of the rules and the possibility to explain the conditional data distribution with antecedent variables  $\mathbf{x}$ . Despite this general modeling idea, the model output as defined in (10) is restricted to a normal conditional distribution with zero mean. This restriction may not accommodate documented dynamics of data, such as the existence of extreme observations or skewness in several stock returns series. For this reason, we propose a more flexible fuzzy GARCH model in Section 3 which can be estimated using a maximum likelihood approach.

### 3. Proposed Fuzzy GARCH model

In this paper we present a new flexible fuzzy GARCH model [33], named FGARCH( $L, p, q$ ) where the output  $y_t$  and conditional variance  $h_t$  are defined

by each of  $l$ -th fuzzy rule

$$R_l : \text{If } \mathbf{x} \text{ is } F_l \text{ then } y_{t,l} \mid x_t, h_{t,l} \sim NID(\mu_l, h_{t,l}), \quad (14a)$$

$$\text{with } h_{t,l} = \alpha_{0,l} + \sum_{i=1}^q \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^p \beta_{j,l} h_{t-j}, \quad (14b)$$

where  $h_t$  is given by (8b) and the fuzzy sets  $F_l$  are defined by membership functions  $u_{t,l}$  as function of the antecedent variable  $x_t$ .

For the  $l$ -th fuzzy rule, the output consequents are defined by a GARCH( $p, q$ ) model which has a normal distribution with mean  $\mu_l$  and variance  $h_{t,l}$ , with parameters given by:

$$\begin{aligned} y_{t,l} &= \sqrt{h_{t,l}} \epsilon_t, \\ h_{t,l} &= \alpha_{0,l} + \sum_{i=1}^q \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^p \beta_{j,l} h_{t-j}, \\ h_{t-j} &= \sum_{l=1}^L g_{t-j,l} h_{t-j,l}, \text{ for } j = 1 \dots, p. \end{aligned} \quad (15)$$

where  $g_{t,l} = u_{t,l} / \sum_{l=1}^L u_{t,l}$  is the normalized membership function.

The output of this FGARCH model is:

$$y_t \mid h_t, x_t \sim \sum_{l=1}^L g_{t,l} NID(\mu_l, h_{t,l}), \quad (16)$$

which can be interpreted as a fuzzy combination of normal densities. Depending on the normalized membership functions  $g_{t,l}$ , the output has several distributional forms, such as a normal density, a skewed density or a bimodal density. Comparing the output of (16) and the output (8), it is clear that the outputs follow different inference mechanisms. In (8) the output is a probabilistic normal distribution with zero mean and the variance modelled with a fuzzy system. Conversely, the output (16) combines probabilistic and fuzzy uncertainty, resulting in a combination of normal distributions dependant on a set of fuzzy rules. The model defines the whole output density including the mean and variance.

The output of the proposed model has a proper conditional distribution, similar to a finite mixture of normal densities, under the condition that mem-

bership values satisfy

$$g_{t,l} \geq 0, \forall l, t \quad (17a)$$

$$\sum_{l=1}^L g_{t,l} = 1, \forall t. \quad (17b)$$

These conditions ensure that the probability density, hence the likelihood of observation  $t$  can be written conditional on past observations and past variance.

A second concern in the proposed FGARCH( $L, p, q$ ) model is to obtain positivity and stationarity conditions  $h_{t,l}$  for every rule and at every time period, since the output of the rules in (14a) are not defined otherwise. A sufficient condition for this is to incorporate standard GARCH model conditions for each rule  $l = 1, \dots, L$  in the model:

$$\alpha_{0,l} > 0, \alpha_{i,l} \geq 0, \beta_{j,l} \geq 0, \sum_{i=1}^q \alpha_{i,l} + \sum_{j=1}^p \beta_{j,l} < 1, i = 1, \dots, q, j = 1, \dots, p, \quad (18)$$

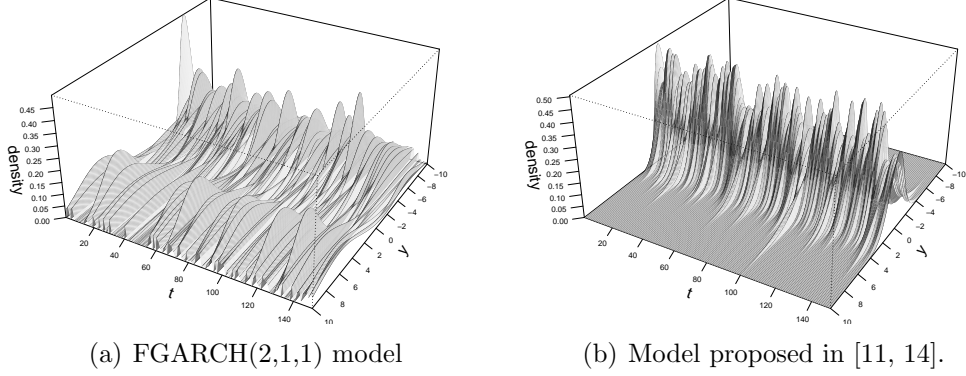
Note that these conditions should also hold for the fuzzy GARCH models proposed in the literature [11, 13, 14, 16, 17], although they have not been explicitly considered.

### 3.1. Interpretation of the model

Intuitively, the conditional distribution of the proposed model in (16) is a smooth combination of normal distributions. This combined density is similar to a finite mixture of normal densities, with combinations relying on the antecedent variables. The estimation of the proposed model, however, is more straightforward and the linguistic interpretation provided by this model is unique.

In relation to the previous fuzzy GARCH models in the literature, the proposed model is more general and can capture several different dynamics of data: In (14a) and (14b), output  $y$  shows a smooth transition between normal densities, with possible different mean and variances. Hence the density of each observation might be multimodal or skewed, while in the previous fuzzy GARCH models the output density in (10) is a unimodal and symmetric normal density. In the proposed model, the combination of normal densities in the rule output can lead to unimodal or skewed distributions depending on model parameters:

Figure 1: Conditional distributions of simulated data from fuzzy GARCH models.



1. If  $\mu_l = \mu_{l^*}$  for all  $l, l^* \in \{1, \dots, L\}$ , output  $y$  comes from a normal distribution and conditional variance  $h$  changes over time. This case leads to the previous fuzzy GARCH models as defined in (10).
2. If mean parameters  $\mu_l$  are relatively different and  $h_{t,l}$  are relatively small and similar across  $l = \{1, \dots, L\}$ , output distribution is likely to be multimodal.
3. If mean parameters  $\mu_l$  are relatively close to each other and  $h_{t,l}$  are relatively different across  $l = \{1, \dots, L\}$ , output distribution is likely to be skewed.

