Erasmus School of Economics,
Erasmus University of Rotterdam

Master Thesis Computational Economics

Financial Risk Analysis Using Probabilistic Fuzzy Systems

Author:
Patrick Nolten
272711

Supervisor:
Dr. Nalan Basturk

Co-reader:
Dr. Tommi Tervonen

August, 2014

Acknowledgments

I would like to take the opportunity to thank my supervisor dr. Nalan Basturk for guiding me with my thesis. She helped me a lot with suggestions, comments and improvements for my research and stood patient with a lot of faith in me as I tried to finish my work. A thank you also goes to co-reader dr. Tervonen, as he was very flexible in the task he was given. A special thanks to ass. prof. Rui Jorge Almeida. Without him I would never have done this thesis since most research here all originates directly from his previous work at the Erasmus University, I wish him all the best now in Eindhoven. I also would like to give my parents Errol and Sisca and my brother Brian credit for all their love and patience, as they kept faith in me and knowing when and when not to push me so I keep on working during hard times. Last, but not least, my girlfriend Stefanie for being there when I needed you, for entering my life during college and for all the love and strength you make me feel. I thank you and I love you, Steef. And for all my friends and family remaining unnamed, I appreciated all their interest, their involvement and their kindness. Thank you all.

Contents

1 Introduction

There is a great need for risk analysis in the world of finance. Risk and return are important factors in portfolio management, so the need for measuring these two elements are equally present [17]. Credit risk, operational risk and market risk are but a few of the different types of financial risk. When looking at market risk, the measure Value-at-Risk (VaR) shows how much value of an asset during a specified percentage of probability will occur. In order to get decent numbers of VaR, models are developed with each their own statistical traits, estimates and accuracies.

Because financial assets have such different volatilities it is important to implement risk management as one of the activities for financial institutions. Thus, choices can be made with information gathered from risk management so that failures happen between acceptable boundaries. The choice on how to hedge or insure an investment portfolio, adjusting stock selections, protecting from environmental risk, the credit risk or the operational risk [19] are all based on information generated from these risk management. Also, the risk of being in the market is one of the risks to analyze with risk management.

Market risk is the risk one gets when market value of a portfolio is not known as a precise number. There is a need to allocate the market risk to one number and therefore a measure like Value-At-Risk (VaR) has been developed. It recreates the risk into one precise number in order to express and summarise the total market risk of any financial portfolio of different assets. VaR measures what the loss of money would be over a given set of time when the market itself remains the same as in it's past at a given confidence level usually defined in percentages. Especially with the financial crisis of the last few years, financial institutions are obligated to use the VaR in order to communicate their level of market risk to the outside world [21]. This suggests that VaR models with proficient accuracy must be developed for efficient decision making of allocation regarding investment capital in order to cover possible losses.

When having risk defined with volatility, researching risk tends to find patterns in volatility. These stylized facts of volatility, whether realized or not, are interesting to look at and for in detail. One such stylized fact is the so-called leverage effect. This is an asymmetric relationship between new information and volatility [10]. A leverage effect is present when a negative return supposedly increases the estimated volatility more than a positive return with the same size. However, the realized volatility already has some information in itself about surprising negative returns, which means the leverage effect is less present than when using daily variances [25]. This is but one example of stylized facts that can be researched when having estimated conditional densities and volatility.

1.1 Understanding uncertainties in risk management

When considering past returns or other past financial values, the conditional density of a financial time series estimates the distribution of the current value regarding that time series. Such estimation has not only its worth in finance and risk management, but also other academic fields. We can state two explanations from the value of conditional density estimation in finance. As financial return series look to have a low correlation with itself over a long time period, correlation does exist in its conditional variance and other descriptive statistics. Classical models that estimate only one value are not able to incorporate the conditional variance in their estimations, at which point creates the desire to estimate the full conditional density. Also, when having the value of expected return, the value of expected risk accompanied with is not provided by these classical models. Expected risk can be transformed into a value using small portions of the estimated distribution of financial returns, such as Value-at-Risk or Expected Shortfall, and classical models providing point forecasts are not sufficient alone for providing estimated and expected risk.

Not only is estimating risk and return in numbers interesting in the field of risk management and portfolio management, but having linguistic vagueness in return and in particularly risk is an occurring event. It is both the uncertainty in the values as well as the linguistic vagueness, or fuzziness, that makes management in finance challenging and interesting.

Examples in economics

In the financial world, stock prices of a financial index move either up or down every period of time [19]. Still, it is hard to predict every period of time the direction of the stock prices correctly, let alone the size of the change in stock prices. Predicting these movements of prices are made difficult, because of probabilistic uncertainty of prices moving in both directions and the size of the movements. These same movements can also be translated into fuzzy statements as a stock can fall big, rise small or remain more or less stable. There is also the part of interpretation on whether the movement of a stock price during a period of time is beneficial or not. If a stock price goes down, it could still be good news when the change of the stock price is less than expected by analysts. Vice versa, if a stock price goes up less than expected by analysts, than this would be bad news. So there is probabilistic uncertainty in the movement of prices and linguistic uncertainty or fuzziness can be used to describe the type of movement and if it can be considered beneficial or not. From here on, we can create the conditional probabilities that stocks move big or small depending on similar mentioned movements of stocks in the past or other linguistic terms that can be modeled. In this case we specifically are not interested in the value of the stocks movement, only its relative linguistic scale of the movement.

These examples above are just a couple of many potential problems of probabilistic uncertainty and fuzziness within economic context, let alone other fields of research. Thus, there is a strong incentive, at least in the research on economics, to develop methods in order to understand more about both probabilistic uncertainty and fuzziness. In this thesis, we are interested in studying the use and relation of probability and fuzzy uncertainty in risk management.

The problem: modeling daily variances

Common approaches to estimate conditional densities and from thereon VaR are simulation and parametric approaches. When looking at VaR with a time horizon different than one day, usually the daily volatility is square-rooted multiplied with to get multiple day horizon volatility [17]. Basic estimation of conditional densities assume a static volatility throughout time. More complex models lose this assumption and recognizing dynamic volatility in time series. Volatility in this dynamic situation is modeled in more than one way.

Getting an accurate model for all possible financial returns in the form of a distribution is a complex activity, because of various statistical properties. These include asymmetry in their distributions, variances that changes on the fly and one or more fat tails in their distributions. Because of these aspects that create such diverse shapes of distributions, several articles describe models to estimate the density of returns, given past return-information and/or other correlated financial variables. The Generalized Autoregressive Heteroscedasticity model (GARCH) is one of the models that is widely known for estimating return distributions, especially because it primarily estimates the volatility first [11]. This model explains the variance in returns by using historical returns and historical variances.

When GARCH models are preferred, but statistical properties other than the variances are included, various literature talk about extensions of the GARCH models, e.g. Student-t GARCH models is an extended model that includes the fat tails whereas GJR-GARCH tries to model skewness [15, 12]. Different types of GARCH models evolved into other models with similar GARCH properties that can capture other properties of return distributions, such as smooth transition GARCH models and regime-switching GARCH models [16, 18]. GARCH models have also been combined with artificial neural networks [9] and fuzzy systems [34] in different ways. A fuzzy GARCH model is a set of fuzzy rules, each rule firing off a separate autonomic GARCH model [20]. In this model 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.

Combinations of other type of fuzzy systems and GARCH models can also be used to estimate financial returns. Usually they either approximate return series or realized volatility [14]. Realized volatility is the sum of squared intra-daily (e.g. 5 minutes data) returns [5]. This realized volatility also an example of being a point based prediction, whereas return volatility a density based prediction is. Finally, a GARCH model with fuzzy coefficients can model the error term, also by using a set of fuzzy rules [28]. All previous mentioned models combine fuzzy systems with a statistical model. Fuzzy systems fusing with statistical elements are also the probabilistic fuzzy system [4, 27, 26], used to estimate conditional volatility of returns and probability distributions. This type of system that combines fuzzy logic and probabilities can thus also be used for conditional density estimation.

The probabilistic fuzzy system (PFS) is a semi-parametric model where the model structure is flexible and the model parameters can adapt itself to the underlying data distribution, such as a financial time series. It is in this model where data statistics and their elements are combined with a linguistic description of the system behaviour. This combination is encoded with fuzzy rules. PFS is suitable for approximating probability distributions, whether that being time series, temperature distributions or error term distributions. This kind of approximating is also aptly performed on VaR by estimating the probability distribution of the data. PFS has been used to estimate VaR in [4] and [33], where they concentrated on just using models with a single covariate in order to determine its parameters.

1.2 Previous work on probabilistic fuzzy systems

The idea of PFS is not a complete new one according to [22]. Although it is in [22] that probabilistic weighting is introduced in fuzzy systems, it is also noted that several classifiers discussed in previous literature behave and act like PFS, but are not called as such. For the most part, [22] shows how to create classifiers from PFS instead of regular fuzzy systems and both theory and analysis in order to create PFS are illustrated. One classifier that was taken as example of several other studies, although not called as a PFS, was applied on target selection of direct marketing. Classification is not added in this thesis, but it is worth noting that PFS already has shown accurate estimations, even if it is in classification. PFS itself is already documented and analyzed in its theory fully in [22], [23], [6] and [32]. When utilizing PFS for regression, the parameters to optimize are the antecedent membership functions and the conditional probabilities. The proposed method of optimization on both previous mentioned membership functions and probabilities is the maximum likelihood optimization. Also, when using few antecedent membership functions, PFS shows a heavy loss on accuracy compared to using more antecedent membership functions according to [32]. This is the case when regression problems are looked upon with PFS. Using more antecedent membership functions, thus defining more rules may result in benefitting accuracy with rules remaining humanly readable.

