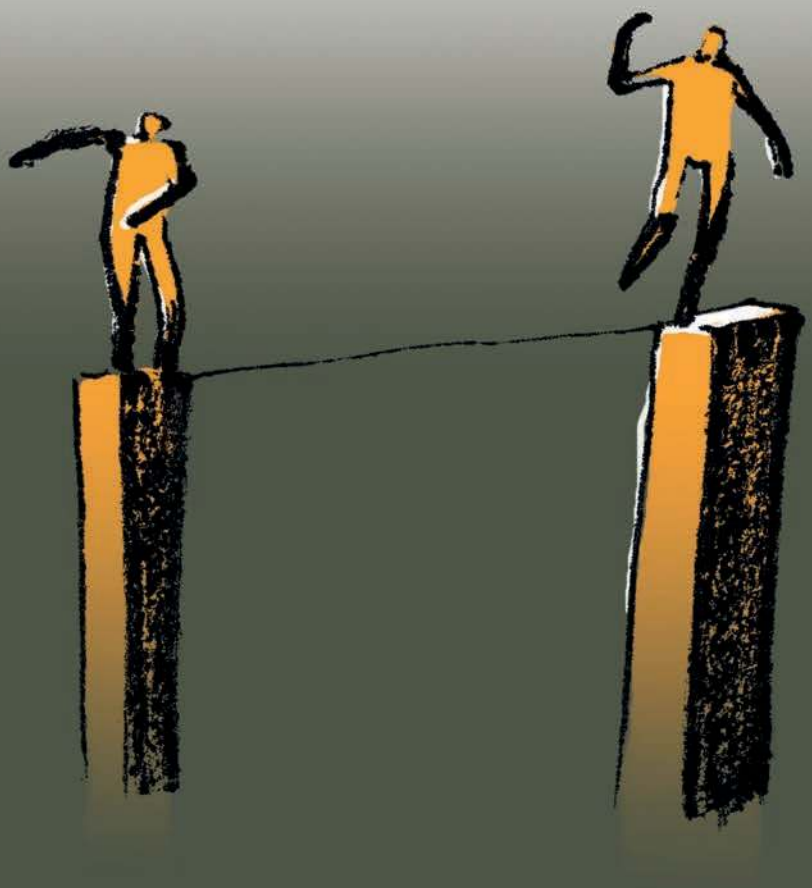
RUI JORGE ALMEIDA

Conditional density models integrating fuzzy and probabilistic representations of uncertainty



Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

Voorwaardelijke dichtheids modellen die vage en probabilistische voorstellingen van onzekerheid integreren

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Thursday 26 June 2014 at 11:30 hours

by

RUI JORGE DE ALMEIDA E SANTOS NOGUEIRA
born in Lisbon, Portugal.

ERASMUS UNIVERSITEIT ROTTERDAM

Doctoral Committee

Promoters:	Prof.dr.ir. U. Kaymak
	Prof.dr. J.M.C. Sousa
Other members:	Prof.dr. P. Groenen
	Prof.dr. T. Martin
	Prof.dr. J. Spronk

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://hdl.handle.net/1765/1

ERIM PhD Series in Research in Management, 310
ERIM reference number: EPS-2014-310-LIS
ISBN 9789058923608
©2014, Rui Jorge de Almeida e Santos Nogueira

Design: B&T Ontwerp en advies www.b-en-t.nl

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

SIKS Dissertation Series No. 2014-27
The research reported in this thesis has been carried out in cooperation with SIKS, the Dutch Research School for Information and Knowledge Systems.

Acknowledgements

The present work, although it bears only my name in the cover, was made possible with the direct and indirect support of many, for which I would like to express my gratitude.

First of all I would like to thank the supervisors of this thesis, Prof.dr.ir. Uzay Kaymak and Prof.dr. João Miguel da Costa Sousa. Both of you always allowed me to follow my own path during my PhD, struggle and persevere. Uzay, thanks a lot for accepting one of João's students as a PhD candidate. Thank you for the continuous unstinting support, guidance and knowledge. Your office door was always open for quick and very frequent meetings, your mail inbox would not retain my emails unanswered for long and you always had a wise and/or witty remark to go with your answers. Thank you also for the pragmatic problem solving approach. All of these and much more I tried to learn from you. João, thank you very much for enticing me to do a PhD with Uzay. Starting even before my master studies, you provided the right encouragement for me to pursue my research. You always had positive, very critical and constructive comments regarding my work. It was always a pleasure to discuss fresh work with you. I would also like to thank Prof.dr. Rommert Dekker for being my initial promotor and guiding me through the first steps of my PhD. Furthermore, I would like to thank Prof.dr. Patrick Groenen, Prof.dr. Trevor Martin, and Prof.dr. Jaap Spronk for the comments on an early version of this thesis and Prof.dr. Didier Dubois, Prof.dr.ir. Jan van den Berg and Prof.dr. Herman K. van Dijk for agreeing to take part in the thesis committee. I am honoured to have such distinguished professors in the committee.

Part of the work in this thesis was made in collaboration with several people. Prof.dr.ir. Jan van den Berg, Dr. Viorel Milea, Dr. Nalan Baştürk and Nick Verbeek thank you for the discussions and fruitful cooperation which resulted in several chapters of this thesis being published. Furthermore, I was fortunate to work in other diverse projects, some of them ongoing, which are not included in this manuscript. Nonetheless, the benefits of these collaborations are reflected in this thesis. Thank you Prof.dr. Bernadette Bouchon–Meunier, Dr. Marie–Jeanne Lesot, Gilles Moyse, Prof.dr. Thierry Denœux, Dr. Didier Coquin, Dr. Lionel Valet, Prof.dr. Herman K. van Dijk, Dr. Lennart Hoogerheide, Dr. André Fialho, Dr. Rúben Pereira, Dr. Milan Lovric and Dr. Tommi Tervonen for sharing your knowledge. Part of the work of this thesis was supported

by the external grants: EUSFLAT Student Grant, IEEE-CIS Walter Karplus Student Research Grant and COST Action IC0702 grant ECOST-STSM-IC0702-180411-006041. I would like to once again thank the external institutions that provided me with these grants for their support.

As a previous member of the Econometric Institute (EI-ESE), Erasmus Research Institute of Management (ERIM) and also benefiting from the support of Tinbergen Institute (TI), I would like to thank all their faculty and staff for the help in all research and administrative matters. I would specially like to thank all with whom I had direct contact, namely the current and previous ERIM directors of doctoral education, Prof.dr. Finn Wynstra and Prof.dr. Patrick Groenen and also from ERIM, Tineke van der Vhee, Olga Rook, Miho Iizuka, Marisa van Iperen and Kim Harte who recently helped me greatly with the final thesis procedure, from TI, Judith van Kronenburg and Ine Driessen, and finally from EI-ESE, Ursula David, Marjon van Hees-Gouweleeuw, Anneke Kop and Carien de Ruiter-Coopmans. Your work may sometimes go unnoticed but it is extremely important for PhD students like myself. I would also like to thank Prof.dr. Albert Wagelmans for providing me with the opportunity to expand my teaching experience and all the great colleagues in the EI-ESE 'computer science' group, namely Yingqian Zhang, Flavius Frasincar, Wim Pajls and Eelco van Asperen. Finally, I would like to thank Shouhe Kuo and all of the student assistants, like Victor, Robert, Putri just to name a few, who helped me greatly with the Skills IT course.

The last stretch of this thesis was completed while taking a new challenge in Eindhoven University of Technology, in the Information System research group of the School of Industrial Engineering. I am extremely grateful to Prof.dr.ir. Paul W.P.J. Grefen for the support and trust placed on me. Furthermore I would like to thank my new colleagues Pieter Van Gorp, Bilge Aydın, Claudia Chituc, Remco Dijkman, Rik Eshuis, Rob Kusters, Ayça Tarhan, Oktay Türetken, Jos Trienekens, Irene Vanderfeesten and Anna Wilbik for the empathic, productive and challenging work environment. Finally, thank you Annemarie van der Aa for receiving all my many requests with a smile, including the improvements for the Dutch summary of this thesis. All and any mistakes are still my responsibility.

The PhD trajectory is shared with many fellow students and friends, some new, some old, some close, some far away, all of whom I would like to acknowledge. First of all I would like to thank my office mates, Karim, Andrei, Peter, Mathijn, Lanah, Jeanine and Nüfer. It was always a pleasure to discuss a huge panoply of subjects with all of you. Karim, thanks also for sharing late afternoons in The Smitse and I hope the Portuguese words that I unintentionally shared with you while programming will be of some use. Nüfer, it is a pleasure to continue our office conversations, now in the train, without any guilt for work not being done instead. Thanks a lot Eelco for introducing me, the new student, to most of the PhD students present early morning on the 9th floor. Life in Rotterdam was more pleasurable thanks to Simon, Inge and more recently

Tobi's welcoming barbecues often with Cem and Anil (with whom I would dance the last bus away), Felix, Katerina and blue eyed Isabella's courteous dinners, Asli who would kindly share her cakes with me, Viorel, Oke, Sun, Eran and Kyle who took my money, Murat (where's your brother's passport?) and Senne, Melek, Julia, Dan Zhang, Agapi, Frederik, Alexander (this time I got your names correct), Berk, Sarah who always has a positive pragmatism in life, Aurelie with whom I like to share relaxed soirées, Cathelijne (brain pause inducing smile blondie), Bas (gnarly, pumped and stoked adventure companion), Katerina (with an ever positive attitude), Tulay, Irina (doubted your existence for a while, glad you are real), Ayca, Ozgen, Tugba, João (who never fails to share the delicious contents of his luggage), Rui, Rui, Ivo, Ines, Elisa, Lilia, Paula, Nicole, Adam, Julian (the latter three for the friendly hospitality in Compiègne) and Susana, whom only in Rotterdam I had the pleasure to know.

I could also not fail to mention those who extend the R'dam hospitality border to wherever we meet: Tommi-Tervomemeter-Portuguese-speaking-fuzzy-lover and Karin in the sunny land of great Lungos, Gianluca in awe inspiring Vulcano, Burcu (BonJour), Selen (Green-Dress), Duygu (Cat-Woman), Duygu, Bengü, Burcu (sisters) in Istanbul and Franzi (Xica) for simple but indulging bon vivant dinners, luckily now only two hours away.

For years Catalunya is one of my favourite destinations in no small part due to the welcoming that I always received. Moltes gràcies Elisabet i Josep Anton, Raquel, Enric, Carlos i Silvia, Gerard, Joan i Maria Teresa, Marc i Maria Jesús, David i Carmina, per donar-me la benvinguda i compartir la gastronomia i la cultura Catalana, Irene i Mercè per l'acolliment en Barcelona i Rut que sempre tenia una abraada llest per a mi. Gràcies Carles per ensenyar-me a veure amb altres ulls els Països Baixos i Monica pel teu somriure i el desig de fer plans per al futur davant l'adversitat que es van convertir en una font d'inspiració per a mi. Un agraïment especial per a Montsant que sempre va encoratjar i em va donar forces abans de l'inici d'aquest estudi de doctorat, fins i tot quan significava algunes dificultats entre nosaltres. La vida dona moltes voltes, però sempre recordo amb afection la teva ajuda i generositat (i perdó pels errors gramaticals).

Talking Portuguese in the streets of Rotterdam can trigger spontaneous reactions, which perdured across France, Belgium and The Netherlands. Thank you Sara for being a social magnet and a cherie. My sincere thanks to Nuno, one of the first persons I met in the 9th floor, who has an uncanny capacity to maintain a train of thought, Carlos, who enjoys dialogues populated with periphrase and hypallage which has led to the dysphemism of being sad, and Joana with whom, although for brief months, we shared fantastic moments. Together, we share the tendency for lengthy banters and light-hearted Tuesdays.

Although I moved away from Lisbon, I am always welcomed back by many friends. Thank you Professora Benilde and Professor Mata for what has already became our traditional lunches. Thanks Caldeira (now my fellow resident), Hugo and Barbara (always supportive and helpful,

not to mention fun), Vinagre, Rodrigo, Mestre, Nelson, Josefina, Bruno, Artur, Pires and all the Gente Gira for the dinner parties, which I greatly enjoy. Thank you also Ana Sininho for the fondness support at the beginning of my PhD, Isabel for always welcoming me with a heart-warming smile, Teresa, luckily our paths crossed again and brought Guilherme and Tico along and finally to Nuno: since young we share a brother-like relationship always evolving with new events and distances and this PhD was no exception.

A special word to my mother Carolina and brother Pedro: A minha escolha de explorar novos caminhos longe de casa não podem ser substituídos por longos telefonemas ou curtas visitas. Muito obrigado Mãe pelo apoio e por entender as minhas escolhas, de alguma forma semelhantes as tuas.

Last, but not least, I would like to thank my Paranimfen, Viorel and Nalan. Viorel, Chief, our predilection for bootleg music, bingo nights and general life discussions (filled with your unique descriptions of even the most mundane events) are treasurable moments in my PhD. Nalan, an indication of the influence you had in this thesis is the number of times your name (which only took me a few months to memorize and few years to correctly pronounce) appears in it. Besides this, we still had time to share many moments and even London-Venice trips, which should continue to expand. Together, the three of us shared late lunches, teaching tips, childhood stories and nights filled with game-theoretical decisions that would end up with musketeers like equilibriums. No matter how I attempt to phrase it, it will hardly convey my thoughts, so I'll summarize it instead: I'm there. Thank you for once again accompanying me.

Rotterdam, May 2014
Rui Jorge de Almeida

Table of Contents

1 Introduction **1**

1.1 Conditional density estimation 1

1.2 Fuzzy and probabilistic representations of uncertainty 2

1.2.1 Fuzzy systems 3

1.2.2 Probability theories 5

1.3 Do probability and fuzzy sets exist? 6

1.4 Research goal 8

1.5 Contributions of this thesis 9

1.6 Thesis outline 10

1.6.1 Work developed in this thesis 10

1.6.2 Thesis organization 12

2 Conditional density estimation using probabilistic fuzzy systems **15**

2.1 Introduction 15

2.2 Probability and statistics of fuzzy events 18

2.2.1 Probability of fuzzy events 18

2.2.2 Statistical issues 20

2.3 Probabilistic fuzzy systems 24

2.3.1 Outline 24

2.3.2 Reasoning 26

2.4 Relation to deterministic fuzzy systems 33

2.5 Discussion 34

2.6 Conclusions 36

3 Analysing probabilistic fuzzy systems **37**

3.1 Introduction 37

3.2 Probabilistic fuzzy systems revisited 39

3.2.1 Fuzzy histogram model 40

3.2.2 Probabilistic fuzzy output model 41
3.2.3 Outputs of probabilistic fuzzy system 42
3.2.4 Necessary conditions for probabilistic fuzzy systems 46
3.2.5 Probabilistic fuzzy systems with multiple outputs 47
3.3 Function equivalence to systems with deterministic output 48
3.3.1 Additive fuzzy models 49
3.3.2 Zero-order Takagi-Sugeno fuzzy models 52
3.3.3 Radial basis function networks 53
3.4 Probabilistic fuzzy system parameters 56
3.4.1 Antecedent membership functions 56
3.4.2 Consequent membership functions 59
3.4.3 Conditional probability parameters 60
3.5 Examples: Synthetic data parameter estimation 63
3.5.1 Function approximation 63
3.5.2 Conditional density approximation 67
3.6 Application on US inflation data 71
3.6.1 Comparison of different estimation methods for the inflation model . . 73
3.6.2 Inflation patterns according to the rule outputs 75
3.6.3 Density estimates of inflation 77
3.6.4 Inflation density estimates for multiple time periods 78
3.7 Conclusions 79

4 Probabilistic fuzzy systems in Value-at-Risk estimation 81
4.1 Introduction 81
4.2 Value-at-Risk models 83
4.2.1 Value-at-Risk 84
4.2.2 Volatility estimation 84
4.2.3 Fuzzy VaR models 85
4.2.4 Model validation 85
4.3 Probabilistic fuzzy systems 86
4.4 Probabilistic fuzzy models of value at risk 90
4.4.1 Mamdani PFS 90
4.4.2 Fuzzy histogram model 92
4.5 Experimental study 94
4.6 Results 98
4.7 Conclusions 103

5 A multi-covariate multi-horizon conditional volatility model using PFS 105
5.1 Introduction . . . 105
5.2 Probabilistic fuzzy systems . . . 107
5.2.1 Fuzzy histogram model . . . 108
5.2.2 Probabilistic fuzzy output model . . . 109
5.2.3 Equivalence of reasoning mechanisms . . . 110
5.2.4 Probabilistic fuzzy systems with multiple outputs . . . 111
5.3 Value-at-Risk estimation . . . 112
5.3.1 Value-at-Risk . . . 112
5.3.2 Volatility estimation . . . 113
5.3.3 Model validation . . . 114
5.4 Multi-covariate multi-horizon probabilistic fuzzy model . . . 115
5.4.1 Model covariates . . . 115
5.4.2 PFS model parameters . . . 116
5.5 Conditional density estimation for the S&P 500 index . . . 119
5.5.1 Fuzzy histogram model interpretation . . . 121
5.5.2 Probabilistic fuzzy output model interpretation . . . 123
5.6 Value-at-Risk estimation for the S&P 500 index . . . 124
5.6.1 Multi-covariate model results for the S&P 500 index . . . 125
5.6.2 Seasonality analysis for the S&P 500 index . . . 128
5.6.3 Multiple horizon forecasts for the S&P 500 index . . . 130
5.7 Conclusions . . . 131

6 Estimation of flexible fuzzy GARCH models for conditional density estimation 133
6.1 Introduction . . . 133
6.2 GARCH model and extensions . . . 136
6.3 Proposed fuzzy GARCH model . . . 140
6.3.1 Interpretation of the model . . . 144
6.3.2 Parameter estimation . . . 146
6.4 Examples: Synthetic data parameter estimation . . . 147
6.4.1 Fuzzy GARCH data with constant mean . . . 148
6.4.2 Fuzzy GARCH data with general time varying mean . . . 151
6.5 Application: Conditional density estimation of S&P500 returns . . . 152
6.6 Conclusion . . . 158

7 Summary and Conclusions **161**
7.1 General findings 161
7.2 Conclusions 162
7.3 Future Research 163

References **165**

Summary in English **183**

Nederlandse Samenvatting (Summary in Dutch) **185**

Curriculum Vitae **187**

Chapter 1

Introduction

"It is better to be vaguely right than exactly wrong."
– Carveth Read, *Logic, deductive and inductive (1898), p. 351*

CONDITIONAL density estimation is an important problem in many areas such as system identification and machine learning, where the predicted density is typically highly non-linear and multimodal (Bishop, 2006), artificial intelligence (John and Langley, 1995), empirical economics (Li et al., 2010), macroeconomic analysis (Diebold et al., 1998), quantitative finance and risk management (Glosten et al., 1993), where financial time-series typically possess non-trivial statistical properties, such as fat tails, asymmetric distributions and changing variation over time (Villani et al., 2009).