We illustrate the difference between the fuzzy GARCH model defined in [11, 14] and the proposed FGARCH( $L, p, q$ ), using simulated data. Figure 1 shows the conditional density of output  $y$  for simulated data from the proposed model (14a) and model (7). In this example, both models have two rules with Gaussian membership parameters  $\{c_{1,1}, s_{1,1}^2, c_{1,2}, s_{1,2}^2\} = \{-2.3, 2.5, 1, 1\}$  defined in (20) and GARCH parameters defined for each rule  $\{\alpha_{0,1}, \alpha_{1,1}, \beta_{1,1}, \alpha_{0,2}, \alpha_{1,2}, \beta_{1,2}\} = \{0.5, 0.25, 0.17, 1.0, 0.50, 0.33\}$ . In the previous fuzzy GARCH model [11, 14]  $\mu = 0$ , while in the FGARCH(2,1,1) model  $(\mu_1, \mu_2) = (-6, 6)$ .

The model in (7) leads to unimodal and symmetric conditional densities while simulated data from the proposed model has a more complex behavior with skewed, asymmetric and bimodal conditional densities.

The proposed FGARCH model makes a clear distinction between linguistic and probabilistic uncertainty. The fuzziness or linguistic vagueness is present in the antecedent of each rule and their combination. By using

fuzzy sets to represent linguistic vagueness, the output density is allowed to vary smoothly, in mean and variance, over time. These smooth changes are related to linguistic labels [42], belonging to one or several fuzzy sets at the same time. The linguistic labels can be used to explain complex systems, such as financial markets [43], with imprecise descriptions of phenomena in a similar way humans do it. Following the concept of granularity [44, 45], a fuzzy linguistic label can be viewed as a set of observation values grouped according to some criteria, in an environment of imprecision, uncertainty and partial truth [46], where each linguistic label has a degree of validity. The probabilistic uncertainty can be captured by the GARCH model. In this extensively studied and good performing model [27], the conditional density of the data is a normal distribution with time varying variance depending on past variance and past observations.

Fuzzy GARCH models can be related to finite mixture of GARCH models. Standard mixture of GARCH models allocate each observation to one GARCH model at a time and the probability of each GARCH model is fixed. More general mixture GARCH models can have smoothly varying regime probabilities [47]. In these models, each observation is allocated to different GARCH models, depending on the regime probabilities explained by explanatory variables. The fuzzy GARCH model also uses such antecedent variables, but in this case the uncertainty is modelled using fuzzy sets, relaxing the restriction of realizing one state at each observation. Even if the mathematical formulation is similar, the interpretation and underlying modeling of uncertainty (fuzzy and probabilistic) is different from mixture GARCH models (probabilistic only).

### 3.2. Parameter estimation

It is possible to estimate the model in (14a) using the maximum likelihood method, given that  $x$  is *predetermined* with respect to  $y$ , i.e. input  $x_t$  is included in the information set at time  $t - 1$ . More specifically,  $x_t$  can for instance take past  $y$  values or can be an exogenous variable.

Given that the type and number of membership functions  $g_{t,l}$  are known, the log-likelihood of data  $y = \{y_{t^*}, \dots, y_T\}$  is:

$$\ln \ell(y \mid I_{t-1}) = \ln \prod_{t=t^*}^T \ell(y_t \mid x_t, h_t) = \sum_{t=t^*}^T \ln \left( \sum_{l=1}^L g_{t,l} \phi(y_t; \mu_l, h_{t,l}) \right), \quad (19)$$

where  $h_t$  is calculated from (14b),  $t^* = \max(p + q) + 1$  and initial variances  $\{h_1, \dots, h_{t^*-1}\}$  are assumed to be known. In practice,  $\{h_1, \dots, h_{t^*-1}\}$  can be set as the unconditional data variance.

In order to calculate the likelihood in (19) it is necessary to specify suitable membership functions, that satisfy conditions (17a) and (17b). In this work, the FGARCH models considered use Gaussian membership functions of the form:

$$u_{t,l} = u_{t,l}(x_t) = \prod_{k=1}^n \exp \left( -\frac{1}{2} \frac{(x_{kt} - c_{k,l})^2}{s_{k,l}^2} \right). \quad (20)$$

These membership functions were chosen because they naturally satisfy conditions (17a) and (17b) since  $g_{t,l} = u_{t,l} / \sum_{l=1}^L u_{t,l}$ ,  $u_{t,l} \geq 0$  for  $l = 1, \dots, L$ ,  $\sum_{l=1}^L u_{t,l} > 0$ . This reduces the need for additional parameter constraints in the gradient search optimization of the maximum log-likelihood estimation (21).

The parameter estimates can be obtained by maximizing the log-likelihood in (19), using gradient search methods. We maximize the log-likelihood with respect to the GARCH parameters  $\theta_{g,l} = \{\alpha_{0,l}, \alpha_{1,l}, \dots, \alpha_{q,l}, \beta_{1,l}, \dots, \beta_{p,l}\}$ , the output mean for each rule  $\mu_l$  and the membership function parameters  $\theta_{u,l}$ , simultaneously. The optimization problem can be defined as:

$$\begin{aligned} \underset{\mu_l, \theta_{g,l}, \theta_{u,l}}{\text{minimize}} \quad & -\log \ell(y \mid I_{t-1}) = -\sum_{t=t^*}^T \ln \left( \sum_{l=1}^L g_{t,l} \phi(y_t; \mu_l, h_{t,l}) \right) \\ \text{subject to} \quad & \alpha_{0,l} > 0, \alpha_{i,l} \geq 0, \beta_{j,l} \geq 0, i = 1, \dots, q, j = 1, \dots, p, \\ & \sum_{i=1}^q \alpha_{i,l} + \sum_{j=1}^p \beta_{j,l} < 1, i = 1, \dots, q, j = 1, \dots, p, \\ & c_l \leq c_{l+1}, l = 1, \dots, L-1, \end{aligned} \quad (21)$$

The first two restrictions ensure a positive variance at each time period, while the last restriction ensures that the membership functions cannot permute labels. The optimization method chosen approximates the whole output density instead of a proxy for the density, such as the mean or variance of the process, lending itself to density estimation. This is also the conventional method to obtain standard GARCH and mixture GARCH models' parameter estimates. We constrain the search space to solutions satisfying the positive

variance condition and membership functions that cover the universe of the input variables in the antecedent space.

We acknowledge that the proposed maximum likelihood estimation of the FGARCH( $L, p, q$ ) parameters has a possible disadvantage of local maxima, similar to standard and extended GARCH models. The problem of local optima is often more pronounced in highly parametrized models. For this reason, we concentrate on a FGARCH( $L, 1, 1$ ) model in the remaining of this paper. The simple parametrization of the underlying GARCH model is also based on the findings that a GARCH(1,1) model is very hard to beat in practice [27]. One exception to this is the asymmetric GARCH models [26] which are naturally considered in a FGARCH( $L, 1, 1$ ) model.

#### 4. Examples: Synthetic data parameter estimation

In this section we illustrate the performance of the proposed FGARCH model and discuss the estimation issues using a known data generating process to simulate data. Doing so, allows us to study the approximation capabilities of the FGARCH model, *i.e.* recover the same density function. It also shows the sensitivity to the initialization of the maximum likelihood estimation procedure, as explained in Section 3.2, on the FGARCH model.