There are a couple of articles about PFS like [22] and [32], but few about PFS used on randomized data sets like a normal pdf. The ones that discuss PFS, are usually about how PFS is applied on real life data sets like in [33]. In [33], PFS is used as a means of risk management. In particular, their empirical study tries to get a good estimate of the Value-at-Risk (VaR), a risk measure of financial assets. The reason of using PFS is that the VaR has a probability distribution function which does not fit within parametric modeling.

PFS is a semi-parametric model, since it uses the statistical elements of the data with descriptive rules, that create a systematic behavior. By using back-testing, [33] shows acceptable validation of using PFS on VaR. Also, the estimates utilizing PFS are not rejected by backtesting compared to the estimates utilizing a Generalized Auto Regressive Heteroscedasticity model (GARCH), another method to measure VaR with. GARCH is an econometric model that estimates time series with the focus on error variance and extracting characteristics from those series, whereas PFS is a model that can be built with similar intentions, but originates from research in soft computing. In [33] it is argued that PFS fits better than their benchmark concerning the properties of their data being stock returns of several stock indices.

A continuation of this research is done in [4]. In here, the added element is a change in the PFS in order to get a better view on the edges of the probability distribution. This is accomplished by initializing the consequent membership functions in the output space of the PFS more towards the edges of the data set. This is different from [33] and other classical approaches of initializing consequent membership functions in the output space, because they use equal spacing between those membership functions. From thereon, [4] also compares the backtesting results of estimating VaR to the same results from GARCH models. The conclusion remains consistent with [33]: whereas GARCH models are sometimes rejected, models of PFS never are. Also, the added feature of differing the membership functions in the output space besides equal spacing is an extra quality of PFS concerning adjusting to different data sets. This feature will also be used in this thesis later on and more explained in chapter 5.

One of the latest papers about PFS is [6]. Probabilistic fuzzy systems have the ability to refine linguistic concepts of the user and with these refinements approximate conditional probability density functions. In such a way, both the probabilities that are uncertain as well as fuzzy linguistic descriptions are implemented in the model. Among others, [6] suggested an additive reasoning scheme for PFS. In order to get a crisp output, the uncertainty of fuzzy linguistic description is calculated by defuzzification while at the same time the uncertainty of probabilities is computed by an averaging step which leads to an expected output of the PFS. There are some pre-calculations required for this kind of estimation. First, the probability of a consequent fuzzy event given an antecedent fuzzy system is calculated. Second, the centroid points of the consequent fuzzy sets are computed. And third, the degree of fulfillment of the fuzzy rules is determined. The output of a PFS is the same as a zero-order Takagi-Sugeno fuzzy system with conditional expectation of the defuzzified output membership functions selected as the consequent variables. This shows that such fuzzy systems, being deterministic or probabilistic are successful in the approximation of various functions.

The most work on PFS and VaR is done in [33], [4], [2] and [1]. All these papers show that VaR of different stocks and indices can be estimated using PFS. They show that their VaR is rarely rejected compared to other models, such as Garch(1,1). However, neither of them have looked upon stylized facts of volatility and every paper is different from their selected stocks and indices. This thesis will therefor be an addition to previous work of PFS applied in finance. Also, by estimating conditional densities, not only can VaR be calculated, but interestingly other measures and quantiles can be derived from these densities.

1.3 Research questions

As far as modeling variances go, probabilistic fuzzy systems is a viable way to estimate probability distributions. There is still ground to cover when looking at how well a PFS performs, in particular at stock returns. So our main research question is:

> "Is there any ability in a probabilistic fuzzy system to be successfully applied on performing financial risk analysis?"

From hereon, the following sub questions can be derived:

1. Does a probabilistic fuzzy system has the ability to be successfully applied on stock data to perform financial risk analysis?
2. How do we estimate conditional probability distributions and volatilities from probabilistic fuzzy system for financial stocks?
3. Is it possible to perform risk analysis with stylized facts utilising the probabilistic fuzzy system?
4. Can a probabilistic fuzzy system create models used to forecast different time periods?
5. How well does a probabilistic fuzzy system estimate the statistical properties of a stock?

Now that the sub questions are present, a canvas of the methodology can be presented. Within the subsection of methodology, the structure of the thesis will be shown and the boundaries of the scope on the thesis are set.

1.4 Methodology

First, we look what models have been used in order to estimate VaR, variances and conditional probability distributions. There are several statistical models, where GARCH is most common. PFS is the main focus in this thesis, so some work is also presented about this model. This already answers whether PFS has the ability to successfully be applied on financial data, such as SP500 in [6]. So our first sub question can be answered positively, because in previous literature the model has shown applications regarding financial estimations. In this thesis, instead of stock index data, the focus will be on stock data. Next, in detail will be seen how PFS works and what can be improved to get better results in the estimation. For all stocks that we estimate the distributions from, we use the same model created from PFS. Furthermore, the selected stock data will be gained from financial databases and some descriptions will be shown on how the stocks actually moved for a long period of time. From this data covariates are created similar to [31] and [24]. Using the returns and covariates, we also get some descriptive statistics. It is here that the workings and formulations of the PFS becomes clear and how it estimates the conditional probability distributions, volatilities and quantiles for estimating VaR.

Experiments are done by varying the mentioned covariates, sorting on different weekdays and leverage effects. As shown in [25], there is some noteworthy information hidden in the returns sorted by the day of the week, and with PFS the relation between on one hand day of the week and the other hand returns and the covariates is being examined. This answers the question on whether PFS can support stylized facts in risk analysis. Also, we look at multiple day horizons in order to look at market risk capital management. This kind of management looks at how much money should be set aside in order to cover market risk over a longer horizon of time. Remember that VaR usually looks at immediate time ahead, but with market risk capital management, VaR becomes also interesting when looking further the future in rather than the moment at hand. Using these multiple day horizons, the PFS will show how it deals with forecasting different time periods. Finally we compare the models of PFS with a Garch(1,1) model with the widely known Kupiec test and the Christoffersen's Markov test from [13]. This should reveal the effectiveness of PFS and support the evidence that the model is viable within finance.

Structure

Chapter One is an introduction about why PFS is interesting in order to use in financial risk management, what previous and similar work has been done and what the focus is in this thesis. Chapter Two is about PFS in specific detail, what kind of optimization is used and what the value is of using PFS in estimating stock returns and the variances. Chapter Three discusses the experiments that are threefold, where we first look at estimating the VaR one day ahead and compare between stocks. Second, the comparison between covariates used as input are the base for the second experiment. Third, we look at estimating the VaR ten days ahead where also results stocks and covariates are compared, which can be mentioned beforehand as similar with the first experiment. All three experiments are explained beforehand in chapter Three. Finally, this will all be concluded and highlighted in Chapter Four.

2 Probabilistic Fuzzy Systems

The model that is used in the thesis to perform financial risk management is called the probabilistic fuzzy systems. This model is an extended version of the regular fuzzy system, where probability signifies the uncertainty within the fuzzy rules. The first notion of having a probabilistic element within a fuzzy event was in [34]. Before the model is explained in detail, there will be an example on how a set of fuzzy rules are linguistically built. This example is broken in two versions where the first version is a similar example like in [26]. The second version is a larger version of the first one, more specifically the larger input space. Both versions of examples illustrates the conversion from linguistic statements based on expert knowledge to PFS. After this example, the details of PFS are shown in full. A look is given on how a fuzzy histogram model is the precursor to the probabilistic output model and how the additive reasoning scheme result in similar input and output of the models. Next, the parameters of a PFS are explained and lastly the optimization methods for those parameters are depicted. These parameters are one by one being set or optimized within the experiments in chapter 3. A lot of the theory can be found also in [4], [3], [1], [23], [29], and [30]. Finally, the practicality of the PFS modeling regarding the volatility and VaR estimation is explained. A brief note on how the comparison of the PFS and GARCH(1,1) is also given, as both types of models share similarities regarding the usefulness of volatility estimation. All this information about transforming linguistic advice into probabilistic fuzzy rules for the system, the fuzzy clustering method and the likelihood optimization answers the sub question on how we estimate the distributions and volatilities with the mentioned models.

2.1 Transforming a linguistic advice into probabilistic fuzzy rules

A probabilistic fuzzy system is a set of rules where the input space includes one or more fuzzy antecedents and the output space includes multiple consequents, each one referring to a probability distribution function. This set of rules can be derived from expert knowledge, where the derivation is similar to the fuzzy systems with fuzzy events as its consequents. The expert knowledge now can include some conditional probabilistic uncertainty, which is more common than expert knowledge with absolute certainties included or expert knowledge based on one-on-one relationships.