This work considers the general problem of conditional density estimation, *i.e.* estimating and predicting the density of the response variable as a function of covariates. The semi-parametric models studied and developed in this work combine fuzzy and probabilistic representations of uncertainty, while making very few assumptions regarding the functional form of the estimated density or changes across the space of covariates. These models possess sufficient generalization power to approximate a non-standard density and ability to describe the underlying process using simple linguistic descriptors despite the complexity and possible non-linearity of these processes.

1.1 Conditional density estimation

Conditional density estimation is an approximation of the probability density $\hat{f}(\underline{y}|\mathbf{x})$ of a stochastic output variable $\underline{y}$ given an observed vector of attribute value $\mathbf{x}$. A conditional density estimator provides an entire density function for the target variable, while a regression estimator provides a deterministic prediction, the expectation $E[\underline{y}|\mathbf{x}]$. Regression analysis will not ade-

quately portrait a system if the conditional distribution possess non-trivial statistical properties, such as multimodality, asymmetry, or heteroskedastic noise. In these cases the estimate of the conditional distribution contains more information. The full output distribution allows several quantities of interest to be extracted, including the expectation, modes, moments and outlier boundaries. Furthermore, conditional density estimation makes it possible to quantify and visualize the prediction intervals that contain the target variable with a specified probability.

The estimation of conditional distributions is an important problem in empirical economics, such as macroeconomic and financial applications. It is often desirable to estimate not only the expected inflation levels conditional on covariates but also the complete inflation density since it can be used to obtain an estimated range for inflation. Based on these predictions, a central bank can adjust the monetary policy instruments accurately (Diebold et al., 1998). A similar reasoning applies in the approximation of financial returns' distribution where 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 left-tail quantiles of the estimated returns distribution, such as Value-at-Risk or Expected Shortfall (Jorion, 2006), and it cannot be assessed from models providing point forecasts.

Different methods can be used to estimate conditional densities, conditional on past information, or other macroeconomic variables. A popular approach where volatility, and hence the return distribution, changes dynamically based on distribution assumptions is the Generalized Autoregressive Heteroskedasticity (GARCH) model (Bollerslev, 1986). 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 such as smooth transition GARCH models (González-Rivera, 1998) and regime-switching GARCH models (Haas et al., 2004). In time series settings, non-parametric estimation of conditional densities is often desirable for forecasting, as knowledge of true distribution rarely exists (Yatchew, 1998; Fan, 2005). Semi- and non-parametric methods, such as quantile regression (Koenker and Hallock, 2001; Koenker, 2005), kernel density estimation (Fan et al., 1996; Bashtannyk and Hyndman, 2001), are widely used in econometrics.

This thesis proposes semi-parametric conditional density estimation models, that can incorporate possible non-linear relations between variables, while allowing for a parsimonious and interpretable description of the dynamic behaviour of the system.

1.2 Fuzzy and probabilistic representations of uncertainty

An important aspect of the models considered in this thesis is the combination of different representations of uncertainty. Many researchers have argued that fuzziness and randomness

are actually describing the same phenomena or at least they presume that fuzzy set theory is a generalization of probability theory or the other way around (Thomas, 1995; Goodman and Nguyen, 2002). However the concepts behind fuzzy set theory and probability theory are different (Zadeh, 1968, 1995; Bertoluzza et al., 2002). The concepts behind fuzzy set theory and probability theory need a suitable practical interpretation in order to be used meaningfully. In the following sections the concepts of fuzzy sets and probability are explained.

1.2.1 Fuzzy systems

As a mathematical notion, a fuzzy set F on a finite universe U is unambiguously defined by a membership function $u_F : U \to [0, 1]$. The mathematical object representing the fuzzy set is the membership function $u_F(x)$ indicating the grade of membership of element of $x \in U$ in F.

The elicitation of membership grades and interpretation of fuzzy sets, membership grades and fuzzy rules is the base of much confusion and negative comments. Recently, there has been an effort to clarify the different meanings of fuzzy sets (Dubois and Prade, 1997), membership grades (Dubois and Prade, 2012) and fuzzy rules (Dubois and Prade, 1996).

Fuzzy sets are usually related to vagueness. This vagueness is not defined as uncertainty of meaning but instead as the standard definition of vagueness with the possession of borderline cases (Sorensen, 2013), (see Dubois et al. (2005) for a detailed discussion on this subject). In the literature, fuzzy sets are used to represent three different concepts: gradualness, epistemic uncertainty and more recently bipolarity (Dubois and Prade, 1997, 2012). These basic concepts of fuzzy sets differ greatly from each other (Dubois and Prade, 2012) and they can be summarized as

Gradualness Refers to the original idea of Zadeh (1965) that many categories in natural language are a matter of degree, including truth. The fuzzy set is used as representing some precise gradual entity consisting of a collection of items. Such fuzzy sets are conjunctive and can be called ontic fuzzy sets [1]. The gradualness is indicated through membership. The transition between membership and non-membership is "gradual rather than abrupt" (Zadeh, 1965). The gradualness can be linked to different situations:

1. The boundaries of the set are precisely known, but it is not possible to measure it (or indicate it) precisely. An example is the definition of a meaningful area (*e.g.* forest zone) in a grey level image. Inherently, the boundary of this zone is gradual.
2. It is possible to measure each element of the set precisely (*e.g.* position of the trees), the boundaries of the set are known, but a (crisp) definition of its boundaries is not

[1] This description follows the ideas presented in Dubois and Prade (2012) where ontic sets and epistemic sets follow an analogy with ontic and epistemic actions in cognitive robotics (Herzig, 2003).

precise. Following the above example, the gradualness in this case is a result of the density of trees slowly decreasing in peripheral zones.

3. The uncertainty is linked to a fuzzy predicate referring to a gradual concept (*e.g.* "dense" forest zone). In this case the boundaries are known (even if fuzzy), the measure of each element is precise, but the fuzzy predicate indicates gradualness.

Epistemic uncertainty Refers to the idea of partial or incomplete information. The base is that sets are epistemic constructions and represent incomplete information about the world. As such it is described by a set of possible values of some quantity of interest, one of which is the right one, while elements outside this set are considered impossible. This idea is the basis of possibility theory (Zadeh, 1978). An example is that an agent only has a rough idea of the size s of a forest zone, and provides an interval $[a, b]$ as containing the right value of s. Such an interval is the disjunction of mutually exclusive elements. The interval itself is subjective (it is the knowledge of the agent), but has no intrinsic existence, even if it refers to a real fact.

Bipolarity This recent interpretation by Dubois and Prade (2012) refers to the idea that information can be described by distinguishing between positive and negative sides, possibly handled separately, as it seems to be the case in the human brain. In this case the membership scale of a fuzzy set is a univariate bipolar scale (Dubois and Prade, 2012).

Following the different interpretations of fuzzy sets, the degree of membership $u_F(x)$ of an element x in a fuzzy set F can be used to express degree of similarity, degree of preferences (in utility functions) and degree of uncertainty (Dubois and Prade, 1997). These interpretations can be summarized as:

Degree of similarity The membership degree $u_F(x)$ represents the degree of proximity of x to prototype elements of F. This view is used in clustering analysis and regression analysis, where the problem is representing a set of data by the proximity between pieces of information. It is also at used in fuzzy rule-based control techniques, where the similarity degrees between prototype situations described in the condition parts of the rules and the current one are the basis for the interpolation mechanism between the conclusions. A simple example (Dubois and Prade, 1997) is the classification of cars of known dimensions in categories of $F = \{\text{big cars}, \text{regular cars}, \text{small cars}\}$. If the prototype of the category *big cars* is a *Mercedes Class S*, then we can construct a measure of distance between any car to this prototype, where the distance is a measure of similarity.

Degree of preference The membership degree $u_F(x)$ represents an intensity of preference of object x, to a set F of preferred objects. Alternatively, F represents a set of values of a

decision variable x' and $u_F(x)$ represents the feasibility of selecting x as a value of x'. This view of fuzzy sets as criteria or flexible constraints is used in fuzzy optimization and decision analysis. An example is an agent buying a big car. In this case the membership degree will reflect the degree of satisfaction of cars chosen by the agent to the class of *big cars*, according to the criterion *size*. In this case the membership indicates the preference of the agent.

Degree of uncertainty The membership degree $u_F(x)$ represents the degree of possibility that a parameter p has a value x given that all that is known is that p is F. This view is used in possibility theory and is applied in expert systems, and artificial intelligence. An example is when an agent says that he saw a big car. The variable whose value is the name of the big car is uncertain, all we know is that size is big. In this situation the membership grade of a given car (which can be measure precisely) to the class of big cars reflects our degree of possibility that this kind of car is the same as the one seen by the person. When this membership degree is high, we are still uncertain about which particular large car the agent saw. If the membership degree is low then any large car can be rejected as a very implausible candidate.

1.2.2 Probability theories

In mathematical terms a probabilistic measure Pr[2] of an experiment ϵ yet to be performed, is a mapping $2^U \rightarrow [0,1]$ that assigns a number $\mathrm{Pr}(A)$ of event A to each subset of U, satisfying the Kolmogorov axioms. $\mathrm{Pr}(A)$ is the probability that a generic outcome of ϵ, an ill-known single-valued variable x, hits set A. If the outcome of ϵ is such that $x \in A$, then we we say that event A as occurred. In this case there is uncertainty about the occurrence of any particular x and consequently of event A. This uncertainty is described by $\mathrm{Pr}(A)$. All probability theories generalize the "law of the excluded middle", where an element either belongs to a (well-defined) set A, its complement A^c, but not both.

There are several interpretations of $\mathrm{Pr}(A)$. Although the semantics of probability theory are clear and well understood, they are not unique, leading to differing views on the semantics of probabilities (de Finetti, 1974; Hesse, 1975; Khrennikov, 1999). The concept of numerical probability emerged around the 17th century (Hacking, 1975), related to games of chance and reliability of testimonies. Classical probability theory is based on symmetry arguments (*e.g.* six outcomes in a normal looking die), while frequentist probabilities represents a physical random phenomenon over long-run frequencies and accounts for variability of (precise) observations.

[2] In this work we follow de Finetti notation of 'Pr' which referred to probability, price and prevision, interchangeably.

Following a different perspective, a subjective probability represents the epistemic state of an agent, assuming degrees of belief, related to a status of uncertainty. It can be equated to exchangeable betting rates. Bayes' theorem is the logical tool to update the probability in the light of new pieces of information.

It is important to note that subjective probability theory does not possess a single formalism on how to obtain a particular $\Pr(A)$, which may be non-numerical (de Finetti, 1972). Uniformity of judgements between individuals can occur with respect to games of chance (impressions of symmetry) or where statistical historical data (frequencies) are available. Nonetheless the degree of difference between individuals subjective probabilities will depend on the particular circumstances under which these judgements are elicited, *i.e.* the initial probabilities are the opinions of the individual expressing a judgement (de Finetti, 1972). The theory of imprecise probability was proposed as a solution for the difficulty of consistently accounting for incomplete information using a single subjective probability (Walley, 1991).

1.3 Do probability and fuzzy sets exist?

This question was already answered negatively by de Finetti (1974) and Dubois and Prade (2012). The general idea is that probability as a numerical measure does not exist independently of the human mind and may be subjective regarding an individual, similarly to the epistemic view of fuzzy sets (Dubois and Prade, 2012). The objective of this section is not to try to answer this question but instead discuss other aspects of modeling that arise from this question and are used in this thesis.

In this thesis we follow the approach presented in Nau (2001), where this 'non-existence' is not a problem for statistical inference, decision analysis or economic modeling. Both concepts are useful to model different aspects of uncertainty. Furthermore, this work follows Zadeh's idea that "probability theory and fuzzy logic are complementary rather than competitive" (Zadeh, 1995). This cooperation can take the form of theoretical developments (Baldwin et al., 1996; Singpurwalla and Booker, 2004; Coppi et al., 2006; Coletti and Scozzafava, 2006) or of new models containing fuzzy sets and probabilities (Meghdadi and Akbarzadeh-T., 2001; Liu and Li, 2005; Zhang and Li, 2010, 2012) or the ones presented in this work. Claims on the superiority of a certain theory as a superior representation of uncertainty (e.g. Klir, 1994), which appear to be caused by misunderstandings (Dubois and Prade, 1993), are not addressed in this work. All theories discussed in this work are rigorous approaches for modeling uncertainty from different perspectives, particularly uncertainty represented by fuzzy sets and probability theory, leading to a deeper understanding of the concept of uncertainty.

In economics uncertainty is commonly modelled using probabilities. In some cases the use of the calculus of probability appears as if there was no question whatsoever about the validity of such use (Rudra, 1966). Probability theory is based on a rigorous mathematical construct, has proved useful in many applications and has clear semantics, although not unique as explained in Section 1.2.2. Furthermore, historically, there are several theories in economics which are intrinsically connected with the calculus of probability. An example is the classical theory of consumer decision making (von Neumann and Morgenstern, 1944), which is based in a set of axioms stated in the language of probabilities for maximization of expected utility. Econometrics encompasses a vast array of mathematical models used in empirical economics, macroeconomic analysis, quantitative finance and risk management, where the goal of the analysis is often the left tail of a probability density. Fuzzy systems are typically used for approximating deterministic functions, in which the stochastic uncertainty is ignored. The models studied in this thesis are used to produce a simplified and imperfect substitute of reality as observed via precise data modelled using fuzzy sets, and have as output a probability density function, which makes them suitable for the aforementioned analysis. In this work these fuzzy sets were interpreted with their relation to gradualness although other interpretations are possible. Interestingly, the elicitation of these sets has a similar problem to that of elicitation of subjective probabilities: they depend on the individual. As chapter 5 and chapter 6 show, the models studied are robust and can adapt to different elicitations of the input fuzzy sets.

An important issue with economic modelling is that the considered variables should have a relation to the problem under study, based on economic theory or empirical evidence. This relation may not be direct or linear, due to the inherited complexity or impossibility of performing direct measures. To solve this problem, proxies are used in different economic problems:

- When exchange-traded derivatives for jet fuel are not available, futures contracts on commodities related with jet fuel, such as crude and heating oil are used instead (Hull, 2000).
- The GDP per capita is used as a proxy of quality of life in a country (Montgomery et al., 2000; Becker et al., 2005)
- Proxy variables, such as market-to-book assets ratio are used for a firm's investment opportunity set (Adam and Goyal, 2008).

Fuzzy systems lend themselves to this type of modelling using proxy variables. They have been successfully adopted in many domains and used extensively in commercial products used daily such as microwaves and washing machines. These types of systems have the advantage that they can be used to describe knowledge of the process in the form of rules, without very strict assumptions, which is very natural for human to understand. Furthermore, these type of

models can have a well understood parsimonious structure *e.g.* autoregressive-moving-average (da Costa Sousa and Kaymak, 2002).

In summary, this work presents conditional density models which integrate fuzzy and probabilistic representations of uncertainty. The concepts behind fuzzy set theory and probability theory are based on rigorous mathematical constructs and, although not unique, well understood semantics, which have been useful in many applications. The combination of these concepts leads to models that can deal with the concept of gradualness or epistemic uncertainty and also the concept of stochastic uncertainty. In this thesis, these models are shown to be useful to model non-linear relations without strict assumptions where regression density estimation is the goal of the analysis.

1.4 Research goal

The purpose of this thesis is to develop new models capable of flexible estimation of conditional densities without strict distribution assumptions. Such models aim to incorporate possible non-linear relations between variables, while allowing for a parsimonious and interpretable description of the dynamic behaviour of the system. Keeping this approach in mind the research goals of this thesis are threefold:

1. Establishing different models that encompass fuzzy and probabilistic representations of uncertainty capable of conditional density estimation. Such models aim to incorporate possible non-linear relations between variables, steaming from actual non-linearities in the system or caused by the use of proxy variables. Furthermore, these models allow for a parsimonious and linguistic description of the dynamic behaviour of the system.

2. Providing a formal description and analysis of these models, considering their different parts and elements. Furthermore, studying the properties, estimation issues, model interpretation, and differences with other similar models.

3. Application of these models combining fuzzy systems and probabilities to financial problems following well studied economic relations. Comparing the findings from these applications to previous work on similar financial problems.

1.5 Contributions of this thesis

The contributions presented include the development and study of models that integrate fuzzy and probabilistic representations of uncertainty for conditional density estimation as well as their application to real world financial analyses. These contributions can be summarized as:

- Development of two new systems, namely the fuzzy GARCH and probabilistic fuzzy systems (PFS), for conditional density estimation, that combine fuzzy and probabilistic representations of uncertainty. Such systems can capture different properties of data, such as fat tails, skewness and multimodality in one single model.
- Formal description of probabilistic fuzzy systems. Two possible and equivalent reasoning mechanisms are presented, which lead to two different interpretations of this type of systems.
- Analysis of the necessary conditions for a probabilistic fuzzy systems, such that the estimated output density is a proper probability density function and subsequent higher moments derived from this density exist.
- Description of the relation of probabilistic fuzzy system with different types of non-linear deterministic systems that have universal approximation capability, such as fuzzy systems and radial basis functions.
- Analysis and explanation of the parameters of PFS, as well as approximation capabilities of these systems in synthetic examples for function approximation and conditional density estimation.
- Application of PFS in multi-horizon estimation of quarterly U.S. inflation where point estimates and the density estimates of inflation are relevant for a comprehensive analysis of volatility and mean changes, and for policy making.
- Application of different PFS in the estimation of Value-at-Risk using multi-covariate and seasonal models. Particular relevance is given to the interpretation of PFS and its use in the study of stylized facts, such as seasonality and volatility clustering.
- Proposal and formal description of fuzzy GARCH model for conditional density estimation.
- Discussion of properties, estimation, relation to similar models and different interpretation of the proposed fuzzy GARCH model.

- Application of the fuzzy GARCH model for density forecast of the S&P 500 daily returns series. Possibly complex effects of current market information on future returns are explained using simple linguistic descriptors in combination with the well studied GARCH-type rule base system.