We consider two sets of simulated datasets. First, we consider data simulated from the previously studied fuzzy GARCH models defined by (7) and (8), which restricts the output density to a normal conditional distribution with zero mean, used *e.g.* in [16]. We show that the FGARCH model proposed in this paper can correctly capture the properties of this data. Second, we consider simulated data from the proposed FGARCH model defined by (14a) and (14b). In both cases, we simulate 3000 data points from the model considered for  $L = 2$  and  $L = 3$  rules. We maximize the log-likelihoods with respect to the GARCH parameters  $\theta_{g,l} = \{\alpha_{0,l}, \alpha_{1,l}, \dots, \alpha_{q,l}, \beta_{1,l}, \dots, \beta_{p,l}\}$ , the output mean for each rule  $\mu_l$ , and the Gaussian membership parameters  $\theta_{u,l} = \{c_l, s_l^2\}$  for  $l = 1, \dots, L$ , simultaneously, as defined in (21).

Given the number of model parameters, a straightforward approach to decrease the possibility of obtaining local optimum is to consider several initializations for parameter estimation and choose the best model. This will also show the sensitivity of the optimization procedure on the proposed fuzzy GARCH model to the initialization. For all estimations considered, we estimate model parameters starting from 100 random initial points. From these

Table 1: Simulated FGARCH model data with constant mean: percentage of observations in respective distribution tails.

Simulated data from FGARCH(2,1,1) model					
	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
mean	0.010	0.047	0.098	0.201	0.397
90%	(0.009, 0.013)	(0.045, 0.047)	(0.090, 0.100)	(0.194, 0.202)	(0.388, 0.399)
Simulated data from FGARCH(3,1,1) model					
	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
mean	0.010	0.047	0.104	0.206	0.394
90%	(0.008, 0.011)	(0.044, 0.049)	(0.102, 0.106)	(0.203, 0.209)	(0.389, 0.399)

repetitions, the estimation providing the highest likelihood value is considered the global maximum and labelled as ‘best’. To provide an indication of different local minima and its effect in models’ performance, we report the average estimates, for both the model parameters and distribution tails, as well as the 90% interval of the estimates around the average value.

#### 4.1. Fuzzy GARCH data with constant mean

In this section, we use a FGARCH model, where the output density is restricted to a normal conditional distribution with zero mean, as the data generating process. This model is equivalent to the previously studied fuzzy GARCH models defined by (7) and (8). We estimate the parameters of the proposed FGARCH model defined by (14a) and (14b), without any restrictions to the output mean and distribution.

To study the approximation capabilities of the FGARCH model we compare the true and estimated data densities. The fuzzy GARCH model provides an estimated output density. We compare the quantiles of this estimated density and the percentage of simulated data points corresponding to each quantile. For a good approximation of the output density, the quantiles of this estimated density should match with the quantiles of the data, *e.g.* 5% of the actual observations should fall in the 5% tail of the output density. The mean estimates and 90% intervals of the quantiles  $\hat{\tau}(c)$  for  $c = 1\%, 5\%, 10\%, 20\%, 40\%$  for the simulated datasets are reported in Table 1, for FGARCH(2,1,1) and FGARCH(3,1,1) models.

Table 1 shows that the percentage of observations in estimated tails of the data distribution are close to the theoretical values, although for some quantiles the percentage of observations in some tails are smaller than the theoretical value. This may be a consequence of estimating this model with



Table 2: Parameter estimates and true values for simulated data from a FGARCH( $L,1,1$ ) models restricted to a normal conditional distribution with zero mean. 90% intervals from 100 random initializations are given in parentheses.

FGARCH(2,1,1)					
	$l = 1$		$l = 2$		
	value	estimate	value	estimate	
$\mu$	0.00	0.03 (-0.02, 0.04)	0.00	-0.02 (-0.03, 0.02)	
$\alpha_0$	0.50	0.47 (0.42, 0.66)	1.00	1.00 (0.62, 1.01)	
$\alpha_1$	0.25	0.29 (0.26, 0.53)	0.50	0.58 (0.41, 0.65)	
$\beta_1$	0.17	0.13 (0.13, 0.31)	0.33	0.34 (0.16, 0.34)	
$c$	-3.40	-0.60 (-1.28, 0.12)	3.20	1.21 (0.31, 2.11)	
$s^2$	1.00	0.54 (0.00, 0.82)	1.00	0.48 (0.19, 0.77)	
FGARCH(3,1,1)					
	$l = 1$		$l = 2$		$l = 3$
	value	estimate	value	estimate	value estimate
$\mu$	0.00	-0.02 (-1.09, 0.80)	0.00	-0.09 (-0.82, 0.63)	0.00 0.03 (-0.46, 0.53)
$\alpha_0$	0.50	0.57 (0.00, 1.09)	1.00	0.79 (0.37, 1.23)	1.50 0.86 (0.40, 1.18)
$\alpha_1$	0.17	0.32 (0.00, 0.69)	0.33	0.38 (0.18, 0.62)	0.50 0.29 (0.15, 0.43)
$\beta_1$	0.11	0.23 (0.00, 0.56)	0.22	0.18 (0.00, 0.51)	0.33 0.21 (0.03, 0.50)
$c$	-3.40	-1.07 (-2.59, -0.02)	0.01	-0.20 (-0.60, 0.27)	3.20 0.51 (-0.44 2.95)
$s^2$	1.00	0.48 (0.00, 0.98)	1.00	0.52 (0.09, 0.76)	1.00 0.67 (0.15, 1.43)

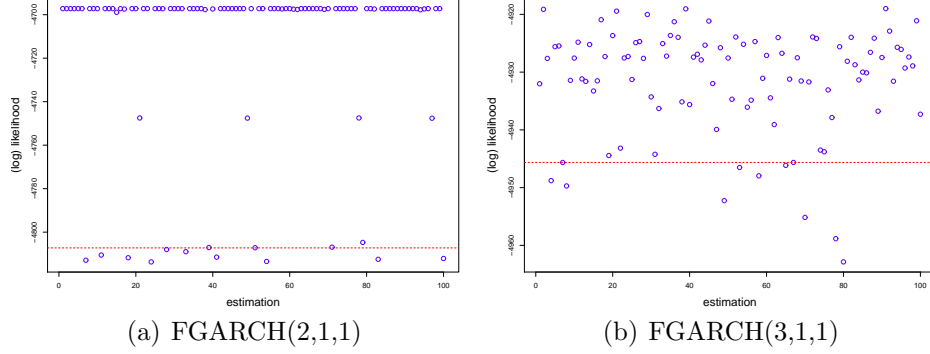
only 3000 data points, which may not be enough to represent the true distribution. A detailed analysis of the effect of the number of observations on the true and estimated tails of the density is left for future research.

Table 2 presents the true parameter values, together with mean estimates and 90% interval of parameter estimates (in parentheses) from 100 parameter estimates with different starting values for optimization.

As Table 2 shows, the effect of the initial points on parameter estimation are not negligible. Estimates of the GARCH parameters ( $\alpha_0, \alpha_1, \beta_1$ ) and rule output means  $\mu_l$  are close to the true parameter values and the 90% interval regardless of the initial points. Estimates of the membership parameters  $c, s^2$  on the other hand, deviate much more from the original values and are more affected by their initialization. It is interesting to note that the overall fit of the FGARCH model is not substantially affected with completely random initializations of all parameters, as the 90% intervals show in Table 1.