Consider the following expert knowledge of a financial analyst about the history of realized variance that affects the current return of a stock. This expert says that when the realized variance of a stock increased with a large jump the previous day, the stock return will likely rise with similar size today and thus the price of the stock could increase. Not only does he say what the stock return likely will do, he will also tell what the stock return would unlikely do depending on the same realized variance. For each of the four different situations of what the realized variance he sees has done, he tells the likely and unlikely occurrences of movement of the stock return. In the quote below, all statements are written down.

All these statements can now be implemented into a fuzzy system as the expert knowledge is suitable for PFS. There is an input variable 'variance yesterday' that can have several fuzzy properties like 'increased large' or 'decreased small'. There is also a set of actions when the conditions in the fuzzy system are met, which will be our output variable. According to our financial analyst, this set consists of four actions: 'stock rises large', 'stock rises small', 'stock falls large' and 'stock falls small', all four the next period of time with some factor of likeliness. This likeliness secures the linguistic uncertainty necessary for PFS. This set of statements translates i.e. to,

IF realized variance yesterday IS increased large THEN stock return today is either: likely to rise large; unlikely to rise small; unlikely to fall small; and likely to fall large;

IF realized variance yesterday IS increased small THEN stock return today is either: unlikely to rise large; likely to rise small; likely to fall small; and unlikely to fall large;

IF realized variance yesterday IS decreased small THEN stock return today is either: unlikely to rise large; likely to rise small; unlikely to fall small; and unlikely to fall large;

IF realized variance yesterday IS decreased large THEN stock return today is either: unlikely to rise large; unlikely to rise small; unlikely to fall small; and unlikely to fall large;

This set of statements at this point is easily implemented in a probabilistic fuzzy system, since there is only one input variable and one output variable. An example of using expert knowledge into a probabilistic fuzzy system in order to help decision makers with stock selection can be found in [33] and [26].

Now that same financial analyst suddenly has more information available regarding his predictions about the stock return. This extra information to his disposal is the current market sentiment which he views as positive, neutral and negative. With the market sentiment added, he tells how stock returns are likely or unlikely to change based on the realized variance yesterday and market sentiment today. He changes his four statements into twelve statements which show all combinations of the four perceived changes of variance and the three levels of market sentiment. All these statements can be implemented in the PFS, regardless on how many input variables there are, which in our example are two input variables. His full set of predictions now form the following pattern:

IF variance yesterday IS increased large AND market sentiment IS positive THEN stock return today is either: likely to rise large; likely to rise small; likely to fall small; and unlikely to fall large;

IF variance yesterday IS increased large AND market sentiment IS neutral THEN stock return today is either: likely to rise large; likely to rise small; likely to fall small; and likely to fall large;

IF variance yesterday IS increased large AND market sentiment IS negative THEN stock return today is either: unlikely to rise large; likely to rise small; likely to fall small; and likely to fall large;

IF variance yesterday IS increased small AND market sentiment IS positive THEN stock return today is either: unlikely to rise large; likely to rise small; unlikely to fall small; and unlikely to fall large;

IF variance yesterday IS increased small AND market sentiment IS neutral THEN stock return today is either: unlikely to rise large; likely to rise small; likely to fall small; and unlikely to small large;

IF variance yesterday IS increased small AND market sentiment IS negative THEN stock return today is either: unlikely to rise large; unlikely to rise small; likely to fall small; and unlikely to fall large;

IF variance yesterday IS decreased small AND market sentiment IS positive THEN stock return today is either: likely to rise large; likely to rise small; unlikely to fall small; and unlikely to fall large;

IF variance yesterday IS decreased small AND market sentiment IS neutral THEN stock return today is either: unlikely to rise large; unlikely to rise small; unlikely to fall small; and unlikely to small large;

IF variance yesterday IS decreased small AND market sentiment IS negative THEN stock return today is either: unlikely to rise large; unlikely to rise small; likely to fall small; and likely to fall large;

IF variance yesterday IS decreased large AND market sentiment IS positive THEN stock return today is either: likely to rise large; likely to rise small; unlikely to fall small; and unlikely to fall large;

IF variance yesterday IS decreased large AND market sentiment IS neutral THEN stock return today is either: unlikely to rise large; unlikely to rise small; unlikely to fall small; and unlikely to small large;

IF variance yesterday IS decreased large AND market sentiment IS negative THEN stock return today is either: unlikely to rise large; unlikely to rise small; likely to fall small; and likely to fall large;

A larger input space becomes apparent as the latest predictions now are built from two input variables them. The actions in the example are bound to happen ‘likely’ and ’unlikely’, but if the expert would have given multiple forms of linguistic vagueness, this wouldn't be a problem as long as there is some order in it. As we will see further on in the thesis, it becomes clear that the starting probabilities as defined by the linguistic uncertainty will change to fit the predictions. This shows that the linguistic uncertainty 'unlikely' changes differently in value within each statement. This gives the modeler more flexibility in their initial linguistic uncertainties as they are not required to be specific, because the optimal linguistic uncertainties will change in order to get better predictions with the use of expert knowledge.

Unlike the linguistic uncertainty, the fuzzy uncertainties are required to build a PFS. Thus, it is important to have some form of fuzziness in your statements, more specifically in both the conditions of the statements as well as the actions of the statements. Still, no numeric values are required for building the PFS, which shows the independence of the model regarding such numeric values as an advantage. From this example it is possible to derive a more formal equation that entails the fuzzy rules regarding multiple input variables and

$$\begin{aligned}\text{Rule } R_q :\ &\text{If } \ \mathbf{x}_1 \text{ is } A_{1,1} \text{ and } \ldots \mathbf{x}_s \text{ is } A_{s,u} \text{ then} \\ &y \text{ is } C_1 \text{ with } \Pr(C_1|A_{1,1}\ldots A_{1,u}\ldots A_{s,u}) \text{ and} \\ &y \text{ is } C_2 \text{ with } \Pr(C_2|A_{1,1}\ldots A_{1,u}\ldots A_{s,u}) \text{ and} \ldots \text{and} \\ &y \text{ is } C_N \text{ with } \Pr(C_N|A_{1,1}\ldots A_{1,u}\ldots A_{s,u}).\end{aligned} \tag{1}$$

For each rule R_q with $q = 1, 2, ...Q$ where Q is the number of rules, if the fuzzy event A_q as antecedent happens, then each fuzzy event C_j as consequents in this rule can happen, where $j = 1, 2, ...N$, N being the number of consequents. Each consequent within a rule can be selected with the conditional probability $Pr(C_j|A_q)$. In our latest example of the financial analyst x is 'stock return on current day', A_q is {'rising large', 'rising small', 'falling large', 'falling small'} with $q \in \{1...4\}$, y is 'stock return on next day' and C_N is {'rising large', 'rising small', 'falling large', 'falling small'} with $j \in \{1...4\}$.

2.2 Theory of probabilistic fuzzy systems

From equation 1, we can declare that both the conditional probabilities as well as the fuzzy uncertainties are captured by a set of rules and statements such as earlier described above. Expert knowledge thus can be language-only in order to get such a system, and benefits from becoming readable. This becomes an apparent advantage over other models that are knowledge-based, since not only are the statistical properties of a regular fuzzy system present, but also the property of readability. This property becomes essential for looking at a stylized fact such as leverage effects later on briefly mentioned in chapter 3.

PFS are an extension of Mamdani, Takagi-Sugeno fuzzy systems, and other fuzzy systems. This is one of the similarities PFS has with other fuzzy systems. The difference between the Mamdani and the Takagi-Sugeno fuzzy system is the consequent membership functions, where Mamdani's output space is fuzzified ($y = C_j$) and Takagi-Sugeno output space are functions of the input ($y = f(x)$). For the purposes in this thesis, Takagi-Sugeno is left out, thus we utilize a Mamdani system to estimate pdfs. Similar to a regular fuzzy system, a PFS has a set of rules with a mechanism that will reason by inferencing the rules. Seeing a PFS using conditional probabilities during modeling, it can be deduced that PFS is stochastic in its output. Therefore, to estimate a pdf like the (multivariate) normal distribution, a system that can handle stochastic output like PFS is a viable alternative.

When looking at equation 1 again, the variables change when now using the same equation for estimating a normal probability distribution. So when considering equation 1, $\mathbf{x}$ is a one-dimensional vector and A_q is a one-dimensional antecedent for rule q that is determined by the membership function of an antecedent $\mu_{A_q}(x)$; y is the variable that depends on $\mathbf{x}$ and comes with $y \in Y$. This is also the variable we want to predict. C_N are the consequents of the fuzzy system. These consequents are determined by $\mu_{C_q}(y)$, the fuzzy degree of consequent layer.

The parameter that regular fuzzy systems do not have and PFS do have is $Pr(C_j|A_q)$. Assuming rule j is fired or activated, $Pr(C_j|A_q)$ is the conditional probability that y will be C_j, which belongs to a fuzzy set. Also, the mapping between the antecedents and consequents are realized by the two 'levels' between the *probability* distribution of the training data set and the *possibility* distribution of the membership functions in the input and output space, as detailed in [34]. This makes sure a construction of interpolation is present in order to get the conditional probabilities of the estimated distribution. For convenience, we shorten the membership functions (mfs) in the input space as the input membership functions (imfs) and likewise the membership functions in the output space as output membership functions (omfs). These imfs and omfs are thus also the mfs used in the antecedents and consequents. Also, for clarity'sake, mf is just stating one membership function, whereas mfs is stating membership functions.