1.6 Thesis outline

1.6.1 Work developed in this thesis

This thesis presents two systems to estimate conditional densities, making very few assumptions regarding the functional form of the estimated density or changes across the space of covariates. An important aspect of these semi-parametric models is that they combine fuzzy sets and probabilistic uncertainty, making them quite simple to estimate and understand. Nonetheless, as the practical applications presented show, these models possess sufficient generalization power despite the complexity and possible non-linearity of the modelled processes. Multi-horizon, multi-covariate and seasonal models are considered. Although these models possess good approximation capabilities, they provide a simple interpretation essential for process understanding. Special attention is given to the interpretation of the models such that they can be useful in many fields such as macroeconomic analysis, quantitative finance and risk management. The bulk of the work presented in this thesis is on the formal description, analysis and practical applications of probabilistic fuzzy systems. The last chapter refers to the combination of fuzzy systems and the Generalized Autoregressive Heteroskedasticity (GARCH) model where the latter is widely used in empirical economics.

In chapter 2 we consider conditional density approximation by fuzzy systems. Fuzzy systems are typically used for approximating deterministic functions, in which the stochastic uncertainty is ignored. We propose probabilistic fuzzy systems (PFS) in which the probabilistic nature of uncertainty is taken into account. These systems take also fuzzy uncertainty into account by their fuzzy partitioning of input and output spaces. We discuss an additive reasoning scheme for probabilistic fuzzy systems that leads to the estimation of conditional probability densities, and prove how such fuzzy systems compute the expected value of this conditional density function. We show that some of the most commonly used fuzzy systems can compute the same expected output value and we derive how their parameters should be selected in order to achieve this goal.

Chapter 3 analyses different aspects of probabilistic fuzzy systems in the context of function approximation and conditional density estimation. We analyse the necessary conditions for a PFS, such that the estimated output density is a proper probability density function and

subsequent higher moments derived from this density exist. These conditions relax the previous assumption of well-formed output space and are not very restrictive or consider a particular definition of conditional probability of fuzzy systems. We consider the relation of probabilistic fuzzy system with different types of deterministic systems that have universal approximation capability. This relation indicates that a PFS is also suitable for problems of function approximation. Furthermore we analyse a PFS as a fuzzy additive system and how a PFS can be obtained from this fuzzy additive system. A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred to PFS, while providing understanding of different aspects of a probabilistic fuzzy system. We analyse the parameters of PFS in synthetic data for function approximation and conditional density estimation. Finally we show that a PFS is suitable to estimate and predict quarterly US inflation. In this problem both the point estimates and the density estimates of inflation are relevant for a comprehensive analysis and for policy making. We show that PFS provides accurate density estimates for this data.

In chapter 4 we consider Value-at-Risk estimation by using probabilistic fuzzy systems. A PFS provides the potential to adapt estimations of probability density to the linguistic framework of the modeller. We study two approaches to designing probabilistic fuzzy VaR models and compare their performance with the performance of a GARCH model. In the first approach, a Mamdani-type probabilistic fuzzy system (Kaymak et al., 2003) is used for estimating the VaR. The model parameters are obtained by a sequential approach in which the location of the antecedent membership functions is determined by using fuzzy clustering and maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The output membership functions are scaled by using a single scaling parameter. In the second approach, an alternative representation of a PFS as a fuzzy histogram is considered. In this case, the membership functions are obtained from the modeler and the conditional probability parameters of the model are then estimated by minimising the test statistic of a back testing method by using a constrained evolutionary optimisation algorithm.

Conditional densities and Value at risk (VaR) values for financial returns have been successfully estimated using single covariate probabilistic fuzzy systems (PFS) as presented in chapter 4. Chapter 5 considers conditional density and VaR estimation based on a PFS model for density forecast of a continuous response variable conditional on a high-dimensional set of covariates. The proposed model is a multi-covariate multi-output PFS model which provides the conditional density forecasts of returns for one day ahead and one month ahead periods. Furthermore, this model allows to analyse seasonal patterns in returns. The additional information and process understanding provided by the different interpretations of the PFS model is illustrated and the model parameters are estimated by a novel two–step process. The proposed

Table 1.1: Chapters that contain some overlap. Bold face indicates detailed overview.

	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
PFS	**Sec. 2.3**	**Sec. 3.2**	Sec. 4.3	Sec. 5.2	
Multi-Output PFS		**Sec. 3.2.5**		Sec. 5.2.4	
VaR Models			Sec. 4.2	**Sec. 5.3**	
GARCH			Sec. 4.2.2	Sec. 5.3.2	**Sec. 6.2**

model is applied to daily S&P500 stock returns. Properties of the estimated conditional density for the S&P 500 index are reported and the performance of the proposed model is compared to the performance of a GARCH model for VaR estimation of the S&P 500 index. It is shown that the validity tests for GARCH models are sometimes rejected, while those of PFS models of VaR are never rejected. Additionally, the PFS model captures both instant volatility changes and periods of high volatility, and leads to less conservative models. It is found that the proposed model indicates seasonal patterns in short and longer horizons as well as conservative VaR in long term forecasts.

In chapter 6 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 modelled through a GARCH model while the fuzziness or linguistic vagueness is present in the antecedent and combination of the rule base system. The conditional distribution of the data can vary smoothly over time in mean and variance, where the smooth changes are related to linguistic descriptors, 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.

1.6.2 Thesis organization

This thesis is a collection of studies on models that encompass fuzzy and probabilistic representations of uncertainty capable of conditional density estimation. Each chapter contains sufficient information to be read independently. Since most of this thesis verses on probabilistic fuzzy systems, some degree of overlap exists between chapters. This overlap is clearly illustrated in Table 1.1.

This thesis is organized as follows. Chapter 2 presents a formal description of probabilistic fuzzy systems for conditional density approximation. In these systems the probabilistic nature and fuzzy uncertainty are taken into account. Probabilistic fuzzy systems are further analysed in chapter 3 in the context of function approximation and conditional density estimation. Furthemore this chapter discusses the application of PFS in the estimation and prediction of quarterly US inflation. Chapter 4 presents two approaches to the design of probabilistic fuzzy systems for Value-at-Risk estimation of multiple stocks. Chapter 5 builds upon the previous approaches and presents a conditional density and Value-at-Risk estimation based on a multi-output PFS model for density forecast of a continuous response variable conditional on a high-dimensional set of covariates. The additional information and process understanding provided by the different interpretations of the PFS model is illustrated, where special attention is given to the interpretation of the models in terms of stylized facts, such as seasonality and volatility clustering. Finally, chapter 6 proposes the combination of the well understood GARCH model with fuzziness or linguistic vagueness in a rule base model.

Alternatively, the reader can follow the chronological order of these studies. The main ideas of probabilistic fuzzy systems were discussed in different applications to real world problems, (e.g. van den Berg et al., 2002a, 2004; Xu and Kaymak, 2008) and (Almeida and Kaymak, 2009a,b) presented in chapter 4, but a formal description and analysis of this type of system was yet to be provided. This definition is provided in chapter 2. The successful application of PFS in conditional density estimation and associated left tail measures of risk, such as Value-at-risk, are presented in chapter 5. Special attention is given to the interpretation of the models in terms of stylized facts, such as seasonality and volatility clustering. Encouraged by this successful application, different aspects of PFS required further analysis. This analysis is presented in chapter 3. Since probabilistic fuzzy systems were only able to outperform GARCH models in some situations (Almeida and Kaymak, 2009a; Almeida et al., 2012a), it was interesting to investigate if this widely used, well understood and simple econometric model could be combined in a meaningful way with fuzzy systems. The resulting model is the fuzzy GARCH model presented in chapter 6.

Chapter 2

Conditional density estimation using probabilistic fuzzy systems[1]

2.1 Introduction

APPROXIMATION of unknown functions from sampled data is an important activity in modern modelling and systems theory. With the advent of modern computer systems, the costs of data collection and storage have been reduced significantly. However, it has become equally important to develop models from the data, which have sufficient generalization power and can describe the underlying process with accuracy despite the non-linearity and the complexity of these processes. The machine learning community has responded to this need by developing various methods such as neural networks (Bishop, 1995), support vector machines (Cristianini and Shawe-Taylor, 2000) and fuzzy systems (Klir and Yuan, 1995), which can be used for non-linear function approximation.

Amongst the systems that have universal approximation capability, fuzzy systems have attracted particular interest due to their ability to provide linguistic descriptions of the modelled process. Encouraged by their success in practical applications, fuzzy sets community has proposed various rule base structures and reasoning mechanisms for fuzzy systems (Mamdani and Gaines, 1981; Takagi and Sugeno, 1985, e.g.), putting the emphasis on the modelling of the linguistic uncertainty and the interpolation capability of fuzzy systems. Some researchers outside the fuzzy set community, however, have felt uneasy about the success of fuzzy systems for function approximation, partly because the connection of these systems to the probabilistic nature of uncertainty in many data sets was unclear (see e.g. the panel discussion by the representatives of three European Networks of Excellence on fields related to computational in-

[1] Parts of this chapter have been published in van den Berg, Kaymak, and Almeida (2012, 2013).

telligence in CoIL2000). Fuzzy systems have thus been seen as being heuristic systems without clear connections to probability theory.

Since fuzzy systems are known to be universal approximators (Kosko, 1994), it is reasonable to assume that they lend themselves for probabilistic analysis, just like other universal approximators known from the literature. The question that needs to be answered is whether fuzzy systems are able to estimate conditional probability density functions (pdf's), and in particular, whether they are able to estimate the conditional expected output values for a given system. If the answer is positive, this can explain the success of fuzzy systems for function approximation in the presence of probabilistic uncertainty.

Various researchers have studied the relation between probabilistic and fuzzy systems, and more generally, between probabilistic and fuzzy modelling (see e.g. Bertoluzza et al. (2002); Grzegorzewski et al. (2002); Thomas (1995) for a collection of papers on these topics). In his perception-based theory of probabilistic reasoning Zadeh (2002) introduces a set of inference schemes for answering all kinds of 'every day questions' where both numerical (measurement-based) and linguistic (perception-based) information are processed. Dubois and Prade have studied the relation between the possibility theory and the probability theory (Dubois and Prade, 2002). However, fuzzy systems for *function approximation* serve another goal than a perception-based analysis and they are also not rooted in the possibilistic interpretation of fuzzy sets.

Kosko has analysed the relation of such fuzzy systems to probabilistic systems (Kosko, 1997). He finds a connection between fuzzy systems and probabilistic systems, but his argument is mainly based on the mathematical similarity of center-of-gravity defuzzification (Klir and Yuan, 1995) to the computation of an expected value in probability theory: normalized membership functions are simply said to define a (discrete) probability density function (Kosko, 1997, pp. 53). Similarly, many researchers have argued that fuzziness and randomness are actually describing the same phenomena or at least they presume that fuzzy set theory is a generalization of probability theory or the other way around. For example, Thomas strongly advocates the proposition that a fuzzy subset is actually a likelihood function (Thomas, 1995), while Goodman and Nguyen extensively discuss the random set representation of membership functions based upon results of so-called α-level sets (Goodman and Nguyen, 2002).

However, fuzzy systems research has shown that the concept of membership and the concept of probability are different (Zadeh, 1968, 1995; Bertoluzza et al., 2002). In the last decade, studies where fuzzy rule-based systems also have probabilistic features that allows them to handle randomness, have received much interest. For example, in Meghdadi and Akbarzadeh-T. (2001); Liu and Li (2005, 2009); Zhang and Li (2010, 2012) probabilistic fuzzy sets are used instead of the regular fuzzy sets, where it is considered that the fuzzy membership grade is a

random variable with a certain probabilistic distribution function. This type of systems is similar to type-2 fuzzy systems (Karnik et al., 1999), where the primary membership function is fuzzy and the secondary function is a probabilistic density function. The combination of these two functions is able to express both fuzzy and stochastic information. This type of models was also combined with neural networks (Li and Liu, 2008) to improve time varying stochastic uncertainty. A similar approach was presented in Hengjie et al. (2011) where a probabilistic fuzzy neural network is introduced. The probabilistic information is incorporated in the antecedent part of fuzzy rules and its impact quantified on the consequent part. In Abonyi and Szeifert (2001); Lee et al. (2008) a fuzzy rule base classification model is obtained through an iterative learning process, where the consequent part of each rule is defined as the probabilities that a given rule represents. Thus, each rule can represent more than one class with different probabilities. Following the concept of random fuzzy variable (Colubi et al., 2002), fuzzy models are developed from the probabilistic and statistical point of view (Zmeškal, 2001). In Hong et al. (2009) a Takagi–Sugeno model combined with probabilistic noise explicitly, is presented. Special focus is placed on density estimation in Helin and Koivisto (2011a), using a GARCH model where the error distribution is obtained from fuzzy rules. The universal-function-approximation capability of fuzzy systems with consideration of probability distributions over possible consequences of an action, have also been used for reinforcement learning (Hinojosa et al., 2011).

In this chapter, we follow an idea similar to van den Eijkel (1999); Meghdadi and Akbarzadeh-T. (2001); Liu and Li (2005); Zhang and Li (2010, 2012) where fuzziness and randomness can co-occur, but following a different approach. The approach used in this chapter has previously been applied to real world problems, (van den Berg et al., 2002a, 2004; Almeida and Kaymak, 2009b; Xu and Kaymak, 2008, e.g.), but a formal description and analysis of this type of systems still needs to be given. In this work we consider the relation of fuzzy systems for conditional density estimation to the probabilistic uncertainty in the data within a framework of probabilistic fuzzy systems, which deal explicitly and simultaneously with two complementary types of uncertainty (fuzziness or linguistic uncertainty and probabilistic uncertainty) based on probability measures for fuzzy events. We show that probabilistic fuzzy systems, as defined in this chapter, estimate conditional pdf's for the output variable, given the inputs to the system. We provide an additive reasoning mechanism for this purpose. We derive expressions for computing the expected output of a probabilistic fuzzy system both in cases where we know the probability distribution in advance and in cases where we need to assess the relevant probabilistic quantities from the data. We further show that a zero-order Takagi–Sugeno (TS) deterministic fuzzy system uses the same expressions for reasoning. Hence, its parameters can be selected such that its output is equal to the conditional expected value of the identified probability density function.

Note that a deterministic function approximator of n variables can be used to generate a given distribution, provided it learns the appropriate mapping and its inputs are augmented to $n+1$ variables by the addition of a uniformly distributed variable r (Werbos, 2009). The application of this idea in fuzzy systems can be found in Kreinovich and Nguyen (2009); Kreinovich et al. (2010). Our approach is related, but different in that we do not consider an additional random input to the system. This new approach has the advantage that it can deal explicitly and simultaneously with fuzziness or linguistic uncertainty and probabilistic uncertainty. This model can estimate a probability density function of a non-linear system while keeping a linguistic link between variables. Besides the information provided by the linguistic interpretation of the rules, the probabilistic fuzzy model proposed allows to gain more information and process understanding given by the different reasoning mechanisms analyzed in this chapter. These advantages are illustrated in a financial application of conditional density estimation.

The outline of the chapter is as follows. In Section 2.2, we give an overview of the concept of probability of fuzzy events, which is at the basis of probabilistic fuzzy systems. In addition, we present some statistical theory of fuzzy events, most notably concerning the notion of fuzzy histogram. We introduce probabilistic fuzzy systems in Section 2.3 and we discuss how reasoning can be made with these systems. An additive reasoning mechanism is introduced. It is explained how conditional expected outputs of such systems can be computed within probabilistic and statistical approaches. In Section 2.4, the relation of probabilistic fuzzy systems to deterministic fuzzy systems is considered. It is shown that the output of both systems can be equivalent in certain cases. We discuss in Section 2.5 several issues related to our findings, and conclude the chapter in Section 2.6.

2.2 Probability and statistics of fuzzy events

Probabilistic fuzzy systems are based on the concept of the probability of a fuzzy event, as defined by Zadeh (1968). In the following subsection 2.2.1, we give a brief introduction to the theory of probability measures of fuzzy events. In the next subsection 2.2.2, we present several results concerning the statistics of fuzzy events that we will need later on.

2.2.1 Probability of fuzzy events

The material in this section assumes a random scalar variable $\underline{x}$ defined on a continuous sample space X. The results for discrete variables and vector variables are analogous.

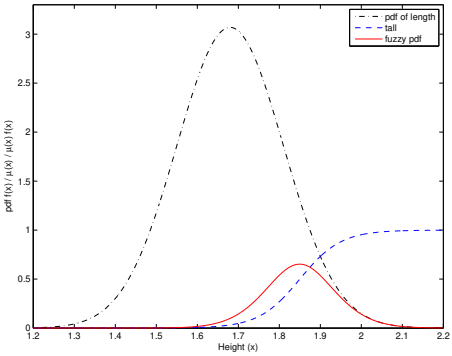


Figure 2.1: The pdf $f(x)$ of the height of Dutch women, the membership function $u(x)$ defining tallness, and the 'fuzzy pdf' $u(x)f(x)$.

A compact subset Γ of X defines an event, and its probability $\Pr(\Gamma)$ is found by integrating the probability density function (pdf) $f(x)$ as

$$\Pr(\Gamma) = \int_{x\in\Gamma} f(x)dx = \int_{-\infty}^{\infty} \chi_\Gamma(x) f(x)dx\,, \tag{2.1}$$

where $\chi_\Gamma(x)$ is the binary characteristic function for the event Γ such that $\chi_\Gamma(x) = 1 \Leftrightarrow x \in \Gamma$ and $\chi_\Gamma(x) = 0$ otherwise. In other words, the probability of an event is given by the expectation of its characteristic function.

By replacing the characteristic function in (2.1) with a membership function $u(x)\colon X \to [0,1]$, the probability measure for crisp events can be extended to a probability measure for fuzzy events. In this case, the probability of a fuzzy event A is found by taking the expectation of the membership function as (Zadeh, 1968)

$$\Pr(A) = \int_{-\infty}^{\infty} u_A(x) f(x)dx = \mathrm{E}(u_A(x))\,. \tag{2.2}$$

Equation (2.2) is illustrated in Fig. 2.1. The height $\underline{x}$ of the population of Dutch women is assumed to be a stochastic variable with a pdf, say $f(x)$, while the fuzzy notion of tallness is defined by a membership function, say $u(x)$. The product $u(x)f(x)$ can be termed a 'fuzzy pdf' which is used to calculate the probability that a Dutch woman is tall according to (2.2). Note that this calculation takes both the probabilistic uncertainty and the fuzzy uncertainty of the notion of tallness into account.

Below we shall consider sample spaces that are fuzzily partitioned in a finite set of fuzzy sets. The reason for this is expressed by the following theorem (Ruspini, 1969; van den Berg et al., 2001):

Theorem 2.1 *Let fuzzy events $A_1, A_2, .., A_J$ form a proper fuzzy partition (Klir and Yuan, 1995) in sample space X implying that*

$$\forall x : \sum_{j=1}^{J} u_{A_j}(x) = 1 \,. \tag{2.3}$$

Then, the sum of the probabilities of the fuzzy events equals one or, in mathematical terms,

$$\sum_{j=1}^{J} \Pr(A_j) = 1 \,. \tag{2.4}$$

Fuzzily partitioned sample spaces having property (2.4) will be termed 'well-defined'.