Since the output distribution from the FGARCH is a combination of GARCH models with different means, through a set of fuzzy rules, the output variance is jointly captured by the unobserved (estimated) variance and the fuzzy antecedents. The obtained models are highly nonlinear and the

Figure 2: FGARCH model with 0 mean: Log-likelihood values form different starting values, and 90% lower band for obtained log-likelihood



optimization method includes constraints on GARCH parameters, hence obtaining a local maximum is likely. Local maxima is less problematic for the GARCH model parameters since this part of the model has a structure to explain part of the unobserved variances, given by the fuzzy antecedent. Due to this model structure, different parameter values typically lead to very different unobserved variances and the estimation of these parameters is not affected severely by the initial points. We note that other GARCH models also suffers from similar initialization issues. The fuzzy parameters, on the other hand, are more susceptible to random initialization. In the FGARCH model it is possible that different fuzzy membership parameters lead to similar output density approximations. Hence the optimization of the fuzzy membership parameters are more sensitive to the initial points for optimization.

Fig. 2 shows the optimal log-likelihood values for the 100 different estimations performed, together with the 90% lower bound for these values. We observe that for the FGARCH(2,1,1) model, approximately the same maximum log-likelihood value is obtained in most of the estimations, despite the differences in parameter estimates, caused by the random initialization. For this case, the maximum variation of the log-likelihood is approximately 2.48%. A similar result can be observed for the FGARCH(3,1,1) model. The max variation for the log-likelihood is approximately 0.09%, but the variation around the maximum log-likelihood value is smaller for the FGARCH(2,1,1). Hence the local optima issue, particularly in the fuzzy membership parameters,

Table 3: Simulated FGARCH model data with time varying mean: percentage of observations in respective distribution tails.

Simulated data from FGARCH(2,1,1) model					
	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
mean	0.010	0.047	0.098	0.201	0.397
90%	(0.009, 0.013)	(0.045, 0.047)	(0.090, 0.100)	(0.194, 0.202)	(0.388, 0.399)
Simulated data from FGARCH(3,1,1) model					
	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
mean	0.011	0.046	0.100	0.198	0.391
90%	(0.009, 0.012)	(0.042, 0.049)	(0.096, 0.105)	(0.191, 0.207)	(0.384, 0.407)

does not substantially affect the maximized likelihood. We conjecture that the smaller maximum variation of the log-likelihood for the FGARCH(3,1,1), when compared with the FGARCH(2,1,1) model, stems from the higher number of rules and consequent overlap between them, that leads to the almost same result.

#### 4.2. Fuzzy GARCH data with general time varying mean

In this section, we simulate data from FGARCH(2,1,1) and FGARCH(3,1,1) models without any restrictions on the output density and perform 100 estimations with different starting values for optimization. Table 3 shows the mean estimates and 90% intervals of the quantiles  $\hat{\tau}(c)$  and Table 4 presents the true parameter values, together with mean estimates and 90% interval of parameter estimates (in parentheses).

Similar to the results in Section 4.1, we observe that the estimated output densities capture the tails of the distribution, as Table 3 shows. Furthermore, as it is presented in Table 4, parameter estimates of the GARCH model are less affected by initialization compared to the estimated fuzzy membership parameters. Nonetheless, we note that there is a larger variation in the obtained 90% intervals for the parameter estimates than obtained in Section 4.1. The initial points have a bigger effect in the parameter estimates, although, like previously, the overall fit of the FGARCH model is similar throughout the experiments. This difference is expected, since the FGARCH models is now capturing more complex data properties, such as time-varying mean and variance, skewness and bimodality. In this experiment, the parameters of the FGARCH(3,1,1) model vary more with the initialization than those of the FGARCH(2,1,1) model, as the 90% intervals shown in Table 4 indicate.

The problem of local optima can be a more severe problem when the

Table 4: Parameter estimates and true values for simulated data from a FGARCH(L,1,1) models with  $L = 2$  and  $L = 3$  rules for time varying mean. 90% intervals from 100 random initializations are given in parentheses.

FGARCH(2,1,1)					
	$l = 1$		$l = 2$		
	value	estimate	value	estimate	
$\mu$	-2.00	-0.51 (-1.65, 1.71)	2.00	0.67 (-1.86, 1.71)	
$\alpha_0$	0.50	0.70 (0.60, 1.05)	1.00	0.89 (0.61, 0.86)	
$\alpha_1$	0.25	0.28 (0.19, 0.45)	0.50	0.37 (0.15, 0.45)	
$\beta_1$	0.17	0.23 (0.18, 0.36)	0.33	0.29 (0.10, 0.36)	
$c$	-3.40	-1.84 (-3.73, -0.30)	3.20	1.31 (-1.20, 3.84)	
$s^2$	1.00	0.76 (0.15, 1.93)	1.00	0.49 (0.14, 0.89)	
FGARCH(3,1,1)					
	$l = 1$		$l = 2$		$l = 3$
	value	estimate	value	estimate	value estimate
$\mu$	-2.00	-0.19 (-1.18, 1.14)	0.00	-0.05 (-0.82, 0.97)	2.00 0.72 (-0.45, 1.64)
$\alpha_0$	0.50	0.49 (0.00, 0.82)	1.00	0.73 (0.15, 0.87)	1.50 0.70 (0.50, 0.90)
$\alpha_1$	0.17	0.24 (0.10, 0.50)	0.33	0.29 (0.19, 0.43)	0.50 0.24 (0.13, 0.38)
$\beta_1$	0.11	0.19 (0.00, 0.47)	0.22	0.20 (0.00, 0.34)	0.33 0.23 (0.14, 0.34)
$c$	-3.40	-1.12 (-3.05, 0.55)	0.01	0.04 (-0.84, 0.81)	3.20 1.29 (-0.59, 3.81)
$s^2$	1.00	0.39 (0.00, 0.80)	1.00	0.46 (0.02, 0.77)	1.00 0.60 (0.19, 1.01)

number of parameters (*e.g.* the number of fuzzy rules or parameters  $p, q$  of GARCH models) increases. Nonetheless, the FGARCH model proposed in this paper achieves good approximation properties, as Table 1 and Table 3 show, even with a small number of parameters, as the models described in these experiments.

## 5. Application: Conditional density estimation of S&P500 returns

In this section, we apply the proposed fuzzy GARCH model to build a conditional density model of S&P 500 returns. This stock market index is based on the market capitalizations of 500 companies publicly traded in the U.S. stock market, as determined by Standard & Poor's. It is considered as an indicator of U.S. equities reflecting the risk and return characteristics of the large capital universe. Conditional density estimation used to study financial market volatility has an important role in financial economics and is at the heart of several subjects, including asset allocation, market timing, risk management, the pricing of assets and portfolio management [48]. Many statistical quantiles such as Value-at-Risk or Expected Shortfall, which are

directly linked to the tail of the return distribution of a portfolio of financial assets, are widely accepted financial risk management tools [2].