Let β_q be the normalized degree of fulfillment which will be stated as

$$\beta_q = \frac{\mu_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} \mu_{A_{q'}}(\mathbf{x})} \tag{2}$$

where μ_{A_q} is the degree of fulfillment of rule R_q. If $\mathbf{x}$ was multi dimensional, a conjunction would be needed in order to establish μ_{A_q}. However, since in this thesis the input vector is one dimensional, there is no need to conjunct separate memberships into one membership using any form of t-norm. Thus, the conditional probabilities can be now be computed by using an additive reasoning scheme with multiplicative aggregation of the rule antecedents as seen in [23]. This is stated as

$$f(y|\mathbf{x}) = \sum_{j=1}^{N} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \Pr(C_j|A_q) \frac{\mu_{C_j}(y)}{\int \mu_{C_j}(y)dy} \tag{3}$$

when the input vector is given. Equation 3 is the main formula that needs to be optimized. This equation does have a constraint with it: the output space should be well-formed. This constraint is needed in order to get total conditional probabilities summ ed up to one. Proof is given in [30]. So the constraint is

$$\sum_{j}^{N} \mu_{C_j}(y) = 1, \forall y \in Y \tag{4}$$

One possible interpretation of PFS is taking into account the assumption that output space Y is fully divided into N fuzzy classes, each with their own membership function μ_{C_j}. We get a pdf built by fuzzy columns, where each fuzzy column $f_j(y)$ represents each fuzzy class. When the interval between columns becomes very small, like in a crisp histogram to approximate better, the fuzzy histogram gets more smoothed. Thus, the fuzzy histogram will be

$$f_j(y) = \frac{\Pr(C_j)\mu_{C_j}(y)}{\int \mu_{C_j}(y)dy} \tag{5}$$

When totalling each fuzzy column $f_j(y)$, the true pdf is approximated better with a smoothed fuzzy histogram than with a crisp histogram as stated in [4].

2.3 Parameters of probabilistic fuzzy systems

There are several methods to modify the parameters of a PFS with [27] holding a description about them. In [27], the possibility frequency method, estimation of triangular membership functions and a number of fuzzy c-means clustering techniques are illustrated. Further on in the thesis, because the experiments use different estimations of the output membership functions, we exclude them from this part of the thesis. Also, during optimizing, the output membership functions vary in favor of the experiments, i.e. we fix them and look at the influence of the probability parameters. For now, the focus is only on the one that influences the model the most in this thesis, which are the fuzzy c-means clustering in the input space and the maximum likelihood optimization on the probability parameters.

Fuzzy c-means clustering

Fuzzy c-means clustering is a method introduced in [8] that clusters data points based on their most possible similarity between the clustered data points. The clustering is defined by giving each data point a membership value that determines by how much a data point is selected for a certain cluster. These memberships creates the added feature that a data point can belong to multiple clusters, whereas in hard clustering the clusters are crisp. Note that the clustering itself does not derive an optimal number of clusters, as the number of clusters is set by the number of rules a PFS contains. Each rule consists of one antecedent, and each antecedent is built by one cluster in this thesis.

The approach in this thesis when using fuzzy c-means clustering is that the points are all graded first towards gaussian mfs, and from thereon create new types of mfs. In chapter 3 we explain the transformation from gaussian mfs to other types used in this thesis, as we also change the type of membership functions within our input space later in chapter 3. This varying in type of antecedents gives more options in the experiments beforehand.

When having data points $x_k, k = 1, 2, ...K$ with K being the number of data points, we minimize the objective formula

$$\begin{array}{rcl} J(X;U,V) & = & \sum_{i=1}^{C}\sum_{k=1}^{K}(\mu_{qk})^m\{||x_k - v_q||^2\} \\ \text{s.t.} & & \sum_{i=1}^{C}\mu_{qk} = 1 \end{array} \tag{6}$$

where C represents the number of clusters. This generalized least squared-objective function tries to find the minimum of the sum of the distances between the data points and their centers multiplied with the allocated memberships. It is those memberships that are changed in order to get the minimum. The constraint in equation 6 is that each data point has a total membership of 1 over all clusters, so that the data point is present in full when assigning them to the clusters. So, we minimize the absolute quadratic distances between data points x_k and centers v_q multiplied with the membership μ_{qk} of the corresponding data point to that center, i.e. if the covariance matrix F_q for cluster q is

$$F_q = \frac{\sum_{k=1}^{K}(\mu_{qk})^m(\mathbf{x}_k - \mathbf{v}_q)^T(\mathbf{x}_k - \mathbf{v}_q)}{\sum_{k=1}^{K}(\mu_{qk})^m} \tag{7}$$

then the (immediately gaussian) input mfs are calculated as

$$\mu_{A_q} = e^{-\frac{1}{2}\mathbf{x}^T\mathbf{F}_q^{-1}\mathbf{x}} \tag{8}$$

We create the input space with fuzzy c-means clustering, because we want to use any information in the data set that shows any form of separation in groups within the data set. Also, fuzzy c-means clustering has shown in [33] and [27] to give adequate results. In this thesis, the sets are random generated, so it is interesting to see how the clustering will take care of non-real life data sets. For a full explanation on fuzzy c-means clustering, see [7].

Maximum likelihood optimization

Here is a brief explanation of the maximum likelihood optimization, but there is a similar,but more detailed explanation in [27]. In this thesis, the membership functions are fixed, whereas the probability parameters are being optimized in order to get the best approximation of the true pdf. This looks like

$$\begin{aligned} L(\theta) &= \prod_i \Pr(y_i|x_i) \\ &= \sum_i^N \log\left(\Pr(y_i|x_i)\right) \\ &= \sum_i^N \log\left(\sum_{j=1}^N \sum_{q=1}^Q \beta_q(\mathbf{x}) \Pr(C_j|A_q) \frac{\mu_{C_j}(y)}{\int \mu_{C_j}(y)dy}\right) \end{aligned} \tag{9}$$

with the constraints

$$\begin{array}{ll} \sum_j \Pr(C_j|A_q) = 1, & \forall q, \\ Pr(C_j|A_q) > 0, & \forall j, q. \end{array} \tag{10}$$

As the conditional probabilities are optimized with the maximum likelihood optimization, we concentrate on the membership functions. Looking at the influence of the probability parameters would create a research worthy on its own. Whereas the antecedents and consequents are being varied beforehand, as will be looked upon only in initialization in chapter 3, the optimized probability parameters are taken as given in our research. Before the experiments are illustrated better with the setup, an understanding of our selected measure the KL-divergence is needed, which is shown in the following chapter. Once the understanding and choice of this measure is clear, the experiments and results are depicted.

3 Results

In this section the results are going to be shown here. First, a look is given at what the results of the PFS models are when calculating Value-at-Risk. Second, the results of looking multiple days ahead will be presented as not only does PFS perform one day ahead predictions, but also ten days ahead as introduced previously in section 1. Both the Kupiec test as well as the Christoffersen's Markov test are applied, so that assessments are presented for both one-day and ten-days ahead. Kupiec test shows how well the frequency is of the exceptions, whereas Christoffersen's Markov test would show whether or not the model shows no correlation between exceptions. It is from here that Kupiec test is deemed more appropiate for our experiments, thus the CM test remains barely mentioned. Third, stocks are compared against each other using the results of the two tests. Fourth, both differences and likeliness are sought between covariates combinations. Finally, a brief summary of suggestions on which stylized facts to look on based on the given results.

3.1 Resulting Value-at-Risk

The stocks that have been chosen are all stocks that in one way or another in a world or national index. They range between different international stock indices, so there is no chosen bound on geography. Some of them are pairwise chosen in that they share a similar sector, being financial or food. This choice has been made so that similarities can be shown from the tests. Every stock acts as a data set containing five or more years of daily returns calculated as the logarithmic difference between current and yesterday's return. From every data set the last year, which assumes a trading year of 255 work days, is chosen as a validation set, whereas all other years of data before are chosen as training data. Note that every validation set uses the same dates, thus being the same size each, whereas training data depends on how many years were available. All data is collected through data banks provided from and at the university library of the Erasmus University Rotterdam.

The covariates are similarly created here as in [31] and [3]. *LastDay* is the return from the last day or otherwise called the previous return from the current day; *LastWeek* is the unweighted average of the previous five days from the current day; *LastMonth* differs only from *LastWeek* in that the number of previous days is twenty. $CloseAbs_{9}5$ and $CloseAbs_{8}0$ are averages of previous absolute returns that decreases in a geometrically way where the first one decreases faster than the latter one, meaning the first one uses more information of the previous returns than the latter. In other words,

$$\text{CloseAbs}_\rho = (1-\rho)\sum_{s=0}^{\infty} \rho^s |r_{(t-1-s)}| \tag{11}$$

where ρ is the decreasing factor. Almost similarly constructed, $CloseSqr_{9}5$ and $CloseSqr_{8}0$ are

$$\text{CloseSqr}_\rho = \sqrt{(1-\rho)\sum_{s=0}^{\infty} \rho^s r_{t-1-s}^2} \tag{12}$$

Furthermore, the popular covariates *MaxMin95* and *MaxMin80* are included, because it indicates a little bit of added information to volatility estimating [31]. It incorporates the logarithmic difference between a maximum and minimum pricing point during a trade day. This means that the minima and maxima during trade days are also collected of the aforementioned data set. There is no need to scale these covariates, since it will alter the effects of the fuzzy clustering method negatively. Now that the covariates are set, the parameters of our models should be set properly. Since there are 9 covariates, the total number of combinations of covariates that can be used as input is 2^9. Therefore, only the most important combinations are used: each covariate separately, the covariates grouped by type (e.g. only the two *CloseSqr* are selected) and all covariates, with one covariate or a specific group of covariates excluded.