In Section 2.3, we will also need to deal with conditional fuzzy probabilities, i.e., the probability of a fuzzy event given the occurrence of another fuzzy event. The underlying definition used is the following one

$$\begin{aligned} \Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} &= \frac{\int_{-\infty}^{\infty} u_{A \cap B}(x) f(x) dx}{\int_{-\infty}^{\infty} u_B(x) f(x) dx} \\ &= \frac{\int_{-\infty}^{\infty} u_A(x) u_B(x) f(x) dx}{\int_{-\infty}^{\infty} u_B(x) f(x) dx} \,, \end{aligned} \tag{2.5}$$

where the intersection of two fuzzy events is modelled by the product t-norm (Klir and Yuan, 1995). It is easy to prove (van den Berg et al., 2002a) that definition (2.5) guarantees that theorem 2.1 also holds for conditional probabilities, i.e,

$$\sum_{j=1}^{J} \Pr(A_j|B) = 1 \,. \tag{2.6}$$

2.2.2 Statistical issues

The result described by (2.2) allows us to assess the probability of a fuzzy event from sampled data by using standard expectation estimators such as the arithmetic mean (Kruse, 1982; van den Eijkel, 1999; van den Berg et al., 2001). According to this approach, the probability for

fuzzy event A can be estimated using

$$\hat{\Pr}(A) = \frac{1}{P}\sum_{p=1}^{P} u_A(x_p)\,, \tag{2.7}$$

when P samples x_p are available. The following theorem shows that the estimate $\hat{\Pr}(A)$ has the properties described in theorem 2.1.

Theorem 2.2 *Let fuzzy events $A_1, A_2, \ldots, A_J$ form a proper fuzzy partition in sample space X. Then, the sum of the estimated probabilities of the fuzzy events* (2.7) *equals one or, in mathematical terms,*

$$\sum_{j=1}^{J} \hat{\Pr}(A_j) = 1. \tag{2.8}$$

Proof: Using the sample space property of being well-defined, i.e. (2.3) holds, we conclude that

$$\begin{aligned}\sum_{j=1}^{J} \hat{\Pr}(A_j) &= \sum_{j=1}^{J} \frac{1}{P}\sum_{p=1}^{P} u_{A_j}(x_p) = \frac{1}{P}\sum_{p=1}^{P}\sum_{j=1}^{J} u_{A_j}(x_p)\\ &= \frac{1}{P}\sum_{p=1}^{P} 1 = \frac{1}{P}P = 1.\end{aligned}$$

■

Conditional probabilities for a fuzzy event A, given another fuzzy event B, can be estimated in a similar way. Inspired by (2.5), such a conditional probability $\Pr(A|B)$ is found by (van den Eijkel, 1999; van den Berg et al., 2001)

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}\,,$$

and can be estimated as

$$\hat{\Pr}(A|B) = \frac{\sum_{p=1}^{P} u_A(x_p)u_B(x_p)}{\sum_{p=1}^{P} u_B(x_p)}\,. \tag{2.9}$$

In classical probability theory, we can approximate a probability density function with a finite support by scaling the characteristic functions of crisp events for a disjoint cover of the support. Such an approximation is called a histogram. Assuming we partition the support into disjoint sets Γ_j, $j = 1, \ldots, J$, the probability density function $f(x)$ is approximated by $\hat{f}(x)$

$$\hat{f}(x) = \sum_{j=1}^{J} \Pi_j = \sum_{j=1}^{J} \frac{\hat{\Pr}(\Gamma_j)\chi_{\Gamma_j}(x)}{\int_{-\infty}^{\infty} \chi_{\Gamma_j}(x)dx}\,,$$

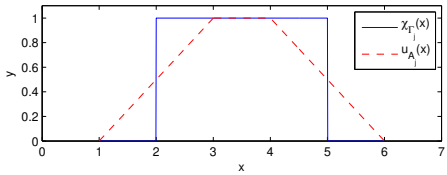


Figure 2.2: A crisp interval and a fuzzy interval of the same size since $\int_{-\infty}^{\infty} \chi_{\Gamma_j}(x)dx = \int_{-\infty}^{\infty} u_{A_j}(x)dx = 3$.

where Π_j represents the jth column of the histogram and where the normalization factor $\int_{-\infty}^{\infty} \chi_{\Gamma_j}(x)dx$ equals the size (in the one-dimensional case, the length) of the set (interval) Γ_j. Similarly, one can approximate the probability density function by scaling the membership functions of fuzzy events that form a proper fuzzy partition of the support as (van den Berg et al., 2001)

$$\hat{f}(x) = \sum_{j=1}^{J} \Lambda_j = \sum_{j=1}^{J} \frac{\hat{\Pr}(A_j) u_{A_j}(x)}{\int_{-\infty}^{\infty} u_{A_j}(x)dx}, \tag{2.10}$$

where each

$$\Lambda_j = \frac{\hat{\Pr}(A_j) u_{A_j}(x)}{\int_{-\infty}^{\infty} u_{A_j}(x)dx} \tag{2.11}$$

represents a 'fuzzified column'. Note that in (2.10) and (2.11), the normalization factor

$$\int_{-\infty}^{\infty} u_{A_j}(x)dx \tag{2.12}$$

of the jth fuzzified column equals the the 'fuzzy length' of the set A_j. We illustrate this approach in Fig. 2.2 showing both a crisp and a fuzzy interval of equal size indicated by equal area under the respective membership functions.

We further make the important observation that (2.10) can also be considered as a weighted additive fuzzy reasoning scheme where the fuzzy membership functions $u_{A_j}(x)$, $j = 1, 2, \ldots, J$ are combined to one fuzzy membership function $u_A(x)$ using the factors $\hat{\Pr}(A_j)/\int_{-\infty}^{\infty} u_{A_j}(x)dx$ as weights:

$$u_A(x) = \sum_{j=1}^{J} \frac{\hat{\Pr}(A_j)}{\int_{-\infty}^{\infty} u_{A_j}(x)dx} u_{A_j}(x)\,. \tag{2.13}$$

Like in the fuzzy histogram interpretation (2.11), we use the normalization factors (2.12) also here, since we want to compensate for different sizes $\int_{-\infty}^{\infty} u_{A_j}(x)dx$.

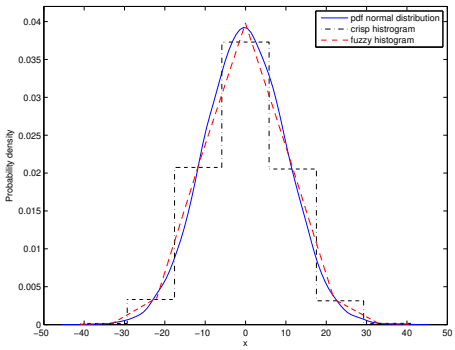


Figure 2.3: A fuzzy histogram better approximates a pdf than a crisp histogram.

Theorem 2.3 *Let X be a well-defined sample space partitioned into J fuzzy sets A_j, $j = 1, \ldots, J$. Then the approximated density function $\hat{f}(x)$ has the (desired) property*

$$\int_{-\infty}^{\infty} \hat{f}(x)dx = 1\,.$$

Proof: Note that for a well-defined sample space, (2.8) holds. Then, by also using (2.10), we conclude that

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{f}(x) &= \int_{-\infty}^{\infty} \sum_{j=1}^{J} \frac{\hat{\Pr}(A_j) u_{A_j}(x)}{\int_{-\infty}^{\infty} u_{A_j}(x)dx} dx \\ &= \sum_{j=1}^{J} \hat{\Pr}(A_j) \frac{\int_{-\infty}^{\infty} u_{A_j}(x)dx}{\int_{-\infty}^{\infty} u_{A_j}(x)dx} = 1\,.\end{aligned}$$

■

Because of overlapping membership functions, fuzzy histograms have a high level of statistical efficiency, better than crisp ones. We show this in Fig. 2.3 where the probability density function (pdf) of the standard normal distribution is approximated by a classical and by a fuzzy histogram using in both cases a partitioning in seven classes. For more details we refer to van den Berg et al. (2004). Besides a high level of statistical efficiency, several classes of fuzzy histograms also have a high level of computational efficiency. An example of such type of fuzzy histogram is one that uses triangular membership functions (Waltman et al., 2005a).

2.3 Probabilistic fuzzy systems

2.3.1 Outline

Probabilistic fuzzy systems combine two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. In previous works, we have presented various types of probabilistic fuzzy systems with the corresponding reasoning schemes (van den Berg et al., 2002a, 2004; Kaymak et al., 2003; van den Berg et al., 2002b). In this chapter, we present a more general formulation where the consequent of each rule is a conditional pdf, given the fuzzy antecedent of the rule. Our probabilistic fuzzy system consists of the rules R_q, $q = 1, \ldots, Q$, of the type

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } f(y|A_q)\,, \tag{2.14}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an input vector, $A_q : X \longrightarrow [0, 1]$ is a fuzzy set defined on X and $f(y|A_q)$ is the conditional pdf of the stochastic output variable $\underline{y}$ given the fuzzy event A_q. The interpretation is as follows: if fuzzy antecedent A_q is fully valid ($x \in \text{core}(A_q)$), then y is a sample value from the probability distribution with conditional pdf $f(y|A_q)$.

If A_q had been crisp events, then only one of the rules would fire and hence only one of the conditional pdf's would be used. The system output can then be written as

$$f(y|\mathbf{x}) = \sum_{q=1}^{Q} \chi_q(\mathbf{x})\, f(y|A_q)\,.$$

In case of fuzzy events, multiple rules may fire and it is more appropriate to take an additive combination of rule outputs.We propose a reasoning mechanism that determines the output of fuzzy system as

$$f(y|\mathbf{x}) = \frac{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x}) f(y|A_q)}{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x})} = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) f(y|A_q)\,, \tag{2.15}$$

where $\beta_q(\mathbf{x}) = u_{A_q}(\mathbf{x}) / \sum_{q=1}^{Q} u_{A_q}(\mathbf{x})$ represents the normalized degree of fulfillment of rule R_q or, in other words,

$$\sum_{q=1}^{Q} \beta_q(\mathbf{x}) = 1\,. \tag{2.16}$$

When $\mathbf{x}$ is n-dimensional, u_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.,*

$$u_{A_q}(\mathbf{x}) = u_{A_{q1}}(x_1) \circ \cdots \circ u_{A_{qn}}(x_n),$$

where x_n is the n-th component of $\mathbf{x}$ and $\circ$ denotes a t-norm. The following theorem shows that the reasoning (2.15) returns a proper pdf.

Theorem 2.4 *Let $R = \cup_{q=1}^{Q} R_q$ be a fuzzy rule base consisting of the rules of type* (2.14)*. Then, the reasoning scheme* (2.15) *computes a pdf, i.e.*

$$\int_{-\infty}^{\infty} f(y|\mathbf{x})dy = 1\,.$$

Proof: Taking the integral over the left-hand side of equation (2.15), we immediately derive the result:

$$\begin{aligned}\int_{-\infty}^{\infty} f(y|\mathbf{x})dy &= \int_{-\infty}^{\infty} \frac{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x}) f(y|A_q)}{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x})} \\ &= \frac{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x}) \int_{-\infty}^{\infty} f(y|A_q)dy}{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x})} = 1.\end{aligned}$$

■

Therefore, if we know the pdf for each rule output, we can calculate the conditional pdf for any input vector $\mathbf{x}$. This formulation is akin to a mixture model, whereby the weights of the mixture are determined by the membership value to the rule antecedents.

For the purpose of function approximation, it is possible to calculate a crisp output for each input vector $\mathbf{x}$ from a conditional probability distribution. To do so, we take a regression approach. The regression hyperplane of y on X is defined (Kecman, 2001) as the location of the mathematical expectations $\mathrm{E}(\underline{y}|\mathbf{x})$ conform

$$\eta_{\underline{y}|\mathbf{x}} = \mathrm{E}(\underline{y}|\mathbf{x}) = \int_{-\infty}^{\infty} y f(y|\mathbf{x})dy\,. \tag{2.17}$$

An interesting characteristic of probabilistic fuzzy system is that besides calculating the expected output, it is also possible to estimate the mode, conditional variance and quantiles, all based on the obtained output probability distribution function. The conditional variance $\sigma^2_{\underline{y}|\mathbf{x}}$ of the output can be calculated conform

$$\sigma^2_{\underline{y}|\mathbf{x}} = \mathrm{Var}(\underline{y}|\mathbf{x}) = \mathrm{E}(\underline{y}^2|\mathbf{x}) - (\mathrm{E}(\underline{y}|\mathbf{x}))^2\,. \tag{2.18}$$

The expected conditional output and conditional variance of the probabilistic fuzzy system is given by the following theorem.

Theorem 2.5 *The expected output of the probabilistic fuzzy system with rule base* (2.14) *is given by the weighted average of the expected output of each rule, i.e.,*

$$\eta_{\underline{y}|\mathbf{x}} = \mathrm{E}(\underline{y}|\mathbf{x}) = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\mathrm{E}(\underline{y}|A_q)\,,$$

and its conditional variance is

$$\sigma^2_{\underline{y}|\mathbf{x}} = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\mathrm{E}(\underline{y}^2|A_q) - \eta^2_{\underline{y}|\mathbf{x}}\,,$$

Proof: Using (2.17), (2.15) and

$$\mathrm{E}(\underline{y}|A_q) = \int_{-\infty}^{\infty} y f(y|A_q)dy\,, \tag{2.19}$$

we conclude

$$\begin{aligned}\mathrm{E}(\underline{y}|\mathbf{x}) &= \int_{-\infty}^{\infty} y \left[\sum_{q=1}^{Q} \beta_q(\mathbf{x}) f(y|A_q)\right] dy = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \int_{-\infty}^{\infty} y f(y|A_q)dy \\ &= \sum_{q=1}^{Q} \beta_q(\mathbf{x})\mathrm{E}(\underline{y}|A_q)\,.\end{aligned}$$

Similarly, using (2.18), (2.17), (2.15) and (2.19)

$$\begin{aligned}\sigma^2_{\underline{y}|\mathbf{x}} &= \int_{-\infty}^{\infty} y^2 \left[\sum_{q=1}^{Q} \beta_q(\mathbf{x}) f(y|A_q)\right] dy - (\mathrm{E}(\underline{y}|\mathbf{x}))^2 = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \int_{-\infty}^{\infty} y^2 f(y|A_q)dy - \eta^2_{\underline{y}|\mathbf{x}} \\ &= \sum_{q=1}^{Q} \beta_q(\mathbf{x})\mathrm{E}(y^2|A_q) - \eta^2_{\underline{y}|\mathbf{x}}\,.\end{aligned}$$

■

2.3.2 Reasoning

In general, the pdf's in the rule consequents are not available, and they must be estimated from the data. We present two equivalent elaborations. In both cases, we suppose that J fuzzy classes C_j form a fuzzy partition of the compact output space Y.

The fuzzy histogram approach

In the first approach, we replace in each rule of (2.14) the true pdf $f(y|A_q)$ by its fuzzy approximation (fuzzy histogram) $\hat{f}(y|A_q)$ yielding the rule set $\hat{R}_q$, $q = 1, \ldots, Q$ defined as

$$\hat{R}_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } \hat{f}(y|A_q)\,, \tag{2.20}$$

where $\hat{f}(y|A_q)$ is defined in line with equation (2.10) conform

$$\hat{f}(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q)u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,. \tag{2.21}$$

In effect, we are using a histogram based on fuzzy events, instead of a usual histogram, to represent the pdf in the rule consequent. A diagram depicting the reasoning of this approach is shown in Fig. 2.4. For any given x_1 we compute estimate $\hat{f}(y|x_1)$ of the conditional probability density function based on a fuzzy histogram $\hat{f}(y|A_q)$. In the figure, only one rule fires for the selected x_1. The crisp system output $\hat{\eta}_{\underline{y}|\mathbf{x}}$ is computed for all x, as the expectation of the estimated conditional probability density function, as it will be presented in Theorem 2.6.

Using the same line of thought as used in subsection 2.3.1, we can calculate an approximation of the expected conditional output of the probabilistic fuzzy output. The corresponding theorem, is the following one.

Theorem 2.6 *The estimated expected output of the probabilistic fuzzy system with rule base* (2.20) *is given by the weighted average of the estimated expected output of each rule according to*

$$\hat{\eta}_{\underline{y}|\mathbf{x}} = \hat{\mathrm{E}}(\underline{y}|\mathbf{x}) = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y|A_q) = \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\Pr}(C_j|A_q)z_{1,j}\,, \tag{2.22}$$

and the estimated conditional variance is

$$\begin{aligned}\hat{\sigma}^2_{\underline{y}|\mathbf{x}} &= \hat{\mathrm{E}}(\underline{y}^2|\mathbf{x}) - (\hat{\mathrm{E}}(\underline{y}|\mathbf{x}))^2 = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) - \hat{\eta}^2_{\underline{y}|\mathbf{x}} \\ &= \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\Pr}(C_j|A_q)z_{2,j} - \hat{\eta}^2_{\underline{y}|\mathbf{x}}\,,\end{aligned} \tag{2.23}$$

where $\hat{\mathrm{E}}(\underline{y}|A_q)$ *is the estimated expected output of each rule,* $(\hat{\mathrm{E}}(\underline{y}|A_q))^2$ *is the estimated variance of the output of each rule,* $z_{1,j}$ *is the centroid of the* j*th output fuzzy set defined by*

$$z_{1,j} = \frac{\int_{-\infty}^{\infty} y u_{C_j}(y) dy}{\int_{-\infty}^{\infty} u_{C_j}(y) dy} \,. \tag{2.24}$$

and $z_{2,j}$ *is defined as*

$$z_{2,j} = \frac{\int_{-\infty}^{\infty} y^2 u_{C_j}(y) dy}{\int_{-\infty}^{\infty} u_{C_j}(y) dy} \,. \tag{2.25}$$

Proof: Using (2.17) with $f(y|A_q)$ replaced by the estimated $\hat{f}(y|A_q)$, and using (2.15) and (2.21), we derive that

$$\begin{aligned}
\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) &= \int_{-\infty}^{\infty} y \hat{f}(y|\mathbf{x}) dy = \int_{-\infty}^{\infty} y \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{f}(y|A_q) dy \\
&= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \int_{-\infty}^{\infty} y \sum_{j=1}^{J} \frac{\hat{\mathrm{Pr}}(C_j|A_q) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy} dy \\
&= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) \frac{\int_{-\infty}^{\infty} y u_{C_j}(y) dy}{\int_{-\infty}^{\infty} u_{C_j}(y) dy} \\
&= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j} \,,
\end{aligned} \tag{2.26}$$

where $z_{1,j}$ is the centroid of the fuzzy set C_j. The estimated expected conditional output $\hat{\mathrm{E}}(\underline{y}|A_q)$ of each rule $\hat{R}_q$ is defined as

$$\hat{\mathrm{E}}(\underline{y}|A_q) = \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j} \tag{2.27}$$

By substituting (2.27) in (2.26), we immediately find equation (2.22).