In this paper, the proposed FGARCH model is applied to 3718 observations of S&P 500 returns from February 18, 1997 to November 23, 2011, calculated as percentage changes of daily closing prices. The training and forecast samples are the first 3218 and the last 500 observations (approximately 2 trading years) and are presented in Fig. 3. In the period considered, it is possible to observe periods of volatility changes and extreme returns, indicating non-trivial statistical properties, such as asymmetric distributions and non-constant variability of returns.

We consider conditional density estimation models for one period ahead forecasts. The proposed FGARCH models approximate the distribution of returns at time  $t + 1$  conditional on the returns at time  $t$ , through the GARCH-type relation and antecedent membership functions. This choice of the fuzzy rule antecedents provides a more complex and non-linear relationship between current returns and past returns than it is assumed by GARCH model. This antecedent variable allows for a linguistic interpretation of different data dynamics on the current returns' conditional density. By using this variable as antecedent, the FGARCH model allows the analysis of other stylized facts, such as volatility clustering and leverage effects. Volatility clustering [49] is considered as the tendency of large changes to be followed by large changes, of either sign, and small changes to be followed by small changes. The leverage effect [50] refers to the asymmetric relation between lagged unexpected returns and volatility, where it is observed that negative return tends to increase subsequent volatility much more than a positive return of the same magnitude. FGARCH( $L,1,1$ ) models with  $L = 2$  and  $L = 3$  rules are estimated, and the results are compared with a standard GARCH(1,1) model. Model performances are assessed by comparing the quantiles  $\hat{\tau}(c)$  of the estimated distribution with the theoretical distribution quantiles  $\tau(c)$ . Each model estimation was repeated 100 times with different initial points. This method allows us to choose the best parameter estimates, which lead to the maximum likelihood value from different initializations.

Table 5 presents the estimated quantiles  $\hat{\tau}(c)$  of the training and forecast samples for the GARCH and FGARCH models. In this table we report the percentage of observations that are included in each  $\tau(c)$  quantile, best and average quantile estimates and the 99% intervals (in parentheses). The best estimates are based on the estimation providing the highest log-likelihood value. This can be seen as the estimation providing the global optimum.

Figure 3: S&P 500 returns in percentages

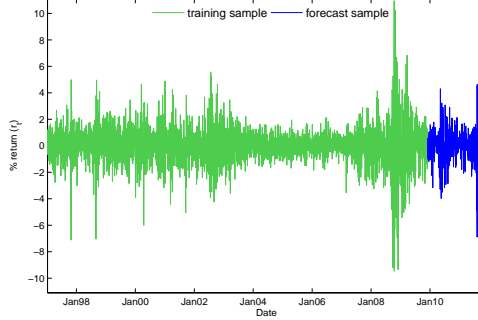


Figure 4 shows the forecast sample and the estimated conditional density quantiles from the GARCH and FGARCH models.

Table 5 shows that the percentage of observations in the respective tails of the returns distribution are close to the true values in most cases. The three models we consider lead to different conditional density estimates. For the distribution quantiles  $\tau(c)$  with  $c = 10\%, 20\%, 40\%$ , FGARCH models perform better than the GARCH model, as the estimated values are closer to the theoretical values. The difference between the FGARCH(2,1,1) and FGARCH(3,1,1) is very small. From the results obtained in Section 4, this small difference is expected since the FGARCH model can capture complex data structures with a small number of rules. For the quantile  $\tau(c)$  with  $c = 1\%, 5\%$  level, all models overestimate this quantile, specially in the forecast sample. This may indicate that the estimated models capture the extreme returns in the training set, thus resulting in more conservative models. For the FGARCH model, a more complex antecedent set including information on past returns and other variables may overcome this issue [51, 52, 32]. This topic is left for future research. Despite this overestimation in the tails, by visually inspecting Fig. 4, we can observe that the FGARCH models' density estimation quickly adapts to changes in the returns. This can be observed, for instance in periods of low returns, where the conditional density obtained by the FGARCH model is closer to the observed returns than those of the GARCH model, indicating a decrease in market risk. The added value of the fuzzy GARCH models are clear in periods of sudden decrease of volatility, for example around October 2010. The standard GARCH model cannot capture these low volatility periods as good as the fuzzy GARCH models. These results are in line with the findings of [30, 32], who show that the

Table 5: Estimated quantiles for the S&P500 data. From 100 estimations with random starting values, we report the percentage of observations at each tail of the distribution according to the average - best estimation and 99% confidence intervals (in parentheses) are reported for each model.

Standard GARCH(1,1)*					
sample	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
<i>training</i>	1.6 (1.6, 1.6)	5.0 (5.0, 5.0)	9.6 (9.6, 9.6)	18.3 (18.3, 18.3)	36.3 (36.3, 36.3)
<i>forecast</i>	2.4 (2.4, 2.4)	6.4 (6.4, 6.4)	9.8 (9.8, 9.8)	17.0 (17.0, 17.0)	32.4 (32.4, 32.4)
FGARCH(2,1,1)					
sample	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
<i>training</i>	1.5 - 1.3 <sup>b</sup> (1.3, 1.7)	5.4 - 5.9 <sup>b</sup> (5.0, 6.0)	10.1 - 10.6 <sup>b</sup> (9.6, 10.7)	18.7 - 19.4 <sup>b</sup> (18.3, 19.4)	36.9 - 37.5 <sup>b</sup> (36.4, 37.6)
<i>forecast</i>	2.5 - 2.6 <sup>b</sup> (2.2, 3.0)	6.6 - 7.0 <sup>b</sup> (6.2, 7.2)	10.0 - 9.4 <sup>b</sup> (9.6, 11.0)	17.1 - 17.4 <sup>b</sup> (16.8, 17.6)	33.5 - 34.6 <sup>b</sup> (32.6, 34.6)
FGARCH(3,1,1)					
sample	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
<i>training</i>	1.5 - 1.5 <sup>b</sup> (1.3, 1.7)	5.6 - 5.9 <sup>b</sup> (5.2, 6.0)	10.2 - 10.6 <sup>b</sup> (9.8, 10.7)	18.8 - 19.5 <sup>b</sup> (18.3, 19.7)	37.1 - 37.3 <sup>b</sup> (36.5, 38.5)
<i>forecast</i>	2.6 - 2.2 <sup>b</sup> (2.2, 3.2)	6.5 - 7.2 <sup>b</sup> (6.0, 7.4)	10.3 - 9.8 <sup>b</sup> (9.8, 11.4)	17.1 - 17.4 <sup>b</sup> (16.6, 18.0)	33.9 - 34.4 <sup>b</sup> (32.8, 35.8)