Because of the chosen initialization for the model, it uses in the input space five gaussian membership functions and in the output space eight triangular membership functions, where two trapezoid membership functions are on the side for capturing outliers. This automatically defines the number of probability parameters to fifty. There is also the possibility for other types of membership functions, but this holds no significance for comparisons between stocks or covariates. Now that everything is set to estimate, first there's a closer look on the stocks compared with each other.

(one figure may be added to illustrate the estimations of pfs; if so, explanation added in separate subsection here -PN)

3.2 Comparisons between stocks

Table 1, 2 and 3 presents the occurrences of faults under the Kupiec test. These numbers are specifically at the three percentages 95%, 97.5% and 99%. Every bold number indicates that the number of faults falls outside the bounds of what is deemed acceptable. A view on the results regarding estimating pdfs on one day ahead is presented first and afterwards a similar view with ten days ahead.

When looking at the stocks individual, it seems that of all the stocks, Tsingtao's is rejected the most. This is largely due to the fact that although they are a beer company for a large amount of time, they are the newcomer on the Chinese index. Also, index stocks with a large history of time series return on have a higher acceptance rate. From here, the derivation that pfs performs better with more data seems just. The inherent downside is that companies already should exist for some time to get proper estimations, whereas start-ups and failed companies are not on the index

Table 1: Single covariates

Estimation pfs current day		Number of exceptions Kupiec test														
Covariates	α	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
	0,95	54	39	**17**	**75**	42	43	**64**	32	**83**	**23**	**22**	**19**	**17**	**8**	**22**
LD	0,975	24	15	**6**	**46**	16	21	**37**	15	**42**	**8**	**10**	**5**	**3**	**5**	**8**
	0,99	7	7	4	**21**	7	7	18	3	13	**1**	2	2	2	3	2
	0,95	**60**	46	**19**	**76**	32	43	**59**	**24**	**84**	**28**	**21**	**17**	**18**	**12**	**24**
LW	0,975	30	16	**7**	**45**	20	14	**34**	**11**	**46**	**9**	12	**5**	**9**	**6**	**9**
	0,99	11	8	3	**19**	9	6	13	3	17	**0**	2	3	**1**	2	3
	0,95	54	44	**13**	**71**	33	43	**65**	**30**	**74**	**27**	**22**	**16**	**19**	**12**	**23**
LM	0,975	26	16	**5**	**35**	20	18	29	15	41	11	11	6	6	5	10
	0,99	12	9	3	16	9	10	15	2	12	**1**	2	2	2	2	3
	0,95	49	51	44	49	47	53	44	48	47	47	50	41	48	**22**	44
CA95	0,975	25	26	16	27	21	27	20	27	22	22	24	15	20	**10**	23
	0,99	12	11	7	12	10	8	5	13	7	10	9	8	6	3	10
	0,95	48	55	33	54	40	49	54	45	59	39	40	**30**	50	**16**	36
CA80	0,975	29	30	15	30	20	23	27	22	32	18	19	**10**	18	5	16
	0,99	12	12	5	11	9	9	10	11	16	5	8	5	4	3	10
	0,95	50	48	46	48	47	48	41	46	49	47	46	40	49	**29**	44
CS95	0,975	26	23	19	27	24	25	21	26	22	23	21	14	20	14	24
	0,99	11	9	8	9	10	7	6	10	5	9	9	8	6	5	11
	0,95	48	52	41	55	39	45	54	41	**58**	41	37	31	47	**23**	38
CS80	0,975	29	28	18	29	21	24	31	23	30	17	17	**11**	14	9	16
	0,99	11	10	6	10	9	9	8	9	14	6	9	6	4	3	9
	0,95	44	46	37	46	38	39	32	47	37	39	38	49	**23**	**18**	31
MM95	0,975	22	20	17	25	18	15	**10**	22	20	23	16	21	**11**	**7**	13
	0,99	10	9	6	9	9	7	2	12	7	9	9	7	3	3	6
	0,95	50	48	39	53	35	38	32	51	46	41	33	41	**25**	**16**	**30**
MM80	0,975	27	24	17	27	18	13	13	25	24	21	15	17	**6**	**7**	**14**
	0,99	11	11	6	9	7	8	4	10	8	6	9	7	3	2	7

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

Table 2: Groups of covariates

Estimation current day		Number of exceptions Kupiec test														
Covariates	α	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
	0,95	49	49	42	52	46	44	51	47	50	41	44	37	53	**22**	42
CA	0,975	25	28	19	28	22	26	24	27	31	21	23	13	19	**9**	22
	0,99	13	12	8	12	10	10	7	12	10	9	12	6	4	3	11
	0,95	50	50	47	52	42	43	51	48	53	42	40	34	56	**27**	44
CS	0,975	25	31	21	27	23	23	25	26	30	23	21	15	21	14	22
	0,99	10	10	8	9	10	9	5	11	9	9	11	7	5	5	11
	0,95	49	48	38	51	43	35	38	52	45	42	35	49	**24**	**19**	31
MM	0,975	29	22	18	21	19	13	13	24	23	22	16	20	**10**	**8**	13
	0,99	9	9	6	6	8	6	6	10	8	6	10	6	3	3	7
	0,95	52	41	**17**	**80**	35	45	**62**	**29**	**87**	**23**	**20**	**15**	**14**	**7**	**20**
LD,LW	0,975	26	15	5	**50**	13	25	30	12	**50**	**6**	**8**	**6**	**6**	**4**	**8**
	0,99	4	5	4	**24**	7	10	13	2	17	**1**	**1**	**4**	**0**	3	3
	0,95	55	39	**18**	**73**	40	44	**66**	33	**80**	**22**	**21**	**15**	**19**	**11**	**24**
LD,LM	0,975	26	17	6	**38**	15	21	38	16	**46**	**6**	**9**	**7**	**3**	**6**	**9**
	0,99	12	7	3	18	9	9	17	2	17	**1**	3	4	**1**	3	3
	0,95	59	36	**16**	**83**	37	45	**66**	**28**	**77**	**23**	**22**	**15**	**19**	**11**	**22**
LD,LW,LM	0,975	25	15	**5**	**53**	18	22	**35**	13	**47**	**4**	**10**	**7**	**6**	**6**	**8**
	0,99	10	6	4	**22**	7	11	16	2	16	**1**	3	4	**1**	2	3
	0,95	**59**	44	44	53	45	49	55	43	**62**	40	**29**	**25**	39	**12**	35
LD,CA95	0,975	28	24	13	**38**	19	25	32	27	**37**	22	18	**10**	14	**5**	15
	0,99	14	9	7	17	7	8	10	13	16	5	8	3	4	3	7
	0,95	**58**	44	39	51	46	46	57	41	**72**	40	**29**	**18**	38	**13**	32
LD,CA80	0,975	28	19	**11**	31	15	26	32	27	**35**	14	15	**7**	13	**5**	14
	0,99	9	7	8	17	6	5	14	13	17	2	4	3	4	3	8
	0,95	49	41	45	55	46	47	47	48	55	43	39	33	52	**15**	39
LD,CA	0,975	28	21	13	32	19	25	27	24	29	22	19	**10**	22	**6**	17
	0,99	7	9	8	13	10	8	7	16	12	6	10	4	8	3	8
	0,95	**58**	42	42	57	47	46	52	45	**59**	43	**29**	**28**	43	13	36
LD,CS95	0,975	27	20	13	29	22	25	30	25	**34**	21	18	12	13	6	17
	0,99	11	7	7	13	8	11	9	13	16	5	8	3	4	3	6
	0,95	57	42	40	49	44	46	52	43	**63**	39	**29**	**28**	38	**14**	36
LD,CS80	0,975	30	18	**11**	33	15	25	29	28	33	14	15	12	13	**5**	15
	0,99	9	7	7	18	7	9	12	14	16	3	7	4	4	3	7
	0,95	46	48	43	52	44	49	51	49	55	46	38	33	53	24	41
LD,CS	0,975	24	26	19	22	21	24	26	25	27	22	18	13	21	12	20
	0,99	8	8	7	12	10	12	7	11	11	7	10	3	7	5	9
	0,95	63	37	17	76	39	43	66	33	83	25	24	19	17	8	23
LD,MM80	0,975	29	13	**6**	**46**	14	23	**40**	14	**45**	**6**	**9**	**5**	**5**	**5**	**9**
	0,99	9	6	4	**23**	7	9	16	2	13	**1**	2	2	**1**	**3**	**3**
	0,95	**63**	37	**17**	**74**	39	42	**64**	34	**85**	**25**	**22**	**19**	**18**	**9**	**22**
LD,MM95	0,975	29	14	**6**	**43**	12	23	**36**	14	**46**	**6**	**10**	**4**	**5**	**5**	**9**
	0,99	9	7	4	**23**	7	9	16	2	14	**1**	3	2	2	2	3
	0,95	**63**	40	**19**	**78**	42	44	**64**	32	**83**	**26**	**24**	**17**	**18**	**8**	**25**
LD,MM	0,975	29	13	**6**	**44**	17	24	**38**	13	**42**	**7**	**9**	**6**	**7**	**5**	**9**
	0,99	10	7	4	**24**	8	8	16	2	13	**1**	**1**	**1**	**1**	2	3