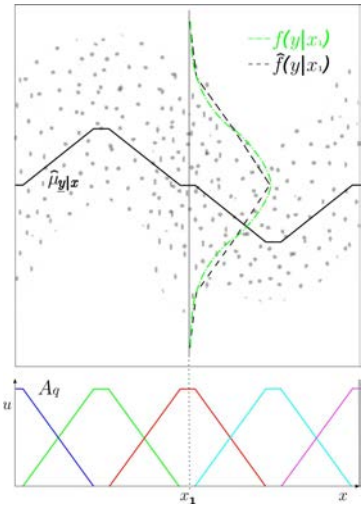


Figure 2.4: Diagram of the fuzzy histogram approach for PFS. The output of the model is a fuzzy histogram $\hat{f}(y|A_q)$ from which the crisp system output $\hat{\eta}_{\underline{y}|\mathbf{x}}$ is computed.

In the same manner, using (2.18) with $f(y|A_q)$ replaced by the estimated $\hat{f}(y|A_q)$, and using (2.15) and (2.21), we derive that

$$\begin{aligned}
\hat{\sigma}^2_{\underline{y}|\mathbf{x}} &= \int_{-\infty}^{\infty} y^2 \hat{f}(y|\mathbf{x})dy - (\hat{\mathrm{E}}(\underline{y}|A_q))^2 \\
&= \int_{-\infty}^{\infty} y \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{f}(y|A_q)dy - (\hat{\mathrm{E}}(\underline{y}|A_q))^2 \\
&= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \int_{-\infty}^{\infty} y^2 \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q)u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} dy - \hat{\eta}^2_{\underline{y}|\mathbf{x}} \\
&= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \hat{\Pr}(C_j|A_q) \frac{\int_{-\infty}^{\infty} y^2 u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} - \hat{\eta}^2_{\underline{y}|\mathbf{x}} \\
&= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) \hat{\Pr}(C_j|A_q) z_{2,j} - \hat{\eta}^2_{\underline{y}|\mathbf{x}},
\end{aligned}$$

where $z_{2,j}$ is defined by (2.25). ■

For modelling purposes, the parameters $\hat{\Pr}(C_j|A_q)$ and $z_{1,j}$ can be computed once offline. The evaluation of the expected output then requires the evaluation of $\beta_q(\mathbf{x})$ for a given $\mathbf{x}$ and the evaluation of (2.22), which can be very fast.

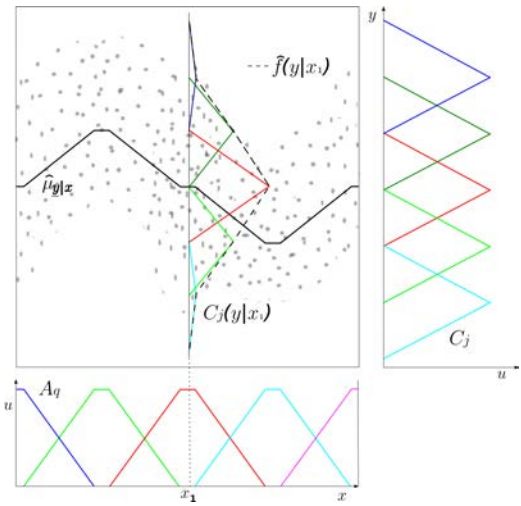


Figure 2.5: Diagram of the probability fuzzy output approach for PFS. Given the occurrence of fuzzy antecedent A_q, the fuzzy output events C_j are weighted with the conditional probability $\hat{\Pr}(C_j|A_q)$.

Note further that the proof of theorem 2.6 involves both an averaging step to deal with the probabilistic uncertainty as present in the pdf and a defuzzification step to handle the fuzzy uncertainty as present in the membership functions used. These two separate steps are needed to let the output of the fuzzy system be a crisp value.

The probabilistic fuzzy output approach

In the second approach, we decompose each rule (2.14) to provide a stochastic mapping between its fuzzy antecedents and its fuzzy consequents. The rules are written in the following form.

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } \underline{y} \text{ is } C_1 \text{ with } \hat{\Pr}(C_1|A_q) \text{ and} \\ \dots \\ \underline{y} \text{ is } C_J \text{ with } \hat{\Pr}(C_J|A_q). \end{aligned} \tag{2.28}$$

The interpretation is depicted in Fig. 2.5 and can be summarized as follows. If x_1 belongs to the fuzzy antecedent A_q, the fuzzy output event C_j occurs with an associated probability $\hat{\Pr}(C_j|A_q)$. For each individual rule, the expected output of each fuzzy rule $u_C(y|A_q)$ is calculated by scaling the fuzzy output C_j and then aggregated them into $u_C(y|\mathbf{x})$. For x_1 the scaled output sets $C_j(y|x_1)$, are depicted in Fig. 2.5. The crisp output $\hat{\eta}_{\underline{y}|\mathbf{x}}$ is obtained by defuzzifying the obtained

expected conditional fuzzy output $u_C(y|\mathbf{x})$. All the calculations are presented in Theorem 2.7. The advantage of using the rule base (2.28) instead of (2.20) is its transparency: the output of each rule is formulated in linguistic terms (namely $C_1, C_2, \ldots,$ and C_J) instead of probability density functions. The link to the linguistic knowledge of experts is then clearer.

Although the fuzzy rule bases (2.20) and (2.28) are different, we can prove the following theorem expressing that, under certain conditions, the two corresponding probabilistic fuzzy systems implement the same crisp input-output mapping.

Theorem 2.7 *Consider the probabilistic fuzzy system with rule base* (2.28) *and let the fuzzy additive reasoning scheme* (2.13) *be used to calculate its expected fuzzy output. Then, the expected output of the probabilistic fuzzy system with rule base* (2.20) *equals the defuzzified output of the probabilistic fuzzy system with rule base* (2.28).

Proof: Consider the system with the probabilistic fuzzy rule base (2.28). We first calculate the conditional expected fuzzy output $u_C(y|A_q)$ of each individual rule, i.e., the expected fuzzy membership function given the occurrence of A_q. By applying (2.13), we can write in this conditional case

$$u_C(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\mathrm{Pr}}(C_j|A_q)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} u_{C_j}(y)\,. \tag{2.29}$$

Using additive fuzzy reasoning (2.15) and substituting (2.29), we find the expected fuzzy membership function given the occurrence of $\mathbf{x}$, i.e.,

$$u_C(y|\mathbf{x}) = \frac{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x}) u_C(y|A_q)}{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x})} = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \frac{\hat{\mathrm{Pr}}(C_j|A_q)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} u_{C_j}(y)\,. \tag{2.30}$$

From this we first conclude, using (2.6), (2.8) and (2.16), that

$$\begin{aligned}\int_{-\infty}^{\infty} u_C(y|\mathbf{x})dy &= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \frac{\hat{\mathrm{Pr}}(C_j|A_q) \int_{-\infty}^{\infty} u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} \\ &= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) = 1\,. \end{aligned} \tag{2.31}$$

Having done all these preparations, we can now calculate the crisp output $\hat{\mathrm{E}}(\underline{y}|\mathbf{x})$ for each $\mathbf{x}$ by defuzzifying $u_C(y|\mathbf{x})$ as given by (2.30) while using the last result (2.31) and definition (2.24):

$$\begin{aligned}\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) &= \frac{\int_{-\infty}^{\infty} y u_C(y|\mathbf{x})dy}{\int_{-\infty}^{\infty} u_C(y|\mathbf{x})dy} = \int_{-\infty}^{\infty} y u_C(y|\mathbf{x})dy \\ &= \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q) \int_{-\infty}^{\infty} u_{C_j}(y)ydy}{\int_{-\infty}^{\infty} u_C(y|\mathbf{x})dy} \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) \hat{\Pr}(C_j|A_q) z_{1,j} \,. \end{aligned} \tag{2.32}$$

Comparing (2.22) to (2.32) shows that both expressions are equal. ■

The proofs of theorems 2.6 and 2.7 show a lot of similarities. However, looking carefully, we observe differences in the interpretation. In the proof of Theorem 2.6, we compute first an estimate $\hat{f}(y|\mathbf{x})$ of the conditional probability density function $f(y|\mathbf{x})$. This estimate is based on a fuzzy histogram. Then, the crisp system output is computed as the expectation of the estimated conditional probability density function. In the proof of Theorem 2.7, however, the crisp system output is computed by defuzzifying the expected conditional fuzzy output $u_C(y|\mathbf{x})$. The expected conditional fuzzy output is computed by first calculating the expected output of each fuzzy rule $u_C(y|A_q)$ and then aggregating them into $u_C(y|\mathbf{x})$. Note that the same type of fuzzy additive reasoning is applied in both schemes which eventually yields the same crisp input-output mapping.

Note further that (2.30) in the proof of Theorem 2.7 shows that - within the probabilistic fuzzy output interpretation - probabilistic fuzzy systems are an example of the additive fuzzy systems discussed in Kosko (1994). Using the notation of this chapter, additive fuzzy reasoning according to equation (1) of Kosko (1994) can be written as

$$u_C(y|\mathbf{x}) = \sum_{q=1} w_j u_C(y|A_q). \tag{2.33}$$

If we choose for the weights w_j in (2.33)

$$w_j = \frac{u_{A_q}(\mathbf{x})}{\sum_{q=1}^{Q} u_{A_q}(\mathbf{x})},$$

then we obtain (2.30) of this chapter, which makes clear that the probabilistic fuzzy systems as presented in this chapter have universal approximation capabilities according the theory as provided in Kosko (1994).

We finally note here that re-arranging (2.22) or (2.32) results into

$$\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) = \sum_{j=1}^{J} z_{1,j} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{\mathrm{Pr}}(C_j|A_q) = \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|\mathbf{x}) z_{1,j} \,,$$

where again fuzzy additive reasoning in line with definition (2.15) has been applied. The latter result shows that the expected system output is equal to the conditional expectation of the defuzzified fuzzy sets.

2.4 Relation to deterministic fuzzy systems

In this section, we consider the relation of the probabilistic fuzzy system described in Section 2.3 to deterministic fuzzy systems. In particular, we are interested in the relation between the expected output of a probabilistic fuzzy system and the deterministic output of a zero-order Takagi–Sugeno system (Takagi and Sugeno, 1985).

Theorem 2.8 *A zero-order Takagi–Sugeno fuzzy system with Q rules, antecedent fuzzy sets A_q and consequent parameters b_q computes the expected value of the conditional pdf provided that the parameters b_q are equal to the expected defuzzified output of the probabilistic fuzzy system, i.e. provided that*

$$b_q = \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j} \,. \tag{2.34}$$

Proof: The proof is provided by re-arranging (2.22) and comparing it to the output of a zero-order Takagi–Sugeno system. The output of a zero-order deterministic Takagi–Sugeno system is given by

$$\gamma(\mathbf{x}) = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) b_q \,.$$

Re-arranging (2.22) gives

$$\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j} = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) b_q \,,$$

with

$$b_q = \sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j} \,.$$

■

Therefore, by selecting the consequent parameters of the TS model in a specific way, one can approximate the expected output of the underlying system that has generated the data. Note that in many cases the parameters of TS fuzzy systems are optimized to minimize an error function, and hence optimality can be achieved in practical situations. This can explain the success of TS fuzzy systems for function approximation.

2.5 Discussion

The previous sections have shown that probabilistic fuzzy systems with an additive fuzzy reasoning scheme are able to approximate the conditional output pdf's for function approximation. This same input-output mapping is found by defuzzification of the expected fuzzy output of a probabilistic fuzzy system having a rule base with probabilistic fuzzy consequents.

We further found that the expected output of the probabilistic fuzzy systems discussed is equal to the output of deterministic zero-order TS fuzzy systems, provided that the consequent parameters are selected according to (2.34). This property provides motivation for the success of additive fuzzy systems for function approximation. Note that in addition to the probabilistic nature of the data, probabilistic fuzzy systems let the analyst explicitly model linguistic concepts through the use of antecedent fuzzy sets A_q and the consequent fuzzy sets C_j: see the rule base (2.28). This allows the model to estimate the underlying probabilistic structure from the data, while the model is calibrated to the linguistic description of the user. The other way around, is also possible to design the fuzzy system in an expert-driven manner. In that case, the calibration can be data-driven and be based on the estimation of the statistical quantities.

In addition to regular pdf's and conditional pdf's, probabilistic fuzzy models allow one to answer questions such as "what is the probability that the output is large given that the input is small" ($\hat{\Pr}(C_j|A_q)$) or "what is the probability that the output is medium given a particular input" ($\hat{\Pr}(C_j|\mathbf{x})$). Analyzing answers to these questions can provide additional information in a particular problem (van den Berg et al., 2004, e.g.). Another advantage of probabilistic fuzzy systems over conventional fuzzy systems is that besides estimating a crisp output, it is also possible to estimate probabilistic confidence bounds.

Although we have discussed that the probabilistic fuzzy systems can approximate conditional pdf's, we have not analyzed the accuracy of this approximation. In general, the accuracy of the approximation of the conditional pdf's can be increased by increasing the number of consequent fuzzy sets C_j on the output domain, by choosing a better fuzzy partitioning of the input or output space, or by selecting better-shaped membership functions. The latter selection problem resembles that of finding adequate basis functions when applying radial basis functions networks (Bishop, 1995) for kernel regression. We already mentioned that using a fuzzy

partition already improves the approximation of the conditional pdf significantly (van den Berg et al., 2004). Similarly, increasing the number of rules will improve the accuracy of interpolation between the rules. On the other hand, the danger that the resulting system overfits the (normally noisy) data (Bishop, 1995) should be dealt with as well.

A related issue that we have not discussed in this chapter is that of optimal design. Although the probabilistic fuzzy system approximates conditional pdf's, the resulting fuzzy system need not be optimal in terms of the number of rules, the definition of antecedent membership functions and consequent membership functions. Particular choices can provide better interpolation for different data sets. For example, in Almeida and Kaymak (2009a); Xu and Kaymak (2008) the influence of the location of output membership functions was investigated. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying with more membership functions placed towards the origin (Xu and Kaymak, 2008) or towards the edges of the universe of discourse (Almeida and Kaymak, 2009a). This varying placement allows to better capture the variability in regions with more membership functions. The design of a PFS is an issue that needs to be studied closely in the future. Furthermore, we have ignored *a priori* distribution of the data in this chapter. This information can be incorporated in probabilistic fuzzy systems through rule weighting, as discussed, for instance in van den Berg et al. (2002a).

In conjunction with defining the number of rules, antecedent and consequent membership functions, it is also necessary to estimate the conditional probabilities in a probabilistic fuzzy system. The calculation of conditional probabilities using (2.9) does not maximize the likelihood of the data set and may lead to biased results (Waltman et al., 2005b). Assuming that the samples in the data set are independent of one another and that the membership functions in the rule antecedent A_q and the rule consequent C_j have been defined, the probability parameters $\hat{\Pr}(C_j|A_q)$ that maximize the likelihood of the data set can be obtained by maximizing the function

$$L = \sum_{p=1}^{P} \ln\left(\Pr(y_p|\mathbf{x}_p)\right) , \tag{2.35}$$

where P is the number of samples in the data set (Waltman et al., 2005b). A suitable initialisation for iterative optimisation for maximum likelihood estimation is given by direct estimation from the data by using (2.9). Other search heuristics can be used to estimate the probability parameters $\hat{\Pr}(C_j|A_q)$, such as genetic algorithms (Almeida and Kaymak, 2009b). In this work, although the results were satisfactory, the authors noted that the objective function chosen was a problem with multiple minima. Thus the solution could converge to local optimums.

In this chapter, we have concentrated on the results for the expected output of probabilistic fuzzy systems and their equivalence to deterministic fuzzy systems. However, it is also impor-

tant to consider the higher moments in the estimations, since these will be influenced by the choice of the membership functions and other parameters. In addition, it is interesting to look at possibilities to develop statistical inference procedures for fuzzy quantities like fuzzy events. Finally, the precise relation of the probabilistic-fuzzy framework proposed here to that of radial basis function networks and that of kernel estimation require a deeper study. We leave this important work for future research.

2.6 Conclusions

Probabilistic fuzzy systems are able to approximate conditional pdf's, while at the same time calibrating the model to the linguistic conceptualization of the model maker. As such, they deal explicitly with both the fuzziness in the linguistic descriptions and the probabilistic uncertainty. We have proposed an additive reasoning scheme for probabilistic fuzzy systems. The expected output of these fuzzy systems is shown to be computable where both a defuzzification and an averaging step are needed to get rid of both uncertainties and to terminate in a crisp output. The complete reasoning is based on the possibility to calculate (a) the probability of a consequent fuzzy event given an antecedent fuzzy event, (b) the centroid points of the consequent fuzzy sets, and (c) the degree of fulfillment of the fuzzy rules. A zero-order TS fuzzy system can produce the same output as the expected output of a probabilistic fuzzy system, provided that its consequent parameters are selected as the conditional expectation of the defuzzified output membership functions. Our results provide insight why additive deterministic fuzzy systems such as TS systems have proven to be so successful for function approximation purposes.

Chapter 3

Analysing probabilistic fuzzy systems[1]

3.1 Introduction

PROBABILISTIC fuzzy systems (PFS) can deal explicitly and simultaneously with fuzziness or linguistic uncertainty and probabilistic uncertainty. A probabilistic fuzzy system follows an idea similar to van den Eijkel (1999); Meghdadi and Akbarzadeh-T. (2001); Liu and Li (2005); Zhang and Li (2010, 2012) where the different concepts (Zadeh, 1968, 1995; Dubois and Prade, 1997; Bertoluzza et al., 2002) of fuzziness and randomness can co-occur. This model can estimate a probability density function or provide an approximation of a non-linear system from sampled data while keeping a linguistic link between input and output variables. Besides the information provided by the linguistic interpretation of the rules, the probabilistic fuzzy model proposed allows to gain information and process understanding on the approximated system given by two different reasoning mechanisms.