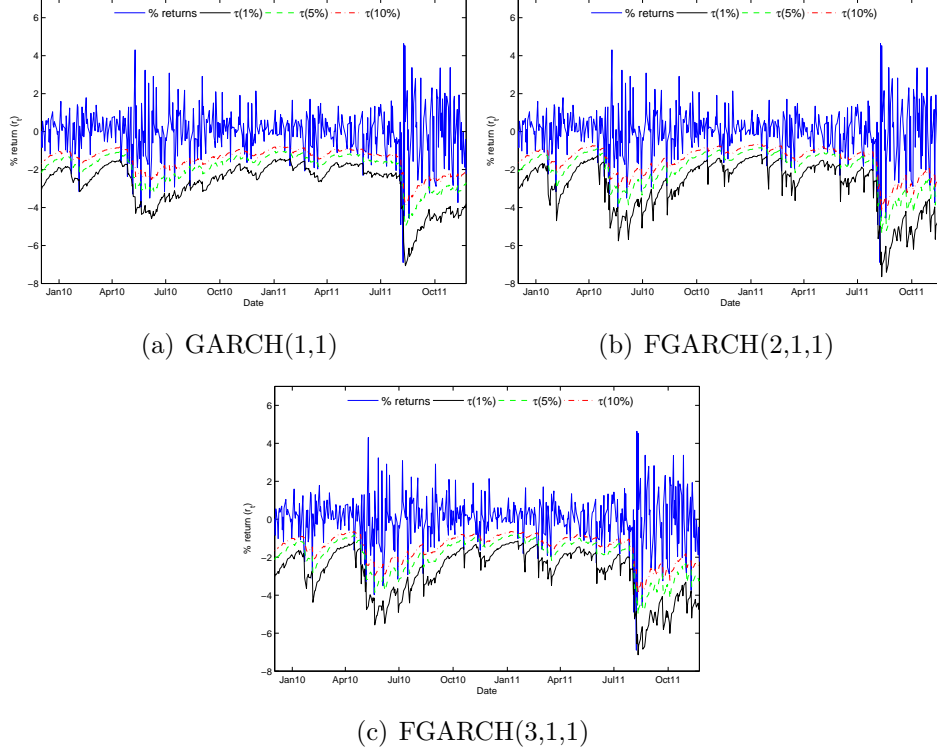
\* The differences in estimated values are negligible, with the reported digits.

<sup>b</sup> Best estimation providing the highest log-likelihood value.

standard GARCH model cannot capture such complex behavior.

The proposed FGARCH model provides a linguistic interpretation of the gradual changes in return density, producing a simple understanding of the underlying changes. From the 100 estimations with random starting values, the ‘best’ estimation result providing the highest log-likelihood value were selected to illustrate the model interpretation. In order to see the behavior of each individual GARCH model, we also report the unconditional standard deviation, calculated as the square root of (2). The rule-based of the

Figure 4: Quantile estimates for S&P 500 data, using the GARCH(1,1) model and FGARCH( $L,1,1$ ) models with  $L = 2$  and  $L = 3$  rules

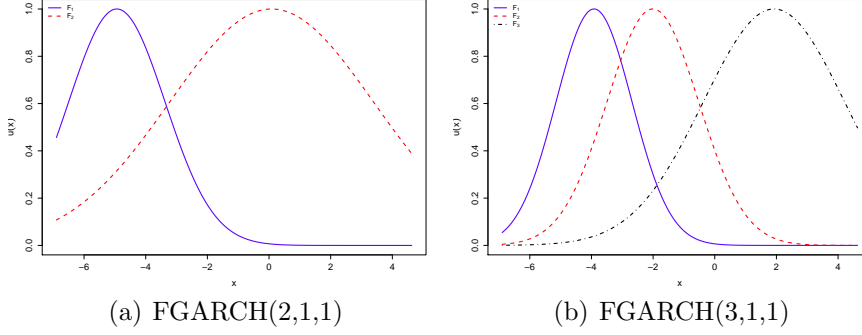


FGARCH(2,1,1) is given by

$$\begin{aligned}
 R_1 : & \text{ If } \mathbf{y}_{t-1} \text{ is } F_1 \text{ then } y_{t,1} \mid y_{t-1}, h_{t,1} \sim N(-0.255, h_{t,1}) \\
 & \text{ with } h_{t,1} = 3.247 + 0.029y_{t-1}^2 + 0.970h_{t-1}, \sqrt{\bar{h}_1} = 54.2645, \\
 & \text{ and } c_1 = -4.936, s_1^2 = 1.565, \\
 R_2 : & \text{ If } \mathbf{y}_{t-1} \text{ is } F_2 \text{ then } y_{t,2} \mid y_{t-1}, h_{t,2} \sim N(0.016, h_{t,2}) \\
 & \text{ with } h_{t,2} = 0.000 + 0.000y_{t-1}^2 + 0.922h_{t-1}, \sqrt{\bar{h}_2} = 0.0007 \\
 & \text{ and } c_2 = 0.083, s_2^2 = 3.306.
 \end{aligned} \tag{22}$$



Figure 5: Membership functions for S&P 500 data, using the GARCH(1,1) model and FGARCH( $L,1,1$ ) models with  $L = 2$  and  $L = 3$  rules



The rule-base model for the FGARCH(3,1,1) model is given by

$$\begin{aligned}
 R_1 &: \text{If } \mathbf{y}_{t-1} \text{ is } F_1 \text{ then } y_{t,1} \mid y_{t-1}, h_{t,1} \sim N(-0.102, h_{t,1}), \\
 &\quad \text{with } h_{t,1} = 2.768 + 0.019y_{t-1}^2 + 0.981h_{t-1}, \sqrt{h_1} = 199.9176 \\
 &\quad \text{and } c_1 = -3.916, s_1^2 = 1.234, \\
 R_2 &: \text{If } \mathbf{y}_{t-1} \text{ is } F_2 \text{ then } y_{t,2} \mid y_{t-1}, h_{t,2} \sim N(0.104, h_{t,2}), \\
 &\quad \text{with } h_{t,2} = 0.002 + 0.000y_{t-1}^2 + 1.000h_{t-1}, \sqrt{h_2} = 6.9384, \\
 &\quad \text{and } c_2 = -2.010, s_2^2 = 1.489, \\
 R_3 &: \text{If } \mathbf{y}_{t-1} \text{ is } F_3 \text{ then } y_{t,3} \mid y_{t-1}, h_{t,3} \sim N(-0.034, h_{t,3}) \\
 &\quad \text{with } h_{t,3} = 0.002 + 0.000y_{t-1}^2 + 0.887h_{t-1}, \sqrt{h_3} = 0.1421 \\
 &\quad \text{and } c_3 = 1.920, s_3^2 = 2.293.
 \end{aligned} \tag{23}$$

The estimated membership values for both models are presented in Fig. 5. For comparison purposes, the GARCH(1,1) is given by

$$\begin{aligned}
 y_t &\mid y_{t-1}, h_t \sim N(0.005, h_t), \\
 \text{with } h_t &= 0.013 + 0.077y_{t-2}^2 + 0.917h_{t-1}, \text{ and } \sqrt{h} = 1.433.
 \end{aligned} \tag{24}$$

The GARCH(1,1) model defines a normal distribution with changing variances for the return series, while the FGARCH models define separate GARCH models combined using the fuzzy antecedents. The standard GARCH model in (24) leads to a mean around 0 and an unconditional standard deviation of 1.433 for returns. The FGARCH(2,1,1) and FGARCH(3,1,1) models provide different means and unconditional standard

deviations for each rule, as well as volatility structure given by the each rules GARCH parameters.