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

Table 3: All covariates at the same time

Estimation current day		Number of exceptions Kupiec test														
Covariates	α	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
	0,95	51	43	44	52	42	42	37	48	46	45	41	42	50	**20**	43
All, but LD	0,975	27	27	17	28	24	20	17	26	25	22	23	18	25	**10**	16
	0,99	10	11	6	13	9	9	4	12	7	10	11	6	6	3	11
	0,95	**60**	**45**	**25**	**64**	40	49	54	39	**74**	37	**26**	**20**	**22**	**8**	**29**
All, but CA,CS	0,975	30	16	**6**	33	19	26	26	21	29	**11**	12	**11**	**8**	**5**	**11**
	0,99	9	7	4	18	10	9	10	6	16	2	3	4	**0**	3	5
	0,95	47	50	42	55	43	48	55	47	54	41	35	32	46	**18**	40
All, but CA,MM	0,975	29	24	17	32	19	23	22	23	30	19	19	15	18	**8**	21
	0,99	9	8	8	12	10	10	7	13	14	5	8	3	6	3	9
	0,95	45	54	42	49	39	47	44	53	52	46	38	33	48	**17**	38
All, but LW	0,975	24	30	19	21	17	24	16	28	28	21	18	14	21	**7**	20
	0,99	8	10	7	9	9	12	6	12	9	7	8	3	6	3	9
	0,95	50	41	39	**58**	43	50	53	49	**67**	41	39	33	46	**16**	40
All, but CS,MM	0,975	29	22	13	**34**	23	24	21	22	27	16	20	15	18	**7**	20
	0,99	13	9	8	17	12	8	8	14	16	6	8	4	5	3	8
	0,95	47	53	42	56	36	46	45	52	48	48	35	31	49	**18**	37
All, but LM	0,975	25	30	18	25	17	24	19	29	21	23	21	13	23	**7**	19
	0,99	9	10	7	13	11	11	3	12	10	9	11	2	6	3	9
	0,95	43	36	43	51	39	46	42	45	53	42	34	42	41	**21**	38
All, but CA95	0,975	26	21	19	25	20	21	17	24	26	17	19	18	16	**10**	17
	0,99	8	7	6	12	12	10	7	16	12	6	9	8	4	4	9
	0,95	41	54	43	50	40	44	36	52	49	43	39	34	48	**17**	38
All, but CA80	0,975	23	28	19	25	19	21	15	29	24	19	18	14	20	**8**	20
	0,99	8	9	8	11	11	11	7	13	8	7	8	3	5	3	10
	0,95	49	53	42	52	36	49	38	51	**61**	42	34	**29**	42	**16**	37
All, but CS95	0,975	28	29	18	22	18	23	17	25	27	20	19	13	18	**5**	20
	0,99	8	9	8	10	9	11	6	16	10	7	11	3	5	3	8
	0,95	43	53	46	52	40	50	39	54	48	42	35	34	44	**17**	38
All, but CS80	0,975	23	28	19	25	20	23	18	27	24	19	20	14	19	**8**	20
	0,99	7	9	8	12	11	11	8	13	8	7	10	3	5	4	10
	0,95	48	51	40	56	40	48	44	48	51	45	38	**29**	51	**17**	38
All, but MM95	0,975	26	30	18	26	18	20	20	26	25	20	17	13	20	**5**	20
	0,99	9	9	8	11	9	9	7	12	11	7	8	2	6	3	10
	0,95	45	52	40	52	39	48	48	50	51	45	37	31	52	**18**	38
All, but MM80	0,975	24	29	18	22	20	24	17	26	25	22	22	13	22	**7**	20
	0,99	8	9	8	9	9	12	5	13	10	8	11	3	6	3	10
	0,95	45	53	42	52	37	48	44	53	**66**	44	37	**30**	50	**17**	37
All cov	0,975	24	29	18	25	19	24	16	28	**35**	20	20	13	21	**6**	20
	0,99	8	9	8	7	10	11	6	13	14	7	11	2	6	3	10

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

(anymore). It's also a worthwile note that the Deutsche Bank and ING, banks that operate in similar markets, have almost identical acceptance and rejections during the out-of-sample period. This can be related to the financial crisis that happened just before and a little bit during the sample period. In overall, the stocks all react in the same way on whether the model gets accepted or rejected; apart from Tsingtao, there is no clear cut on several stocks getting rejected where others get accepted. However, between covariates and combinations of them, there seems to be some pattern visible.

3.3 Comparisons between covariates combinations

When putting a view on Table 1, immediately it becomes apparent that cutting of the history for a day or a week, seems to have a negative effect on the model. The effect is most severe when using only the average of last week's returns on a 95% VaR point. However, when more returns are used in order to create a covariate, the model gets accepted more often and more regular. All the other covariates that are used more in other literature with variance estimation look better than using just a small set of returns in the form of 5 or 20 trading days. When looking at the *MaxMin* covariates, from the single covariates Table 1, there is no significant added value in using the intra daily maximum and minimum returns. The absolute and squared averages have similar acceptances. The discount factors 95 and 80 also are very similar in their own workings. On the other hand, when grouping covariates from the same type with their different factors together, they increase almost to total acceptance.

There is also the notion of actually decreasing the acceptances and thus more rejections. When grouping yesterday's return with weekly or monthly averages, either in a group of two or all three combined, the model actually gets worse. An explanation is that information is too severely added double without any weighing, whereas with the closed averages, the information is discounted and thus any double information gets less influence on the model. Furthermore, adding yesterday's return to the *MaxMin* covariates also worsens the results of the Kupiec test. This adds further to the conclusion that yesterday's return is not a good indicator for current day's conditional density function. Finally, when looking at Table 3, one notices that removing the closed averages returns from all the covariates gives a steep increase in rejections. This gives a good sense on the importance of a covariate that tries to incorporate as much of history in one number for proper estimation.

Now that it is detailed that pfs does create acceptable models for estimating Value-at-Risk for one day ahead, the results of estimating probability density functions ten days ahead are analyzed. Everything in the experiments remains the same, whether data or parameters. The only difference now is that during training and validating out-of-sample is the use of the return ten days ahead as output value instead of the return of the current day.

3.4 Estimating Value-at-Risk ten days ahead

When we look at Tables 4, 5 and 6, immediately there is a resemblance apparent with the previous three tables. They all share the most, if not same characteristics as far as comparison between stocks and comparison between density functions go. The model to estimate Tsingtao's Value-at-Risk is most often rejected and between stocks there is no particular clear cut on which one gets accepted in favor of the other. Also the comparison between covariates when looking ten days ahead has similar results as the comparison when looking one day ahead. This reinforces previous literature like [25] that the stylized fact long term memory holds with variance.

Table 4: Single covariates

Estimation tenth day		Number of exceptions Kupiec test														
Covariates	\alpha	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
	0,95	50	41	**18**	**79**	39	41	**61**	**29**	**66**	**29**	**22**	**21**	**21**	**9**	**23**
LD	0,975	23	20	6	**42**	16	17	29	13	**41**	6	13	10	7	4	11
	0,99	8	8	4	12	6	6	10	4	13	**1**	5	3	2	2	4
	0,95	**63**	44	**19**	**72**	39	39	55	**23**	**69**	**30**	**18**	**16**	**26**	**11**	**22**
LW	0,975	31	20	8	**49**	20	16	28	10	**36**	10	9	6	10	4	10
	0,99	4	8	3	**21**	5	7	9	**1**	18	**0**	**1**	2	2	2	3
	0,95	51	46	**21**	**70**	34	46	53	**24**	**69**	**23**	**24**	**18**	**25**	**9**	**23**
LM	0,975	28	21	5	33	19	18	27	12	**36**	8	13	7	5	5	7
	0,99	2	8	3	18	10	3	7	**1**	11	**0**	3	2	2	**1**	4
	0,95	47	45	45	47	46	54	39	53	44	42	44	33	43	**19**	39
CA95	0,975	22	21	20	29	28	28	23	22	18	22	22	12	20	8	13
	0,99	10	12	9	17	8	10	5	10	6	11	8	4	6	3	8
	0,95	46	48	38	53	41	52	46	37	49	41	36	**24**	44	**13**	38
CA80	0,975	24	27	16	23	25	26	23	18	25	23	18	9	20	5	13
	0,99	9	10	6	13	10	9	8	5	6	6	8	4	6	**1**	6
	0,95	44	44	46	48	46	51	40	46	40	43	42	36	41	**20**	38
CS95	0,975	21	17	20	30	23	26	25	19	15	22	21	14	22	12	12
	0,99	10	9	9	16	10	9	6	11	6	9	8	6	8	4	8
	0,95	44	45	42	51	41	54	50	38	43	40	35	31	48	**14**	37
CS80	0,975	22	25	19	23	25	24	22	14	28	21	17	11	18	6	14
	0,99	9	11	7	10	11	9	9	5	10	8	10	4	5	2	7
	0,95	41	41	41	49	39	46	34	45	33	38	35	44	**23**	**17**	**28**
MM95	0,975	20	21	14	29	21	21	9	23	18	20	17	15	11	5	11
	0,99	9	9	4	11	8	8	**1**	10	5	10	7	9	5	3	5
	0,95	44	47	36	42	40	42	33	43	40	39	31	38	**20**	**12**	31
MM80	0,975	23	20	11	23	19	14	9	22	24	21	15	15	10	4	13
	0,99	7	10	5	6	9	9	3	7	9	8	10	6	3	**0**	5