A probabilistic fuzzy system, as it was formally defined in van den Berg et al. (2013), is based on the probability of a fuzzy event and estimates conditional probability density functions for the output variable, given the inputs to the system. Two equivalent additive reasoning mechanisms have been proposed, one based on the concept of fuzzy histograms and another based on the stochastic mapping between fuzzy antecedents and fuzzy consequents. This type of models has been applied to real world problems, (van den Berg et al., 2002a, 2004; Almeida and Kaymak, 2009b; Xu and Kaymak, 2008; Almeida et al., 2012a, e.g.) and a framework to develop PFS from data has been proposed in Tang et al. (2012). The proposed framework uses a heuristic method to partition the input and output spaces into a determined number of fuzzy sets and the probability parameters are estimated based on a definition of conditional probability of fuzzy events which has been show to be biased and inconsistent (Waltman et al., 2005b).

[1] Parts of this chapter have been published in Almeida, Verbeek, Kaymak, and Costa Sousa (2014c).

In this work we analyse different aspects of probabilistic fuzzy systems in the context of function approximation and conditional density estimation. We derive several higher moments of the conditional probability density function (pdf) estimated using a PFS. The role of higher moments has become increasingly important in risk management since traditional measures of risk, such as the variance, do not fully capture the distributions of stock market returns. Other measures of distributional shape such as higher moments can be useful in obtaining a better description of risk. Furthermore, even though two different portfolios may share the same variance, investors may be more interested in skewed portfolios that match their risk profile. Conventional measures of skewness and kurtosis are essentially based on sample averages, and thus also sensitive to outliers. In this work we show that higher moments can be derived from the output conditional probability density of a PFS. Furthermore, we analyze the necessary conditions for a PFS, such that the estimated output density is a proper pdf and subsequent higher moments derived from this density exist. These conditions relax the previous assumption of well-formed output space and are not very restrictive or consider a particular definition of conditional probability of fuzzy systems.

We analyse probabilistic fuzzy system in relation to other different types of systems that have universal approximation capability, but usually estimate only a deterministic output, not a full density. This analysis allows to gain a different insight into PFS. Special attention is given to the relation of PFS with fuzzy additive systems, such as Mamdani fuzzy system with weighted output or a fuzzy relational model, and how a PFS can be obtained from this fuzzy additive system. Furthermore, this relation indicates that a PFS is also suitable for problems of function approximation, although this type of application has received very little attention.

A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred to PFS, while providing understanding of different aspects of a probabilistic fuzzy system. The analysis of the parameters of PFS are further extended for function approximation and conditional density estimation, while suggestions on how to obtain PFS models from sampled data are provided. In this work, the effect of different optimization techniques and designs of PFS on function approximation and conditional density estimation are illustrated using simulated data. Using simulated data we show that PFS provides accurate density approximations and conditional density estimates, and how the number of input and output memberships, choice of the PFS membership functions and estimation methods affect the performance of PFS. Finally, we apply the PFS model for function approximation and conditional density estimation of the quarterly US inflation data and report the obtained results using different PFS parametrization and optimization techniques. The estimation and forecast of the complete inflation density is a major concern for macroeconomic policy makers and financial institutions. This analysis shows that slowly changing patterns in

inflation are accurately captured by the PFS model. The PFS model performs well in one quarter ahead and 1 year ahead forecasts of inflation.

This chapter is organized as follows. In Section 3.2 we summarize the two possible and equivalent reasoning mechanisms of PFS presented in van den Berg et al. (2013) and extend it to the multiple output case. We show that higher moments, such as skewness and kurtosis, of the output conditional probability density can be derived for a PFS. Furthermore, we analyse the necessary conditions for a PFS, such that the estimated output density is a proper pdf and subsequent higher moments of this density exist. This analysis is not based on the probability of a fuzzy event (Zadeh, 1968) and as such relaxes the previous assumption of a well-formed output space. The relation to systems with deterministic output that have universal approximation capability is presented in Section 3.3. A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred to PFS. Furthermore, we analyse a PFS as a fuzzy additive system and discuss how a PFS can be obtained from this fuzzy additive system. Section 3.4 presents an analysis of the parameters of a PFS for function approximation or density approximation. Several suggestions on how to obtain these parameters are provided. The influence of these parameters on the accuracy of a PFS are further discussed in Section 3.5 for simulated data. Finally, Section 3.6 presents a real-world example of point and density estimates of quarterly US inflation. Conclusion and future work are presented in Section 3.7.

3.2 Probabilistic fuzzy systems revisited

Probabilistic fuzzy systems combine two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. In this work we consider that the probabilistic uncertainty relate to aleatoric variability, while fuzziness or linguistic vagueness relate to epistemic uncertainty or to the concept of gradualness (Dubois and Prade, 1997, 2012).

In mathematical terms a fuzzy set F on a finite universe U is defined by a membership function $u_F : U \to [0, 1]$ and $u_F(x)$ is the grade of membership of element of $x \in U$ in F. A probabilistic measure P is a mapping $2^U \to [0, 1]$ that assigns a number $P(A)$ of event A to each subset of U, satisfying the Kolmogorov axioms. $P(A)$ is the probability that an ill-known single-valued variable x hits set A. At the mathematical level the domain of the mapping P is the Boolean algebra 2^U while the set of fuzzy sets is $[0, 1]^U$ (Dubois and Prade, 1993).

The PFS consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions. Assuming that the input space is a subset of $\mathbb{R}^n$ and that the rule consequents are defined on a finite domain $Y \subseteq \mathbb{R}$, a probabilistic fuzzy system consists of

a system of rules R_q, $q = 1, \ldots, Q$, of the type (van den Berg et al., 2013)

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } f(y|A_q)\,, \tag{3.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an input vector, $A_q : X \longrightarrow [0, 1]$ is a fuzzy set defined on X and $f(y|A_q)$ is the conditional pdf of the stochastic output variable $\underline{y}$ given the fuzzy event A_q. The interpretation is as follows: if fuzzy antecedent A_q is fully valid ($x \in \text{core}(A_q)$), where $\text{core}(A) = \{x|u_A(x) = 1\}$ then y is a sample value from the probability distribution with conditional pdf $f(y|A_q)$.

For the purpose of this study, we consider two possible and equivalent reasoning mechanisms of PFS, namely the fuzzy histogram approach and the probabilistic fuzzy output approach (van den Berg et al., 2013). In both cases, we suppose that J fuzzy classes C_j form a fuzzy partition of the compact output space Y.

3.2.1 Fuzzy histogram model

In the fuzzy histogram approach, we replace in each rule of (3.1) the true pdf $f(y|A_q)$ by its fuzzy approximation (fuzzy histogram) $\hat{f}(y|A_q)$ yielding the rule set $\hat{R}_q$, $q = 1, \ldots, Q$ defined as (van den Berg et al., 2013)

$$\hat{R}_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } \hat{f}(y|A_q)\,, \tag{3.2}$$

where $\hat{f}(y|A_q)$ is a fuzzy histogram conform (van den Berg et al., 2001)

$$\hat{f}(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}\,. \tag{3.3}$$

The numerator in (3.3) describes a superposition of fuzzy events described by their membership functions $u_{C_j}(y)$, weighted by the probability $\hat{\Pr}(C_j|A_q)$ of the fuzzy event. The denominator of (3.3) is a scaling factor representing the fuzzified size of class C_j. Because of overlapping membership functions, fuzzy histograms have a high level of statistical efficiency, compared to crisp ones. We show this in Fig. 3.1 where the probability density function (pdf) of the standard normal distribution is approximated by a classical and by a fuzzy histogram using in both cases a partitioning in seven classes. For more details we refer to van den Berg et al. (2004). Besides a high level of statistical efficiency, several classes of fuzzy histograms also have a high level of computational efficiency. An example of such type of fuzzy histogram is one that uses triangular membership functions (Waltman et al., 2005a).

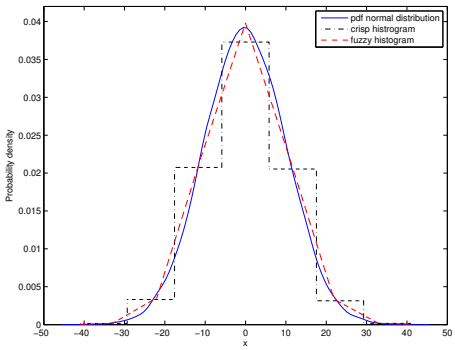


Figure 3.1: A fuzzy histogram approximates a pdf better than a crisp histogram.

Using the definition of fuzzy histogram, the unconditional probability density function of a probabilistic fuzzy system is given by

$$\hat{f}(y) = \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}. \tag{3.4}$$

The interpretation of this type of reasoning is as follows. Given the occurrence of a (multi-dimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, an estimate of the conditional probability density function based on a fuzzy histogram $\hat{f}(y|A_q)$ is calculated.

3.2.2 Probabilistic fuzzy output model

In the probabilistic fuzzy output approach, sometimes also referred to as Mamdani PFS (Kaymak et al., 2003; Xu and Kaymak, 2008; Almeida and Kaymak, 2009b), we decompose each rule (3.1) to provide a stochastic mapping between its fuzzy antecedents and its fuzzy consequents. The rules are written in the following form (van den Berg et al., 2013):

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } \underline{y} \text{ is } C_1 \text{ with } \hat{\Pr}(C_1|A_q) \text{ and} \\ \ldots \\ \underline{y} \text{ is } C_J \text{ with } \hat{\Pr}(C_J|A_q). \end{aligned} \tag{3.5}$$

These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The rules of a PFS also express linguistic information and they can be used to

explain the model behaviour by a set of linguistic rules. This way, the system deals both with linguistic uncertainty as well as probabilistic uncertainty.

The interpretation for the probabilistic fuzzy output approach is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events C_j is likely to occur. The selection of consequent fuzzy events is done proportionally to the conditional probabilities $\hat{\mathrm{Pr}}(C_j|A_q)$, $j = 1, 2, \ldots, J$. This applies for all the rules R_q, $q = 1, 2, \ldots, Q$.

3.2.3 Outputs of probabilistic fuzzy system

Although the fuzzy rule bases (3.2) and (3.5) are different, under certain conditions, the two corresponding probabilistic fuzzy systems implement the same crisp input-output mapping (van den Berg et al., 2013). The output of the fuzzy rules (3.5) is the same as in the rules (3.2), if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents (van den Berg et al., 2004).

Given an input vector $\mathbf{x}$, the output of a probabilistic fuzzy system is a conditional density function which can be computed as

$$\hat{f}(y|\mathbf{x}) = \sum_{j=1}^{J} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{\mathrm{Pr}}(C_j|A_q) \frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}, \tag{3.6}$$

where

$$\beta_q(\mathbf{x}) = \frac{u_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} u_{A_{q'}}(\mathbf{x})} \tag{3.7}$$

is the normalised degree of fulfillment of rule R_q and u_{A_q} is the degree of fulfillment of rule R_q. When $\mathbf{x}$ is n-dimensional, u_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.,*

$$u_{A_q}(\mathbf{x}) = u_{A_{q_1}}(x_1) \circ \cdots \circ u_{A_{q_n}}(x_n), \tag{3.8}$$

where $x_i, i = 1, \ldots, n$ is the i-th component of $\mathbf{x}$ and $\circ$ denotes a t-norm. A t-norm is a binary operation on the interval $[0, 1]$ that satisfies at least the following axioms $\forall a, b, c \in [0, 1]$ (Klir

and Yuan, 1995):

$$a \circ 1 = a, \text{ boundary condition} \tag{3.9a}$$

$$b \leq c \Rightarrow a \circ b \leq a \circ c, \text{ monotonicity} \tag{3.9b}$$

$$a \circ b = b \circ a, \text{ commutativity} \tag{3.9c}$$

$$a \circ (b \circ c) = (a \circ b) \circ c, \text{ associativity} \tag{3.9d}$$

Some commonly used t-norms are the min, product or Łukasiewicz. Since the premise part of each PFS if-then rule does not necessarily include conditions on every element of the input vector, $\mathbf{x}$ may include only elements of the input vector which are conditioned in the premise of rule q, *i.e.* $\mathbf{x}_q \subset \mathbf{x}$. Without loss of generality, to alleviate a cumbersome notation, we will always use the general $\mathbf{x}$ except where necessary for clarity of explanation.

An interesting characteristic of probabilistic fuzzy system is that based on the obtained conditional output probability distribution function $\hat{f}(y|\mathbf{x})$ it is possible to calculate central moments of this distribution of a random variable about the random variable's mean, such as variance, or standardized moments such as skewness and kurtosis. The various moments form one set of values by which the properties of a probability distribution can be usefully characterised. Before we define central moments of the output probability distribution function of a probabilistic fuzzy system, it is necessary to introduce the necessary mathematical formulation.

Let $g(\underline{y})$ be a real-valued function of a continuous random variable $\underline{y}$ with distribution function $f_{\underline{y}}(y)$. The mathematical expectation of $g(\underline{y})$ is denoted by $E(g(\underline{y}))$ and defined as

$$E(g(\underline{y})) = \int_{-\infty}^{\infty} g(y) f_{\underline{y}}(y) dy \tag{3.10}$$

For a continuous univariate probability distribution $f_{\underline{y}}(y)$ the oth moment, with $o \in \mathbb{N}^+$, is defined as

$$m_{o,\underline{y}} = E(\underline{y}^o) = \int_{-\infty}^{\infty} y^o f_{\underline{y}}(y) dy. \tag{3.11}$$

and the oth central moment about the mean η

$$\mu_{o,\underline{y}} = E((\underline{y} - \eta)^o) = \int_{-\infty}^{\infty} (y - \eta)^o f_{\underline{y}}(y) dy. \tag{3.12}$$

The first four moments and central moments of a continuous random variable $\underline{y}$ are

$$m_{1,\underline{y}} = \eta_{\underline{y}}, \quad \mu_{1,\underline{y}} = 0 \tag{3.13}$$

$$m_{1,\underline{y}} = E(y^2), \ \mu_{2,\underline{y}} = E((y-\eta_{\underline{y}})^2) = E(y^2) - \eta_{\underline{y}}^2 = \sigma_{\underline{y}}^2 \tag{3.14}$$

$$m_{3,\underline{y}} = E(y^3), \ \mu_{3,\underline{y}} = E((y-\eta_{\underline{y}})^3) = E(y^3) - 3\eta_{\underline{y}}\sigma_{\underline{y}}^2 - \eta_{\underline{y}}^3 \tag{3.15}$$

$$\begin{aligned} m_{4,\underline{y}} = E(y^4), \ \mu_{4,\underline{y},1} &= E((y-\eta_{\underline{y}})^4) \\ &= E(y^4) - 4\eta_{\underline{y}}E(y^3) + 2\eta_{\underline{y}}^2 E(y^2) + 4\eta_{\underline{y}}^2\sigma_{\underline{y}}^2 + \eta_{\underline{y}}^4 \end{aligned} \tag{3.16}$$

The first moment $m_{\underline{y},1}$ of $\underline{y}$ is the mean $\eta_{\underline{y}}$ and is a measure of centrality, while the second central moment $\mu_{2,\underline{y}}$ is the variance $\sigma_{\underline{y}}^2$. Skewness is defined in relation to the third central moment as

$$\gamma_{1,\underline{y}} = \frac{\mu_{3,\underline{y}}}{\sigma_{\underline{y}}^3}. \tag{3.17}$$

and kurtosis is defined with relation to the fourth central moment

$$\psi_{2,\underline{y}} = \frac{\mu_{4,\underline{y}}}{\sigma_{\underline{y}}^4} \tag{3.18}$$

and the excess kurtosis is defined as

$$\gamma_{2,\underline{y}} = \frac{\mu_{4,\underline{y}}}{\sigma_{\underline{y}}^4} - 3. \tag{3.19}$$

In the case of a probabilistic fuzzy system with rule base (3.2) and (3.5) with conditional output probability distribution function $\hat{f}(y|\mathbf{x})$ the estimated expected output is given by the weighted average of the estimated expected output of each rule according to (van den Berg et al., 2013)

$$\hat{\eta}_{\underline{y}|\mathbf{x}} = \hat{\mu}_{1,\underline{y}|\mathbf{x}} = \hat{\mathrm{E}}(\underline{y}|\mathbf{x}) = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y|A_q) = \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)z_{1,j}\,, \tag{3.20}$$

where $\hat{\mathrm{E}}(\underline{y}|A_q)$ is the estimated expected output of each rule and $z_{o,j}$ is defined as

$$z_{o,j} = \frac{\int_{-\infty}^{\infty} y^0 u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,. \tag{3.21}$$

For the case of $o = 1$, (3.21) is the centroid of the jth output fuzzy set.

The estimated conditional variance $\hat{\sigma}^2_{\underline{y}|\mathbf{x}}$, can be calculated as (van den Berg et al., 2013)

$$\begin{aligned}\hat{\sigma}^2_{\underline{y}|\mathbf{x}} = \hat{\mu}_{2,\underline{y}|\mathbf{x}} = \hat{\mathrm{E}}(\underline{y}^2|\mathbf{x}) - (\hat{\mathrm{E}}(\underline{y}|\mathbf{x}))^2 &= \sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(\underline{y}^2|A_q) - \hat{\eta}^2_{\underline{y}|\mathbf{x}} \\ &= \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)z_{2,j} - \hat{\eta}^2_{\underline{y}|\mathbf{x}} \end{aligned} \tag{3.22}$$

where $(\hat{\mathrm{E}}(\underline{y}|A_q))^2$ is the estimated variance of the output of each rule. Similarly, based on the third and fourth standardized moments, the standardized moments skewness and kurtosis of the conditional density output of a PFS is given by

$$\begin{aligned}\hat{\gamma}_{1,\underline{y}|\mathbf{x}} = \frac{\hat{\mu}_{3,\underline{y}|\mathbf{x}}}{\hat{\sigma}^3_{\underline{y}|\mathbf{x}}} &= \left(\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y^3|A_q) - 3\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y|A_q)\hat{\sigma}^2_{\underline{y}|\mathbf{x}} - \hat{\eta}^3_{\underline{y}|\mathbf{x}}\right)(\hat{\sigma}^3_{\underline{y}|\mathbf{x}})^{-1} \\ &= \left(\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)z_{3,j}\right. \\ &\qquad \left. -3\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)z_{1,j}(\hat{\eta}_{\underline{y}|\mathbf{x}}\hat{\sigma}^2_{\underline{y}|\mathbf{x}}) - \hat{\eta}^3_{\underline{y}|\mathbf{x}}\right)(\hat{\sigma}^3_{\underline{y}|\mathbf{x}})^{-1} \\ &= \left(\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)\left(z_{3,j} - 3z_{1,j}\hat{\eta}_{\underline{y}|\mathbf{x}}\hat{\sigma}^2_{\underline{y}|\mathbf{x}}\right) - \hat{\eta}^3_{\underline{y}|\mathbf{x}}\right)(\hat{\sigma}^3_{\underline{y}|\mathbf{x}})^{-1}, \end{aligned} \tag{3.23}$$

and the kurtosis is

$$\begin{aligned}\psi_{2,\underline{y}|\mathbf{x}} = \frac{\hat{\mu}_{4,\underline{y}|\mathbf{x}}}{\hat{\sigma}^4_{\underline{y}|\mathbf{x}}} &= \left(\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y^4|A_q) - 4\hat{\eta}_{\underline{y}|\mathbf{x}}\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y^3|A_q)\right. \\ &\quad \left. + 2\hat{\eta}^2_{\underline{y}|\mathbf{x}}\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(y^2|A_q) + 4\hat{\eta}^2_{\underline{y}|\mathbf{x}}\hat{\sigma}^2_{\underline{y}|\mathbf{x}} + \hat{\eta}^4_{\underline{y}|\mathbf{x}}\right)(\hat{\sigma}^4_{\underline{y}|\mathbf{x}})^{-1} \\ &= \left(\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)\left(z_{4,j} - 4z_{3,j} + 2\hat{\eta}^2_{\underline{y}|\mathbf{x}}z_{2,j}\right)\right. \\ &\quad \left. + 4\hat{\eta}^2_{\underline{y}|\mathbf{x}}\hat{\sigma}^2_{\underline{y}|\mathbf{x}} + \hat{\eta}^4_{\underline{y}|\mathbf{x}}\right)(\hat{\sigma}^4_{\underline{y}|\mathbf{x}})^{-1}, \end{aligned} \tag{3.24}$$

where $\hat{\mathrm{E}}(\underline{y}|A_q)$ is the estimated expected output of each rule, $(\hat{\mathrm{E}}(\underline{y}|A_q))^2$ is the estimated variance of the output of each rule.