In each rule, the different fuzzy sets combined with the unconditional volatility provides a clear indication of the presence of leverage effects. For the FGARCH(2,1,1) presented in (22), rule 1 shows that the unconditional volatility after negative returns is very high. This indicates that the effect of negative returns on variance is very high. This does not seem to be the case for positive returns, since the unconditional volatility of rule 2 is very low. As Fig. 5 shows, for values above 0 the effect of rule 2 is almost exclusive. For the FGARCH(3,1,1) rule 3 indicates that the effect of positive returns above 2 is the smallest one, while the unconditional volatility of very negative returns is very high. In both models, the GARCH parameter  $\alpha_1$  is lower than in the GARCH(1,1) model presented in (24), since the effect of the past returns in variance is already modeled by the fuzzy antecedents in the rule-base model. These results are in line with the findings [37].

An indication of the existence of volatility clustering and volatility persistence [22] can be related to the effect of past conditional volatility,  $\beta_{1,l}$  in each GARCH model. For both FGARCH models it is possible to observe that the effect of conditional volatility is larger than in the GARCH model (24), except for rule 3 of FGARCH(3,1,1). This rule indicates that after returns above 2 reduce the volatility in the next period. Rule 1 of FGARCH(3,1,1) model (23) captures extreme negative events followed by very high volatility the next day. Mean returns in these volatile days is also negative. Rule 2 of this model is very interesting, since it shows an almost absolute persistence in volatility, but as the fuzzy antecedents show in Fig. 5, this rule is always combined with the other two rules. Rule 1 of the FGARCH(2,1,1) model presented in (22), shows that low returns lead to a persistent effect in volatility. Rule 2 of this model, indicates a high effect of past volatility for returns above 0 but the persistence is lower than in rule 1. Although both models show good conditional density approximation capabilities, they provide different linguistic interpretations. Thus, for the considered application, the choice between models will depend on the desired level of linguistic interpretation.

It is interesting to note that in the above analysis of the FGARCH models, each rule was analyzed independently, providing different interpretations of the conditional density evolution. Despite the simple structure of the FGARCH model, the long run behavior indicates that the system will alternate between rules, leading to a complex non-linear dynamic behavior. In

the long run, the local volatility of the GARCH models defined in each rule will revert to its unconditional volatility. The FGARCH model, on the other hand, due to the fuzzy antecedents and the unconditional volatility defined by the GARCH structure, will not converge to a single unconditional volatility level, but instead vary between the unconditional volatility of each rule.

## 6. Conclusion

This paper studies the properties, estimation issues and interpretation of a new flexible fuzzy GARCH model for conditional density estimation. These models provide linguistic interpretation of the rules and the possibility to explain the conditional data distribution with antecedent variables  $\mathbf{x}$ . Furthermore, the use of GARCH models in rule outputs, allows the system to capture time dependency in the conditional distributions in a flexible way. Previous fuzzy GARCH models were restricted to a normal conditional distribution. This restriction may not accommodate the documented dynamics of data, such as the existence of extreme observations or skewness in several stock returns series. For this reason, we propose a more flexible fuzzy GARCH model. In this model, the distribution of the returns are allowed to vary in mean and variance smoothly over time, where the smooth changes are related to linguistic descriptors. We relate this model with existing fuzzy and probabilistic GARCH models and provide an interpretation of the model, from a statistical and fuzzy linguistic point of view. These models have the advantage that they can be estimated by maximizing a tractable likelihood function, which in turn overcomes the estimation issues appearing in pure probabilistic flexible GARCH models. Another advantage is that the model provides a linguistic interpretation of the smooth changes in return density, providing another view for understanding of the underlying changes. We illustrate the model capabilities using synthetic datasets exhibiting different data properties and real data on S&P 500 returns. We show that the proposed model is suitable for analysis of the returns distribution, as it captures the underlying data distribution in all cases we consider. In future work, we plan to extend the proposed model to include multiple outputs to capture the joint conditional distribution of several variables.

## References

- [1] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31 (1986) 307–327.
- [2] P. Jorion, *Value at Risk: the new benchmark for managing financial risk*, 3 ed., McGraw–Hill, New York, 2006.
- [3] L. Glosten, R. Jagannathan, D. Runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48 (1993) 1779–1801.
- [4] T. Bollerslev, A conditionally heteroskedastic time series model for speculative prices and rates of return, *The Review of Economics and Statistics* 69 (1987) 542–547.
- [5] G. González-Rivera, Smooth transition GARCH models, *Studies in Nonlinear Dynamics and Econometrics* 3 (1998) 61–78.
- [6] M. Haas, S. Mittnik, M. S. Paoletta, A new approach to Markov-switching GARCH models, *Journal of Financial Econometrics* 2 (2004) 493–530.
- [7] D. Ardia, N. Basturk, L. F. Hoogerheide, H. K. Van Dijk, A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihoods, *Computational Statistics & Data Analysis* 56 (2012) 3398–3414.
- [8] L. Bauwens, A. Preminger, J. Rombouts, Theory and inference for a Markov switching GARCH model, *The Econometrics Journal* 13 (2010) 218–244.
- [9] Z. He, J. Maheu, Real time detection of structural breaks in GARCH models, *Computational Statistics & Data Analysis* 54 (2010) 2628–2640.
- [10] R. Donaldson, M. Kamstra, An artificial neural network-GARCH model for international stock return volatility, *Journal of Empirical Finance* 4 (1997) 17–46.
- [11] A. A. Popov, K. V. Bykhanov, Modeling volatility of time series using fuzzy GARCH models, in: *9th Russian–Korean International Symposium on Science and Technology, KORUS2005*, 2005, pp. 687–692.

- [12] J.-C. Hung, F.-Y. Hsu, Adaptive fuzzy asymmetric GARCH model applied to stock market, in: *Convergence and Hybrid Information Technology*, 2008. ICCIT '08. Third International Conference on, volume 1, 2008, pp. 991–996.
- [13] J.-C. Hung, Forecasting volatility of stock market using adaptive fuzzy-GARCH model, in: *Computer Sciences and Convergence Information Technology*, 2009. ICCIT '09. Fourth International Conference on, 2009, pp. 583–587.
- [14] J.-C. Hung, A fuzzy GARCH model applied to stock market scenario using a genetic algorithm, *Expert Systems with Applications* 36 (2009) 11710–11717.
- [15] J.-C. Hung, A fuzzy asymmetric GARCH model applied to stock markets, *Information Sciences* 179 (2009) 3930–3943.
- [16] J.-C. Hung, Adaptive fuzzy-GARCH model applied to forecasting the volatility of stock markets using particle swarm optimization, *Information Sciences* 181 (2011) 4673–4683.
- [17] J.-C. Hung, Applying a combined fuzzy systems and GARCH model to adaptively forecast stock market volatility, *Applied Soft Computing* 11 (2011) 3938–3945.
- [18] L. Geng, J. Ma, Tsk fuzzy inference system based GARCH model for forecasting exchange rate volatility, in: *Fuzzy Systems and Knowledge Discovery*, 2008. FSKD '08. Fifth International Conference on, volume 3, 2008, pp. 103–107.
- [19] B. R. Chang, Applying nonlinear generalized autoregressive conditional heteroscedasticity to compensate anfis outputs tuned by adaptive support vector regression, *Fuzzy Sets and Systems* 157 (2006) 1832–1850.
- [20] B. R. Chang, Resolving the forecasting problems of overshoot and volatility clustering using anfis coupling nonlinear heteroscedasticity with quantum tuning, *Fuzzy Sets and Systems* 159 (2008) 3183–3200.
- [21] R. Das, K. K. Ang, C. Quek, A synergy of econometrics and computational methods (GARCH-RNFS) for volatility forecasting, in: *Evolutionary Computation (CEC)*, 2010 IEEE Congress on, 2010, pp. 1–8.