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

Table 5: Grouped covariates

Estimation tenth day		Number of exceptions Kupiec test														
Covariates	\alpha	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
CA	0,95	49	46	46	49	45	51	40	41	46	43	43	**27**	45	**17**	42
	0,975	24	23	19	26	25	25	22	21	25	24	20	12	24	9	13
	0,99	7	12	7	17	10	12	5	6	7	11	9	6	8	3	8
CS	0,95	46	44	44	45	45	52	42	40	45	41	40	31	47	**20**	40
	0,975	22	22	21	27	24	23	21	17	17	25	21	12	25	11	17
	0,99	8	10	6	15	12	11	3	6	8	11	9	7	10	4	8
MM	0,95	43	43	36	43	42	41	32	42	40	40	34	41	**22**	**15**	**29**
	0,975	21	22	13	25	20	16	10	23	15	22	16	15	11	4	11
	0,99	9	11	5	9	11	9	**1**	9	8	8	9	8	4	2	5
LD,LW	0,95	50	44	**15**	**79**	38	46	**65**	**28**	**75**	**29**	**19**	**18**	**18**	**8**	**23**
	0,975	17	17	6	**44**	17	21	29	14	**43**	6	12	9	6	3	8
	0,99	3	6	4	**19**	5	10	7	2	17	**0**	5	3	**1**	3	2
LD,LM	0,95	52	47	**20**	**67**	34	48	**58**	**27**	**63**	**25**	**19**	**22**	**19**	**9**	**22**
	0,975	24	20	5	**38**	19	15	23	12	**42**	6	12	10	8	4	11
	0,99	6	9	3	17	8	6	9	3	15	**1**	5	2	2	**1**	4
LD,LW,LM	0,95	53	45	**19**	**60**	35	45	**63**	**27**	**68**	**23**	**19**	**20**	**18**	**9**	**20**
	0,975	27	20	8	**37**	17	17	23	12	**37**	7	13	8	6	4	9
	0,99	4	6	3	16	8	8	6	3	16	0	4	2	2	2	2
LD,CA95	0,95	50	52	38	54	46	47	45	43	**60**	42	33	32	50	**10**	35
	0,975	30	27	14	**34**	25	21	20	21	**34**	22	16	12	12	4	18
	0,99	10	11	5	11	10	7	7	9	12	4	9	5	7	**1**	7
LD,CA80	0,95	42	46	34	57	41	45	44	38	**59**	36	**30**	**29**	46	**10**	33
	0,975	26	25	12	26	23	17	22	17	**34**	16	16	10	16	5	17
	0,99	9	9	5	7	9	6	9	8	15	5	9	5	7	**1**	7
LD,CA	0,95	44	47	40	50	45	47	37	41	49	45	41	**29**	52	**15**	37
	0,975	26	27	21	25	22	24	16	17	28	19	17	10	25	5	17
	0,99	6	11	7	13	11	8	7	7	7	6	10	3	9	**1**	8
LD,CS95	0,95	46	51	42	**58**	50	47	40	42	54	42	35	34	49	**11**	38
	0,975	26	26	13	33	26	23	16	18	33	21	18	16	15	5	18
	0,99	9	10	7	11	11	8	6	9	12	4	9	8	9	**1**	8
LD,CS80	0,95	45	46	36	57	42	48	44	35	55	37	**30**	**29**	48	**13**	34
	0,975	25	26	15	26	23	19	17	15	**34**	19	16	14	17	5	16
	0,99	9	9	5	8	10	6	8	7	17	5	9	6	7	**1**	8
LD,CS	0,95	44	55	39	53	43	51	36	43	47	46	38	**29**	55	**17**	34
	0,975	24	28	17	28	25	21	16	25	26	22	17	11	25	8	20
	0,99	8	12	6	10	11	7	5	9	6	7	10	2	10	2	8
LD,MM95	0,95	53	41	**20**	**78**	36	43	**61**	**30**	**69**	31	**22**	**22**	**21**	**9**	**22**
	0,975	26	17	6	**43**	16	16	28	13	**49**	8	14	10	7	4	10
	0,99	7	7	4	**23**	6	6	9	3	18	**1**	5	3	2	**1**	2
LD,MM80	0,95	52	42	**19**	**79**	36	42	**62**	31	**68**	**30**	**21**	**22**	**21**	**8**	**23**
	0,975	25	17	6	**43**	15	16	30	13	**47**	8	15	10	7	4	12
	0,99	7	7	4	**19**	7	6	9	3	17	**1**	6	3	2	2	3
LD,MM	0,95	54	42	**24**	**75**	36	42	**63**	**30**	**68**	32	**21**	**21**	**21**	**8**	**22**
	0,975	23	17	6	**46**	16	18	30	13	**44**	8	15	10	7	4	10
	0,99	8	7	4	**21**	8	7	11	3	17	**1**	4	3	2	**1**	2

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

Similar to one day ahead forecasts, Kupiec test is more interpretable in this case compared to the CM test since the former has a clear rejection region for the null hypothesis of a valid model. Finally, it is important to note that the effects of covariates are similar in one day ahead and ten days ahead forecasts. In other words, the input variables for different forecast horizons are not substantially different

Table 6: All covariates

Estimation tenth day		Number of exceptions Kupiec test														
Covariates	\alpha	AEG	AKZ	ABF	BRC	BAY	BP	DB	HEI	ING	PHI	RDS	SAB	SIE	TSI	UNI
	0,95	47	44	43	49	43	47	39	38	44	42	42	34	51	**18**	39
All, but LD	0,975	24	25	18	26	20	23	16	21	15	24	23	15	21	7	12
	0,99	10	13	7	9	12	10	3	8	2	12	11	5	10	3	7
	0,95	50	45	**27**	51	40	50	45	33	**66**	32	**25**	33	**22**	**8**	**23**
All, but CA,CS	0,975	33	23	10	29	19	21	19	13	**34**	15	12	12	7	4	12
	0,99	10	9	5	11	10	7	5	5	**19**	4	7	5	2	**1**	3
	0,95	45	50	39	46	41	49	35	40	53	39	37	**28**	56	**18**	33
All, but CA,MM	0,975	21	26	16	24	27	19	13	20	28	18	17	11	22	8	20
	0,99	7	8	6	12	11	8	5	6	14	8	10	2	10	3	9
	0,95	45	56	43	47	39	51	38	46	39	43	34	**29**	53	**15**	31
All, but LW	0,975	22	25	18	23	24	25	14	20	22	20	20	14	23	6	21
	0,99	12	9	7	11	13	7	5	7	10	9	8	2	9	3	9
	0,95	44	44	39	46	43	50	45	38	53	40	36	39	48	**14**	40
All, but CS,MM	0,975	27	28	18	23	25	18	16	16	30	17	16	15	23	5	15
	0,99	8	10	7	12	10	9	5	7	17	8	10	4	8	2	9
	0,95	44	55	40	50	38	53	31	46	**61**	39	38	**29**	50	**13**	**30**
All, but LM	0,975	23	23	16	26	22	25	15	21	31	24	20	14	26	6	20
	0,99	12	10	7	11	11	7	3	7	8	11	9	3	9	3	7
	0,95	37	47	38	47	38	49	32	44	49	41	35	46	48	**18**	37
All, but CA95	0,975	21	27	17	23	20	19	15	21	26	22	17	16	20	8	14
	0,99	6	11	8	10	10	10	3	8	13	8	9	8	9	4	8
	0,95	40	53	42	50	41	47	**29**	46	38	33	34	**30**	51	**16**	32
All, but CA80	0,975	16	24	17	26	25	21	15	19	19	24	18	12	22	8	20
	0,99	9	9	7	10	13	10	3	8	8	12	9	3	10	3	9
	0,95	45	53	39	45	36	46	**29**	41	50	37	36	**30**	51	**16**	31
All, but CS95	0,975	23	25	17	28	22	20	13	20	28	24	18	13	24	6	20
	0,99	8	10	7	10	9	10	6	8	11	14	9	2	9	3	9
	0,95	42	54	40	49	40	50	**29**	46	**62**	34	37	**29**	53	**18**	31
All, but CS80	0,975	18	26	17	28	23	22	14	19	**36**	24	19	11	23	8	20
	0,99	7	9	7	11	12	10	6	7	16	12	9	3	9	3	9
	0,95	42	52	44	47	41	50	35	41	47	45	36	**30**	54	**14**	31
All, but MM95	0,975	21	25	17	27	24	25	15	21	23	21	19	11	28	6	20
	0,99	12	10	6	10	12	9	4	7	10	7	9	2	9	3	9
	0,95	43	53	43	50	41	50	32	41	43	35	33	**30**	54	**15**	32
All, but MM80	0,975	21	25	19	29	23	26	15	21	23	25	18	12	27	6	20
	0,99	10	9	8	13	12	8	5	7	9	13	9	2	9	3	9
	0,95	41	53	41	48	38	51	34	44	39	36	33	**30**	54	**15**	32
All cov	0,975	20	24	18	27	24	26	16	20	22	24	19	13	24	6	21
	0,99	10	9	7	12	12	8	5	7	10	13	9	2	9	3	9

Each number shows the Kupiec-test result for a stock estimated by a pfs using a specific covariates combination. A bold number indicates that the number of failures falls out of bounds, therefore pointing towards an overestimating or underestimating model. The lower bound, ideal number and upper bound are: regarding 95% $31 < 43 < 57$; regarding 97.5% $12 < 22 < 33$; regarding 99% $2 < 9 < 18$.