3.2.4 Necessary conditions for probabilistic fuzzy systems

There are conditions that a probabilistic fuzzy system has to follow such that the system output will be a proper probability density function $\hat{f}(y|\mathbf{x})$ and the crisp outputs, expected value $\hat{\mathrm{E}}(\underline{y}|\mathbf{x})$ and conditional variance $\hat{\mathrm{E}}(\underline{y}^2|\mathbf{x})$, exist. The necessary conditions are summarized in the following theorem.

Theorem 3.1 *Let $R = \cup_{q=1}^{Q} R_q$ be a fuzzy rule base consisting of the rules of type* (3.1)*. The necessary conditions for the output of a probabilistic fuzzy system of the form* (3.2) *or* (3.5)*, with an input space $\mathbf{X}$ partitioned in $q = 1, \ldots, Q$ fuzzy sets A_q and output space Y partitioned in $j = 1, \ldots, J$ fuzzy sets C_j, to be a proper probability distribution function* i.e. $\int_{-\infty}^{\infty} f(y|\mathbf{x})dy = 1$ *and that the four moments defined by* (3.20)*,* (3.22)*,* (3.23) *and* (3.24) *exist are:*

$$\sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) = 1 \tag{3.25}$$

$$\hat{\mathrm{Pr}}(C_j|A_q) \geq 0 \tag{3.26}$$

$$u_{A_q}(\mathbf{x}) > 0 \tag{3.27}$$

$$u_{C_j}(\mathbf{x}) > 0 \tag{3.28}$$

$$\sum_{q=1}^{Q} \beta_q(\mathbf{x}) = 1 \tag{3.29}$$

$$\int_{-\infty}^{\infty} u_{C_j}(y)dy < \infty \; j = 1, \ldots, J \tag{3.30}$$

Proof: Condition (3.25) and (3.26) are regular conditions that probabilities should satisfy, while conditions (3.27) and (3.28) are regular conditions that membership values should satisfy. Conditions (3.25), (3.29) and (3.30) ensure that the output of the system is a proper pdf,

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{f}(y|\mathbf{x})dy &= \int_{-\infty}^{\infty} \beta_q(\mathbf{x})\hat{f}(y|A_q)dy \\ &= \int_{-\infty}^{\infty} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \frac{\sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q)u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} dy \\ &= \sum_{j=1}^{J}\sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q) \frac{\int_{-\infty}^{\infty} u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} = 1.\end{aligned} \tag{3.31}$$

Condition (3.30) also ensures that the four moments defined by (3.20), (3.22), (3.23) and (3.24), as well as other higher moments exist. Given that all moments

$$\hat{m}_{1,\underline{y}|A_q} = \mathrm{E}(\underline{y}^n|A_q) = \int_{-\infty}^{\infty} y^n \hat{f}(y|A_q)dy \tag{3.32}$$

and central moments

$$\hat{\mu}_{n,\underline{y}|A_q} = \hat{E}((\underline{y}|A_q - \eta)^n) = \int_{-\infty}^{\infty} (y-\eta)^n \hat{f}(y|A_q)dy. \tag{3.33}$$

depend on fuzzy histogram (3.3)

$$\hat{f}(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\mathrm{Pr}}(C_j|A_q)u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} < \infty\,. \tag{3.34}$$

■

We note that (3.25) and (3.26) are functions defined on the set of conditional events, without assuming any given algebraic structure. Examples of definitions of conditional probabilities of fuzzy events that satisfy the classical axioms of conditional probabilities as given by de Finetti (1949) and Popper (1959) can be found in Coletti and Scozzafava (2006) and Baldwin et al. (1996).

Although not strictly necessary, a desirable characteristic of a probabilistic fuzzy system may be that the sum of the probabilities of the fuzzy consequents equals one. For this it is necessary to ensure that the output space is well-formed, *i.e.* the output membership values satisfy (van den Berg et al., 2004)

$$\sum_{j=1}^{J} u_{C_j}(y) = 1, \qquad \forall y \in Y. \tag{3.35}$$

This condition limits the output of a probabilistic fuzzy system as Section 3.4.2 shows.

3.2.5 Probabilistic fuzzy systems with multiple outputs

The results presented in the previous sections can be extended for the case of multiple outputs, following the distinction between fuzzy input and conditional density output of (3.2) and stochastic mapping between fuzzy antecedents and fuzzy consequents of (3.5). The basic idea is that each one of the d outputs will have an independent probability density function conditional on the same input variables $\mathbf{x}$, making the output of each PFS rule is defined by multiple

densities. The fuzzy histogram model rules can be written as

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f_1(y) \text{ is } f_1(y_1|A_q) \text{and} f_2(y) \text{ is } f_2(y_2|A_q) \ldots \text{and} f_d(y) \text{ is } f_d(y_d|A_q)\,, \quad (3.36)$$

and the probabilistic fuzzy output rules are

$$\begin{aligned} \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } \underline{y_1} \text{ is } C_{1,1} \text{ with } \hat{\Pr}(C_{1,1}|A_q) \text{ and} \ldots \underline{y_d} \text{ is } C_{d,1} \text{ with } \hat{\Pr}(C_{d,1}|A_q) \text{ and} \\ \ldots \\ \underline{y_1} \text{ is } C_{1,J} \text{ with } \hat{\Pr}(C_{1,J}|A_q) \text{ and} \ldots \underline{y_d} \text{ is } C_{d,J} \text{ with } \hat{\Pr}(C_{d,J}|A_q). \end{aligned} \quad (3.37)$$

For all d outputs, $\hat{\Pr}(C_{d,J}|A_q)$ must satisfy necessary conditions (3.25) and (3.26), while $C_{d,J}$ must satisfy (3.28) and (3.30).

If conditional probability parameters are obtained by maximizing the likelihood function as explained in Section 3.4.2, the likelihood function should take into account the multiple output densities defined by each rule, and combine these densities when deriving the likelihood function. If the multiple outputs of each rule output are assumed to be independent of each other, derivation of the likelihood is straightforward, *i.e.* one only has to multiply the conditional densities obtained in each rule output. This assumption is not very restrictive, since the independence only implies that the 'unexplained' part of the output is independent, given the relation with antecedent variables.

The second method of obtaining parameter estimates described in Section 3.4.2 is based on minimizing the mean squared error. This optimization does not require an explicit assumption on the independence of multiple outputs of each rule. However, since the objective function is the average of squared errors, each squared error has the same weight in the objective function. Possible dependency between these errors is not incorporated in this objective function. Therefore minimizing the mean squared error is implicitly similar to maximizing the likelihood function which is based on independent errors.

3.3 Function equivalence to systems with deterministic output

In this section, we consider the function equivalence of the probabilistic fuzzy system described in Section 3.2 to systems with deterministic output. In particular, we are interested in the relation between probabilistic fuzzy systems and Takagi–Sugeno (TS) fuzzy systems (Takagi and Sugeno, 1985), Mamdani fuzzy systems (Mamdani and Gaines, 1981) and radial basis function network (Broomhead and Lowe, 1988; Moody and Darken, 1989). The relation between

these systems is well known (Hunt et al., 1996), but the relation with probabilistic fuzzy system helps to explain its success for function approximation. A practical relevance of the functional equivalence result is that learning algorithms, optimization techniques and design issues can be transferred across different paradigms. Furthermore, it also allows to interpret models transversely across different modeling paradigms.

The results presented in this section allow to analyse and understand different aspects of a probabilistic fuzzy system, for the purpose of designing such systems.

3.3.1 Additive fuzzy models

In addition to the probabilistic reasoning presented in Section 3.2.2, in this section we departure from this definition and instead derive a probabilistic fuzzy system from an additive fuzzy system. This deterministic fuzzy system has rule base multiplicative implication and additive aggregation, where the crisp model output is obtained using the center of gravity defuzzification method.

Let $R = \cup_{q=1}^{Q} R_q$ be a rule base for a additive fuzzy system of the type

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then} \quad & \underline{y} \text{ is } C_1 \text{ with } w(A_q, C_1) \text{ and} \\ & \dots \\ & \underline{y} \text{ is } C_J \text{ with } w(A_q, C_J)\,, \end{aligned} \tag{3.38}$$

where $w(A_q, C_j) \in \mathbb{R}_{\geq 0}$ are non-negative weights. The system defined by (3.38) is similar to the standard additive model (Kosko, 1997, 1998) but in the former, the consequents are not directly dependent on x.

Although the fuzzy rule base system defined by (3.38) can be obtained by replacing the conditional probabilities $\hat{\Pr}(C_j|A_q)$ by non-negative weights $w(A_q, C_j) \in \mathbb{R}_{\geq 0}$ in the fuzzy rule system (3.5), the crisp output of both systems is different, as the following theorem shows.

Theorem 3.2 *Let $R = \cup_{q=1}^{Q} R_q$ be a fuzzy rule base as defined by* (3.38) *such that $u_{A_q}(\mathbf{x}) > 0, \forall q$ and the output space follows* (3.35)*, and the rule base uses multiplicative implication and additive aggregation. Then the crisp model output y^* obtained using the center of gravity defuzzification method is*

$$y^* = \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j} z_{1,j}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j}}\,, \tag{3.39}$$

where $z_{1,j}$ is given by (3.21) *and $v_{1,j}$ is the area of the jth output fuzzy set defined by*

$$v_{1,j} = \int_{-\infty}^{\infty} u_{C_j}(y)dy\,. \tag{3.40}$$

Proof The center of gravity defuzzification method is given by

$$y^* = \frac{\int_{-\infty}^{\infty} y\chi(x,y)dy}{\int_{-\infty}^{\infty} \chi(x,y)dy}\,, \tag{3.41}$$

where $\chi(x,y)$ is the output of the fuzzy system under consideration. For the case of the additive fuzzy system (3.38) using with multiplicative implication and additive aggregation the output is

$$\chi(x,y) = \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)u_{C_j}(y)\,. \tag{3.42}$$

Substituting (3.42) into (3.41) we obtain

$$\begin{aligned}
y^* &= \frac{\int_{-\infty}^{\infty} y \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)u_{C_j}(y)dy}{\int_{-\infty}^{\infty} \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)u_{C_j}(y)dy} \\
&= \frac{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)\int_{-\infty}^{\infty} yu_{C_j}(y)dy}{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)\int_{-\infty}^{\infty} u_{C_j}(y)dy} \\
&= \frac{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)\int_{-\infty}^{\infty} u_{C_j}(y)dy\frac{\int_{-\infty}^{\infty} yu_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}}{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)\int_{-\infty}^{\infty} u_{C_j}(y)dy} \\
&= \frac{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)v_{1,j}z_{1,j}}{\sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})w(A_q,C_j)v_{1,j}}\,.
\end{aligned} \tag{3.43}$$

□

Starting from an additive fuzzy system defined in (3.38), it is possible to obtain a probabilistic fuzzy system. Before formalizing this result we introduce the following definition of a probability kernel.

Definition 3.3 *A kernel is a mapping $\mathcal{K} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ from the measurable space $(\mathbf{X}, \mathcal{X})$ to the measurable space $(\mathbf{Y}, \mathcal{Y})$. The kernel $\mathcal{K}$ is a probability kernel if it is defined as a probability measure on $(\mathbf{Y}, \mathcal{Y})$.*

Given this definition we can now prove that a probabilistic fuzzy system can be obtained starting from the fuzzy system defined in (3.38).

Theorem 3.4 *If the mapping $w(A_q, C_j)$ is defined as a probability kernel and each output consequent C_j are functions defined on a random variable space then the output of the PFS is a conditional probability density for y given x. Under this definition, the fuzzy rule base in* (3.38) *has a functional equivalent to the PFS in* (3.5) *and the crisp output* (3.39) *has a functional equivalent to the conditional output of the PFS in* (3.6).

Proof The defined non-negative weights $w(A_q, C_j) : (\mathbf{X} \times Y) \to R_{\geq 0}$ form a kernel on the measurable space $(\mathbb{R}^n \times \mathbb{R})$. If $w(A_q, C_j)$ is also defined as a probability measure on $(Y, \mathcal{Y})$, such that $\sum_{j=1}^{J} w(A_q, C_j) = 1, \forall q = 1, \ldots, Q$ then according to Definition 3.3, $w(A_q, C_j)$ is a probability kernel. We recall that using (3.7) we obtain $\sum_{q=1}^{Q} \beta_q(\mathbf{x}) = 1$. Furthermore, since the output fuzzy sets C_j are admissible functions for defining random variables then they are limited to those for which a probability distribution exists. A simple form to ensure this is to normalize them

$$u_{C_j'} = \frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} . \tag{3.44}$$

The output of the fuzzy system $\chi(x, y)$ in (3.41) is then a conditional density function for Y given X such that:

$$\begin{aligned}\int_{-\infty}^{\infty} \chi(x, y)dy &= \int_{-\infty}^{\infty} \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) u_{C_j'} dy \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) \frac{\int_{-\infty}^{\infty} u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} = 1 .\end{aligned} \tag{3.45}$$

In the case that $w(A_q, C_j)$ is defined as a probability kernel, the additive fuzzy system defined by the rule base (3.38) is a probabilistic fuzzy system as presented in (3.5). Furthermore, the center of gravity output (3.39) of the additive fuzzy system has a functional equivalent to the expectation of the conditional output of the PFS (3.6)

$$\begin{aligned} y^* &= \frac{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j} z_{1,j}}{\sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) v_{1,j}} \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) w(A_q, C_j) z_{1,j}. \end{aligned} \tag{3.46}$$

Since $w(A_q, C_j)$ is a probability kernel, (3.46) is equivalent to (3.20).

□

As a result of theorem 3.2 and theorem 3.4, a Mamdani fuzzy model can be regarded as a special case of the fuzzy system defined in (3.38), or equivalently the system defined by (3.5).

A Mamdani fuzzy model is recovered when the system is purely deterministic by setting setting for all $q = 1, \ldots, Q$, $\exists \kappa \in \{1 \ldots, J\}$ such that $\hat{\Pr}(C_\kappa|A_q) = 1$ and $\hat{\Pr}(C_j|A_q) = 0$, $j \neq \kappa$ *i.e.,* only one of the possible consequents is certain for each rule Q.

The previous results have shown that a probabilistic fuzzy system defined by (3.5) can be obtained starting from a additive fuzzy system (3.38). An important aspect is that since $w(A_q, C_j)$ is defined as a probability kernel then it has a functional equivalent to $\Pr(C_j|A_q)$ in (3.5), implying that $\sum_{j=1}^{J} \hat{\Pr}(C_j|A_q) = 1$ and $\hat{\Pr}(C_j|A_q) \geq 0$. In this chapter we do not assume any particular algebraic structure for the conditional probability of fuzzy events. There are several examples of definitions of conditional probabilities of fuzzy events that satisfy the classical axioms of conditional probabilities as given by de Finetti (1949) and Popper (1959) that can be found in Coletti and Scozzafava (2006); Baldwin et al. (1996); Singpurwalla and Booker (2004). This is an important issue that needs to be studied closely in the future.

It is also interesting to note that the system defined by (3.38) can be transformed in a fuzzy relational model (Pedrycz, 1985) when $w(A_q, C_j)$ is replaced by the fuzzy relation $u(A_q, C_j)$. Similarly to a fuzzy relational model, a probabilistic fuzzy system can also be fine tuned by modifying the probability parameters $\hat{\Pr}(C_j|A_q)$, while maintaining the fuzzy input and fuzzy output space constant. We stress that a fuzzy relational model and a probabilistic fuzzy system have different interpretations, based on the nature of the uncertainty of the relation and output being modelled, as described in Section 3.1. In a fuzzy relational model the elements of the relation represent the strength of association between the fuzzy sets, while in the case of a fuzzy probabilistic model they are a stochastic mapping between fuzzy sets. Furthermore, the output fuzzy sets of a probabilistic fuzzy system are defined in the space of a stochastic variable $\underline{y}$. These differences leads to different nature of outputs, albeit under certain circumstances, there is a functional equivalence between both models crisp output. In the general case that $w(A_q, C_j)$ are non-negative weights, or in the case of a fuzzy relational model $u(A_q, C_j)$ are fuzzy relations, the output of such a system is not a proper probability density function.

3.3.2 Zero-order Takagi-Sugeno fuzzy models

The relation between the deterministic output of a zero-order Takagi–Sugeno (TS) system (Takagi and Sugeno, 1985) and the expected output of a probabilistic fuzzy system, was previously studied in van den Berg et al. (2013). In this section we show the main result. A zero-order Takagi–Sugeno fuzzy system with Q rules, antecedent fuzzy sets A_q and consequent parameters

b_q has the same output of a probabilistic fuzzy system, provided that (van den Berg et al., 2013)

$$b_q = \sum_{j=1}^{J} \hat{\Pr}(C_j|A_q)z_j \,. \tag{3.47}$$

This result suggests that the PFS as a deterministic system belongs to a general class of general function approximators, called the basis functions expansion (Friedman, 1991) taking the form

$$y = \sum_{q=1}^{Q} \phi_q(x)b_q \,. \tag{3.48}$$

Radial basis function networks also belong to this class of systems and are discussed in the following section.