- [22] T. G. Andersen, T. Bollerslev, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4 (1997) 115–158.
- [23] T. G. Andersen, T. Bollerslev, F. X. Diebold, P. Labys, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96 (2001) 42–55.
- [24] O. E. Barndorff-Nielsen, N. Shephard, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (2002) 253–280.
- [25] T. G. Andersen, T. Bollerslev, N. Meddahi, Realized volatility forecasting and market microstructure noise, *Journal of Econometrics* 160 (2011) 220–234.
- [26] B. M. Awartani, V. Corradi, Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries, *International Journal of Forecasting* 21 (2005) 167–183.
- [27] P. R. Hansen, A. Lunde, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, *Journal of Applied Econometrics* 20 (2005) 873–889.
- [28] A. Thavaneswaran, K. Thiagarajah, S. Appadoo, Fuzzy coefficient volatility (FCV) models with applications, *Mathematical and Computer Modelling* 45 (2007) 777–786.
- [29] T. Helin, H. Koivisto, The GARCH-fuzzy density method for density forecasting, *Applied Soft Computing* 11 (2011) 4212–4225.
- [30] J. V. Berg, U. Kaymak, R. J. Almeida, Conditional density estimation using probabilistic fuzzy systems, 2013. Accepted for publication in *IEEE Transactions on Fuzzy Systems*.
- [31] R. Almeida, U. Kaymak, Probabilistic fuzzy systems in Value-at-Risk estimation, *Intelligent Systems in Accounting, Finance & Management* 16 (2009) 49–70.

- [32] R. Almeida, N. Basturk, U. Kaymak, V. Milea, A multi-covariate semi-parametric conditional volatility model using probabilistic fuzzy systems, in: Computational Intelligence for Financial Engineering Economics (CIFEr), 2012 IEEE Conference on, 2012, pp. 1–8.
- [33] R. J. Almeida, N. Basturk, U. Kaymak, J. M. C. Sousa, Conditional density estimation using fuzzy GARCH models, in: R. Kruse, M. R. Berthold, C. Moewes, M. A. Gil, P. I. a. Grzegorzewski, O. Hryniewicz (Eds.), Synergies of Soft Computing and Statistics for Intelligent Data Analysis, volume 190 of *Advances in Intelligent Systems and Computing*, Springer Berlin Heidelberg, 2013, pp. 173–181.
- [34] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50 (1982) 987–1007.
- [35] C. G. Lamoureux, W. D. Lastrapes, Heteroskedasticity in stock return data: Volume versus GARCH effects, *Journal of Finance* 45 (1990) 221–229.
- [36] E. Zivot, Practical issues in the analysis of univariate GARCH models, in: T. Andersen, R. Davis, J.-P. Krei, T. Mikosch (Eds.), *Handbook of Financial Time Series*, Springer Verlag, New York, 2009, pp. 113–155.
- [37] C.-S. James Chu, G. J. Santoni, T. Liu, Stock market volatility and regime shifts in returns, *Information Sciences* 94 (1996) 179–190.
- [38] S. Gray, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42 (1996) 27–62.
- [39] F. Klaassen, Improving GARCH volatility forecasts with regime-switching GARCH, *Empirical Economics* 27 (2002) 363–394.
- [40] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modelling and control, *IEEE Transactions on Systems, Man and Cybernetics* 15 (1985) 116–132.
- [41] T. G. Andersen, T. Bollerslev, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39 (1998) 885–905.

- [42] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information Sciences* 8 (1975) 199–249.
- [43] N. Hachicha, B. Jarboui, P. Siarry, A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics, *Information Sciences* 181 (2011) 79–91.
- [44] L. A. Zadeh, Fuzzy sets and information granularity, *Advances in Fuzzy Set Theory and Applications* (1979) 69–129.
- [45] L. A. Zadeh, Is there a need for fuzzy logic?, *Information Sciences* 178 (2008) 2751–2779.
- [46] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems* 90 (1997) 111–127.
- [47] J. Geweke, M. Keane, Smoothly mixing regressions, *Journal of Econometrics* 138 (2007) 252–290.
- [48] X. Huang, A new perspective for optimal portfolio selection with random fuzzy returns, *Information Sciences* 177 (2007) 5404–5414.
- [49] B. Mandelbrot, The variation of certain speculative prices, *The Journal of Business* 36 (1963) 394–419.
- [50] R. F. Engle, V. K. Ng, Measuring and testing the impact of news on volatility, *Journal of Finance* 48 (1993) 1748–1778.
- [51] F. Li, M. Villani, R. Kohn, Flexible modeling of conditional distributions using smooth mixtures of asymmetric student  $t$  densities, *Journal of Statistical Planning and Inference* 140 (2010) 3638–3654.
- [52] M. Villani, R. Kohn, P. Giordani, Regression density estimation using smooth adaptive Gaussian mixtures, *Journal of Econometrics* 153 (2009) 155–173.



<b>ERIM Report Series <i>Research in Management</i></b>	
ERIM Report Series reference number	ERS-2013-013-LIS
Date of publication	2013-07-31
Version	31-07-2013
Number of pages	33
Persistent URL for paper	<a href="http://hdl.handle.net/1765/40785">http://hdl.handle.net/1765/40785</a>
Email address corresponding author	rjalmeida@ese.eur.nl
Address	<p>Erasmus Research Institute of Management (ERIM)  RSM Erasmus University / Erasmus School of Economics  Erasmus University Rotterdam  PO Box 1738  3000 DR Rotterdam, The Netherlands  Phone: +31104081182  Fax: +31104089640  Email: <a href="mailto:info@erim.eur.nl">info@erim.eur.nl</a>  Internet: <a href="http://www.erim.eur.nl">http://www.erim.eur.nl</a></p>
Availability	<p>The ERIM Report Series is distributed through the following platforms:  RePub, the EUR institutional repository  Social Science Research Network (SSRN)  Research Papers in Economics (RePEc)</p>
Classifications	<p>The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems:  Library of Congress Classification (LCC)  Journal of Economic Literature (JEL)  ACM Computing Classification System  Inspec Classification Scheme (ICS)</p>