4 Conclusion

4.1 Summary

It is possible to utilize the probabilistic fuzzy system in order to estimate the probabilistic density functions of stock returns using time series as historical data. With fuzzy c-means clustering, historical returns and covariates derived from those returns are categorized and added as input variables in fuzzy rules so that the output of those rules are a range of possible stock return distributions that act as probabilistic density functions. The added value is that in this thesis the derived covariates are now a substitute for future real life covariates, whether being macro-economic variables or sector driven ratings.

By initializing a sound set of parameters on both input space, probability parameters and output space of the probabilistic fuzzy system, all fuzzy uncertainties can be optimized using the maximum likelihood optimization. When all three sets of parameters are optimized, for each rule a probability density function is estimated and by calculating the quantiles and assuming the shape of a normalized distribution, both Value-at-Risk and volatilities of density functions respectively can be measured.

One of the stylized facts what is focused upon is a multiple time horizon by evaluating not only the performance one day ahead but also ten days ahead. Results regarding this particular show that looking on further horizons like ten days ahead is more preferable using probabilistic fuzzy systems rather than looking only one day ahead. More notably is that this suggests more stylized facts such as leverage effects and day-of-the-week effects can be researched with a model such as the probabilistic fuzzy systems.

By using the Kupiec test and the Christoffersen's Markov test, results illustrate a statistical profound evidence that the probabilistic fuzzy system can be utilized on a variety of financial stocks. Whether or not the results are acceptable in the first test remains differing between stocks and the chosen percentiles of the Value-at-Risk measures. However, considering the probabilistic fuzzy system shows more acceptable than unacceptable numbers, this first attempt tells a promising start for a closer look on estimations with more than one input variable.

4.2 Future research

The probabilistic fuzzy system can be used for more derivatives besides financial stocks and their returns. As long as there is uncertainty in the pay-off, one can assess it's risk by modeling any conditional density function. Therefore, options on stocks are an interesting subject to model with the probabilistic fuzzy system. The next step would be a portfolio of any kind of financial derivatives and assess it's complicated combinations of different financial derivatives with the model. Which brings up the thought that better volatility estimation leads to a better control and selection of portfolios. But not only in the application side of the model can be improved on.

The technical part of the system itself can be researched. In this thesis all our covariates were real numbers, but when the number of covariates consisting of real numbers increase, the time of optimization increases exponentially. By using categorical or even binary covariates, increasing the covariates doesn't weigh as much on the optimization time. Also, besides the maximum likelihood function, other optimizing methods can be looked upon to improve the probabilistic fuzzy system. Examples are the genetic algorithm, the ant algorithm, so finding the right heuristic is a topic on its own. As long as the probability matrix stays readable, comprehensible and sensible before and after optimizing, the heuristic is at least sufficient.

Finally, this research considered having lots of data available when choosing stocks from large indices. However, when the Value-at-Risk or volatility of a start-up or unlisted company should be estimated, the probabilistic fuzzy system should still perform well even when it has less available data. Therefore, a topic to research is how to let the probabilistic fuzzy system handle small data sets like with just one year of daily price information.

References

[1] Rui Jorge Almeida, Nalan Basturk, U Kaymak, and JM da Costa Sousa. Estimation of flexible fuzzy garch models for conditional density estimation. Technical report, Erasmus Research Institute of Management (ERIM), 2012.

[2] Rui Jorge Almeida and Uzay Kaymak. Probabilistic fuzzy systems in value-at-risk estimation. *Intelligent Systems in Accounting, Finance and Management*, 16(1-2):49–70, 2009.

[3] Rui Jorge de Almeida, Nalan Basturk, Uzay Kaymak, and Viorel 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*, pages 1–8, New York, NY, USA, March 2012. IEEE.

[4] Rui Jorge de Almeida and Uzay Kaymak. Tail point density estimation using probabilistic fuzzy systems. In *IFSA-EUSFLAT 2009*, pages 1809–1814, Lisbon, Portugal, July 2009. EUSFLAT.

[5] Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

[6] Jan Berg, Uzay Kaymak, and Rui Jorge Almeida. Function approximation using probabilistic fuzzy systems. *ERIM Report Series Reference No. ERS-2011-026-LIS*, 2011.

[7] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function.* Plenum Press, New York, NJ, USA, 1981.

[8] James C. Bezdek, Robert Ehrlich, and William Full. Fcm: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2–3):191–203, 1984.

[9] Christopher M. Bishop. *Pattern Recognition and Machine Learning.* Springer, New York, NJ, USA, 2006.

[10] F. Black. Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 1976.

[11] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[12] Tim Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547, 1987.

[13] Sean D. Campbell. *A Review of Backtesting and Backtesting Procedures.* Divisions of Research Statistics & Monetary Affairs Federal Reserve Board, Washington, D.C., USA, 2005.

[14] Liyan Geng and Junhai 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, pages 103–107. IEEE, 2008.

[15] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

[16] Gloria González-Rivera. Smooth transition garch models. *Studies in nonlinear dynamics and econometrics*, 3(2):61–78, 1998.

[17] Richard C. Grinold and Ronald N. Kahn. *Active Portfolio Management (Second Edition*. McGraw-Hill, New York, NJ, USA, 1999.

[18] Markus Haas, Stefan Mittnik, and Marc S Paolella. A new approach to markov-switching garch models. *Journal of Financial Econometrics*, 2(4):493–530, 2004.

[19] John C. Hull. *Options, Futures, and other Derivatives (5th edition)*. Pearson Education, Upper Saddle River, NJ, USA, 2002.

[20] Jui-Chung Hung. A fuzzy asymmetric garch model applied to stock markets. *Information Sciences*, 179(22):3930–3943, 2009.

[21] P. Jorion. *Value at Risk : the new benchmark for managing financial risk (2nd edition)*. McGraw-Hill, New York, NJ, USA, 2001.

[22] Uzay Kaymak, Willem-Max van den Bergh, and Jan van den Berg. Fuzzy classification using probability-based rule weighting. In *Proceedings of 11th IEEE International Conference on Fuzzy Systems*, pages 991–996, Honolulu, HI, USA, May 2002. IEEE.

[23] Uzay Kaymak, Willem-Max van den Bergh, and Jan van den Berg. A fuzzy additive reasoning scheme for probabilistic mamdani fuzzy systems. In *The IEEE International Conference on Fuzzy Systems*, pages 331–336, St. Louis, Missouri, USA, May 2003. IEEE.

[24] Feng Li, Mattias Villani, and Robert Kohn. Flexible modeling of conditional distributions using smooth mixtures of asymmetric student t densities. *Journal of Statistical Planning and Inference*, 140(12):3638–3654, 2010.

[25] Martin Martens, Dick van Dijk, and Michiel de Pooter. Forecasting s&p 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25(2):282–303, April 2009.

[26] Patrick Nolten. Estimating distributions with probabilistic fuzzy systems. Bachelor of sciences thesis, Erasmus University Rotterdam, September 2011.

[27] Patrick Nolten and Nick Verbeek. Backtesting methods for estimating value at risk using probabilistic fuzzy systems, March 2008. Seminar Computational Economics 2007-2008.

[28] A Thavaneswaran, K Thiagarajah, and Srimantoorao Semischetty Appadoo. Fuzzy coefficient volatility (fcv) models with applications. *Mathematical and Computer Modelling*, 45(7):777–786, 2007.

[29] Jan Van den Berg, Uzay Kaymak, and Willem-Max Van den Bergh. Probabilistic and statistical fuzzy set foundations of competitive exception learning. In *10th IEEE International Conference on Fuzzy Systems*, pages 1035–1038, Melbourne, Australia, December 2001. IEEE.

[30] Jan Van den Berg, Uzay Kaymak, and Willem-Max Van den Bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35(3):291–305, March 2004.

[31] Mattias Villani, Robert Kohn, and Paolo Giordani. Regression density estimation using smooth adaptive gaussian mixtures. *Journal of Econometrics*, 153:155–173, 2009.

[32] Ludo Waltman. A theoretical analysis of probabilistic fuzzy systems. Master's thesis, Erasmus University Rotterdam, 2005.

[33] Du Xu and Uzay Kaymak. Value-at-risk estimation by using probabilistic fuzzy systems. In *Proceedings of the 2008 World Congress on Computational Intelligence*, pages 2109–2116, Hong Kong, June 2008. IEEE.

[34] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100:9–34, 1978.