3.3.3 Radial basis function networks

The type of network under consideration is described by (Hunt et al., 1996; Figueiredo, 2000)

$$y = f(\mathbf{x}) = \sum_{q=1}^{n_\theta} \theta_q(\mathbf{x})\phi_q(\mathbf{x}_q) \tag{3.49}$$

where n_θ is the number of nonlinear processing units (or radial basis function) $\phi_q(\mathbf{x}_q)$ where each unit input vector $\mathbf{x}_q \subset \mathbf{x}$ and $\theta_q(\mathbf{x})$ is the network weighting function. A normalized form of the network is sometimes used and is described by

$$y = f(\mathbf{x}) = \frac{\sum_{q=1}^{n_\theta} \theta_q(\mathbf{x})\phi_q(\mathbf{x}_q)}{\sum_{q=1}^{n_\theta} \phi_q(\mathbf{x}_q)} \,. \tag{3.50}$$

A common form of basis function is the radial Gaussian form described by

$$\phi_q(\mathbf{x}_q) = \exp\left[-(\mathbf{x}_q - \lambda_q)'\Delta_q(\mathbf{x}_q - \lambda_q)\right] , \tag{3.51}$$

where $\lambda_q \in \mathbb{R}^{n_{\mathbf{x}_q}}$ are the center vector of the basis function, $\Delta_q \in \mathbb{R}^{n_{\mathbf{x}_q} \times n_{\mathbf{x}_q}}$ is a diagonal width parameter matrix $\Delta_q = \mathrm{diag}[\delta_{q1}^{-2} \ldots \delta_{qn_{\mathbf{x}_q}}^{-2}]$ and δ_q is the width of each basis function.

The radial basis function described above is sometimes used in its standard form (Heimes and van Heuveln, 1998; Bugmann, 1998). The standard Gaussian basis function applies the same width parameter in each dimension, *i.e.,* $\Delta_q = \mathrm{diag}[\delta_q^{-2} \ldots \delta_q^{-2}]$, each unit processes the

whole input vector, *i.e.,* $n_{\mathbf{x}_q} = n_{\mathbf{x}}$ and $\phi_q(\mathbf{x}_q) = \phi_q(\mathbf{x})$, and the local models are constants, *i.e.,* $\theta_q(\mathbf{x}) = \theta_q$.

The functional equivalence between a probabilistic fuzzy system and a radial basis function is presented in the following theorem.

Theorem 3.5 *A generalized Gaussian radial basis function defined by* (3.50) *and* (3.51) *is functionally equivalent to a probabilistic fuzzy system defined by* $R = \cup_{q=1}^{Q} R_q$ *rules of the form* (3.2) *or* (3.5) *meeting conditions* (3.25)-(3.30)*, if the following conditions are satisfied:*

1. *The number of radial basis function units is equal to the number of rules in the probabilistic fuzzy system,* i.e., $n_\theta = Q$.

2. *The output of each rule is the expected output of each rule fuzzy histogram,* i.e. $\hat{\mathrm{E}}(\underline{y}|A_q) = \int_{-\infty}^{\infty} y\hat{f}(y|A_q)dy$.

3. *The local models of the radial basis function network are constants,* i.e., $\theta_q(\mathbf{x}) = \theta_q$.

4. *The membership functions within each rule are chosen as Gaussian functions.*

5. *The t-norm operator used to compute each rule's firing strength is multiplication.*

Proof: Under condition 4) each probabilistic fuzzy rule consists of the composition of the univariate Gaussian functions which define the membership values in the premise part of each rule. Each univariate membership function has the form

$$u_{A_{qi}}(x_{qi}) = \exp\left[-\frac{(\mathbf{x}_{qi} - \lambda_{qi})^2}{\delta_{qi}^2}\right],\ i = i, \ldots, n \tag{3.52}$$

and this defines the ith membership value of the qth rule. Under condition 5) the firing strength of each rule is given by

$$u_{A_q}(x_q) = \prod_{i=1}^{n} u_{A_{qi}}(x_{qi}) \tag{3.53}$$

with $q = 1, \ldots, Q$ because of condition 1). We then obtain

$$\begin{aligned} u_{A_q}(x_q) &= \exp\left[-\frac{(x_{q1} - \lambda_{q1})^2}{\delta_{q1}^2} - \ldots - \frac{(x_{qn} - \lambda_{qn})^2}{\delta_{qn}^2}\right] \\ &= \exp\left[-(\mathbf{x}_q - \lambda_q)'\Delta_q(\mathbf{x}_q - \lambda_q)\right] \end{aligned} \tag{3.54}$$

with $\Delta_q = \text{diag}[\delta_{q1}^{-2} \dots \delta_{qn_{\mathbf{x}_q}}^{-2}]$. Under condition 2) the output of each rule of the probabilistic fuzzy system is given by

$$\begin{aligned}\hat{\text{E}}(\underline{y}|A_q) &= \int_{-\infty}^{\infty} y\hat{f}(y|A_q)dy = \int_{-\infty}^{\infty} y\frac{\sum_{j=1}^{J}\hat{\text{Pr}}(C_j|A_q)u_{C_j}(y)}{\int_{-\infty}^{\infty}u_{C_j}(y)dy}dy \\ &= \sum_{j=1}^{J}\hat{\text{Pr}}(C_j|A_q)z_j = b_q \end{aligned} \tag{3.55}$$

We can now write the expected output of a PFS by combining (3.7) and (3.55) as

$$\hat{\text{E}}(\underline{y}|\mathbf{x}) = \frac{u_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q}u_{A_{q'}}(\mathbf{x})}\sum_{j=1}^{J}\hat{\text{Pr}}(C_j|A_q)z_{1,j} = \frac{u_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q}u_{A_{q'}}(\mathbf{x})}b_q. \tag{3.56}$$

Under condition 3) the normalized radial basis function (3.50) becomes

$$y = f(\mathbf{x}) = \frac{\sum_{q=1}^{n_\theta}\theta_q\phi_q(\mathbf{x}_q)}{\sum_{q=1}^{n_\theta}\phi_q(\mathbf{x}_q)}. \tag{3.57}$$

Comparing (3.56) and (3.54) with (3.57) and (3.51) the functional equivalence is established since the firing strength of each rule functionally equates to the activation of the radial basis function, *i.e.,* $u_{A_q}(x_q) = \phi_q(\mathbf{x}_q)$ and $b_q = \theta_q$.

We again note that the premise part of each PFS if-then rule does not necessarily include conditions on every element of the input vector. Thus, $\mathbf{x}_q$ consists only of the elements of the input vector which are conditioned in the premise of rule q. ■

The kind of radial basis function which has a functional equivalent to PFS are not the standard Gaussian RBF because there are no restriction on the widths of the basis functions and each unit in the network has as input only a subset $\mathbf{x}_q$ of the input vector $\mathbf{x}$, but it is necessary to restrict the generalized Gaussian RBF such that weights θ_q are constant. In this case, the weights θ_q can be seen as local models whose validity is defined by the activation value $\phi_q(\mathbf{x}_q)$. The network smoothly joins these local models together through interpolation to form the global model. The basis function are defined on hyper-ellipsoids in the input space, indicating that the univariate functions making up an RBF unit are less interdependent than if they were restricted.

The results presented in this section provide motivation for the success of probabilistic fuzzy systems for function approximation. Although outside the scope of this chapter, the results presented in this section indicate that, similarly to all deterministic systems discussed, a probabilistic fuzzy system will under certain conditions serve as an universal approximator of continuous functions defined on compact domains with arbitrarily high approximation accuracy. We plan to devote our attention to these conditions in a future study.

3.4 Probabilistic fuzzy system parameters

A practical relevance of the functional equivalence described in the previous section is that learning algorithms, optimization techniques and design issues can be transferred to probabilistic fuzzy systems. Similar to Mamdani, relational, and zero order Takagi-Sugeno fuzzy models, PFS have constant consequents and thus also similar interpolation properties. Therefore, the method to construct the antecedent membership functions can be similar to one of these models.

Depending on whether the objective is function approximation or density approximation, the parameters of a probabilistic fuzzy system can be estimated in different ways. Function approximation relies on the interpolation between the antecedents and consequents. However, for the case of density approximation, we may be interested in regions of the space with the same local density model (Almeida and Kaymak, 2009a; Almeida et al., 2012a).

In this section we provide a possible way to obtain the probabilistic fuzzy system parameters using supervised learning. Supervised learning is concerned with the prediction of a quantitative measure of the output variable y, based on a vector $\mathbf{x} = (x_1, \ldots, x_n)$ of n observed input variables. Let $\mathbf{x}$ be an arbitrary vector, and y the corresponding unknown output. In classical regression literature, the objective is to determine the best mathematical expression describing the functional relationship between one response and one or more independent variables. Following the nomenclature used, the problem is to obtain some information on y from the training set $\mathcal{L} = \{(\mathbf{x}_p, y_p)\}_{p=1}^{P}$ of P observations of the input and output variables.

In this chapter we do not consider the aspect of optimal design. Although the probabilistic fuzzy system approximates conditional pdf's and consequently crisp outputs, the resulting fuzzy system need not be optimal in terms of the number of rules, the definition of antecedent membership functions and consequent membership functions. Particular choices can provide better interpolation for different data sets.

3.4.1 Antecedent membership functions

In this work we determine the parameters of the antecedent membership functions by using a fuzzy clustering heuristic, that uses the fuzzy c-means (Bezdek, 1981) or the Gustafson-Kessel clustering algorithm (Gustafson and Kessel, 1979), on the product space of the antecedent variables, to obtain a fuzzy partition matrix $U = [u_{qp}]$ for $p = 1, \ldots, P$ samples.

Conditional density estimation

For the case of conditional density approximation, each cluster obtained by product-space clustering of the identification data set can be regarded as an approximation of the regression density distribution. The antecedent fuzzy sets A_{qi} can be computed analytically in the antecedent product space, or can be extracted from the fuzzy partition matrix by projections. In order to obtain membership functions for the antecedent fuzzy sets A_{qi}, the multidimensional fuzzy set defined pointwise in the ith row of the partition matrix U are projected onto the axes of the antecedent variables x_n. This is expressed by the point-wise projection operator of the form:

$$u_{A_{qi}}(x_{ip}) = \text{proj}_i(u_{qp}). \tag{3.58}$$

The point-wise defined fuzzy sets A_{qi} are then approximated by appropriate parametric functions, such as Gaussian membership functions

$$f(x_q; \lambda_{qi}, \delta_{qi}) = \exp\left(\frac{-(x_{qi} - \lambda_{qi})^2}{2\delta_{qi}^2}\right), \tag{3.59}$$

$$u_{A_{qi}}(x_{ip}) = f(x_{ip}; \delta_{qi}^1, \lambda_{qi}^1, \delta_{qi}^2, \lambda_{qi}^2) = f^1(x_{ip}; \delta_{qi}^1, \lambda_{qi}^1) f^2(x_{ip}; \delta_{qi}^2, \lambda_{qi}^2) \tag{3.60}$$

where

$$f^1(x_{ip}; \delta_{qi}^1, \lambda_{qi}^1) = \begin{cases} \exp\left(\frac{-(x_{qi}-\lambda_{qi}^1)^2}{2(\delta_{qi}^1)^2}\right) & x_{qi} \leq \lambda_{qi}^1 \\ 1 & \text{otherwise} \end{cases} \tag{3.61}$$

$$f^2(x_{ip}; \delta_{qi}^2, \lambda_{qi}^2) = \begin{cases} \exp\left(\frac{-(x_{qi}-\lambda_{qi}^2)^2}{2(\delta_{qi}^2)^2}\right) & x_{qi} > \lambda_{qi}^2 \\ 1 & \text{otherwise} \end{cases} \tag{3.62}$$

or triangular membership functions

$$f(x_q; a_q^1, a_q^2, a_q^3) = \max\left(\min\left(\frac{x_q - a_q^1}{a_q^2 - a_q^1}, \frac{a_q^3 - x_q}{a_q^3 - a_q^2}\right), 0\right). \tag{3.63}$$

In general, it is considered that an advantage of this method over the multidimensional membership functions is that the projected membership functions can always be approximated in such a form that convex fuzzy sets are obtained.

The smoothness of the model output depends directly on the smoothness of the antecedent membership functions. This restricts the choice of the type of the membership functions. For instance, the trapezoidal membership functions result in nonsmooth outputs.

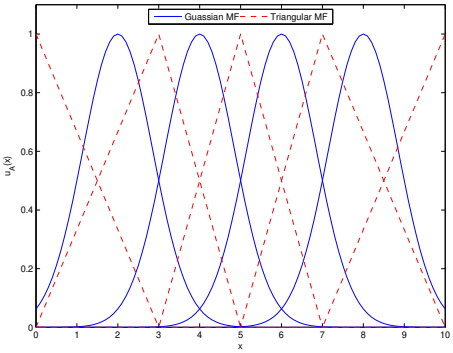


Figure 3.2: Cores of triangular membership function are chosen at the intersection of adjacent Gaussian memberhip functions and at the extreme points of the domain.

Function approximation

For the case of function approximation, the main idea is to construct a system such that the linear submodel represented by one cluster is obtained by interpolation among linguistic fuzzy rules. For this, a heuristic can be used to transform the local submodels such that they will interpolate between rules (Babuška, 1998). Formally this transformation can be described as follows. Consider the antecedents obtained in the previous section, where the fuzzy sets are ordered such that

$$\sup \operatorname{core}(A_q) < \inf \operatorname{core}(A_{q+1}), q = 1, 2, \ldots, Q-1$$

where $\operatorname{core}(A) = \{x | u_A(x) = 1\}$.

This condition also ensures that the cores of the fuzzy sets A_q are disjunct. Denote $A' = \{a'_q | q = 1, \ldots, Q+1\}$ the set of the intersection points of the adjacent fuzzy sets A_q and the infimum and supremum of the domain X:

$$A' = \inf X \cup \{\operatorname{core}(\operatorname{norm}(A_q \cap A_{q+1})) | q = 1, \ldots, Q-1\} \cup \sup X,$$

where $\operatorname{norm}(A)$ denotes the normalization of a fuzzy set, *i.e.*, $A' = \operatorname{norm}(A) \Leftrightarrow u_{A'}(x) = u_A(x) / \sup_{x \in X} u_A(x)$. Triangular membership functions A'_q of the linguistic model can be constructed such that they form a partition and their cores are the points a'_q, defined by:

$$u_{A'_1}(x) = \max\left(0, \min\left(1, \frac{a'_2 - x}{a'_2 - a'_1}\right)\right) , \tag{3.64a}$$

$$u_{A'_q}(x) = \max\left(0, \min\left(\frac{x - a'_{q-1}}{a'_q - a'_{q-1}}, \frac{a'_{q+1} - x}{a'_{q+1} - a'_q}\right)\right) , \tag{3.64b}$$

$$u_{A'_{Q+1}}(x) = \max\left(0, \min\left(0, \frac{x - a'_Q}{a'_{Q+1} - a'_Q}\right)\right) . \tag{3.64c}$$

This idea is illustrated in Fig. 3.2, where the core of five triangular membership functions, as defined by (3.63), are chosen at the intersection of adjacent Gaussian memberhip functions defined by (3.59) and at the extreme points of the domain. In the case of multidimensional A_q the membership functions are derived per variable x_i. To obtain a complete PFS model, it remains to identify the rule consequents for all combinations of the antecedent fuzzy sets and stochastic mapping between input and output fuzzy sets.

3.4.2 Consequent membership functions

The consequent membership functions are constrained according to condition (3.30). A simple way to satisfy condition (3.30) is to define the consequent membership functions in a compact space, for example using triangular membership functions with trapezoidal membership functions at the edges of the domain to ensure that the domain is always covered by the fuzzy partition. Alternatively, it is possible to use a membership function with infinitely large support, such as the Guassian membership function.

Conditional density estimation

The accuracy of the approximation of the conditional probability density function depends on the number of consequent fuzzy sets C_j on the output domain, the fuzzy partitioning of the output space (in conjunction with the input space), or on the shape of the membership functions. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying with more membership functions placed towards the origin (Xu and Kaymak, 2008) or towards the edges of the universe of discourse (Almeida and Kaymak, 2009a). This varying placement allows to better capture the variability in regions with more membership functions. The smoothness of the conditional probability density function depends directly on the smoothness of the consequent membership functions. The output of a probabilistic fuzzy system using Gaussian membership functions will have a smoother probability density function than the equivalent system using triangular membership functions.

We already mentioned that using a fuzzy partition already improves the approximation of the conditional pdf significantly (van den Berg et al., 2004). Similarly, increasing the number of rules will improve the accuracy of interpolation between the rules. Moreover, the danger that the resulting system overfits the (normally noisy) data (Bishop, 1995) should be dealt with as well.

Function approximation

Analysing (3.6) and (3.20), we note that the calculation of $\frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}$ and $z_{1,j}$ can be performed off-line and these sets can be directly replaced by the defuzzified values. Furthermore, the shape of the output fuzzy sets has no influence on the resulting crisp value given by (3.20), since only centroids of these sets are considered. This indicates a high computational efficiency of these models, after identification.

An advantage of a probabilistic fuzzy system is that the outcomes of the individual rules are not restricted to the grid given by the centroids of the output fuzzy sets. This implies that the outputs of a probabilistic fuzzy system as defined in Section 3.2.3 can be fine-tuned without changing the consequent fuzzy sets C_j using the conditional probability parameters $\mathrm{Pr}(C_j|A_q)$. A consequence of this additional degree of freedom is that there are more free parameters, which poses problems in identification. This advantage is lost when using the definition of conditional probability of fuzzy events as defined by Zadeh (1968) and used in Tang et al. (2012), since this definition depends on the location of the antecedent and consequent fuzzy sets.

3.4.3 Conditional probability parameters

In conjunction with defining the number of rules, antecedent and consequent membership functions, it is also necessary to estimate the conditional probabilities in a probabilistic fuzzy system. Let $\mathbf{P} = \{\hat{Pr}(C_1|A_1), \ldots, \hat{Pr}(C_J|A_Q)\}$ be the parameters to be optimized which conform to constraints (3.25) and (3.26). The type of estimation will vary according to the desired objective. For the case of functions approximation the purpose is to minimize the error between the estimated model output and the data, while for the case of density approximation the objective is to consider the estimation of the whole conditional distribution.

Conditional density estimation

Assuming that conditional random variables $y_t \mid x_t$ and $y_k \mid x_k$ are independent for $k \neq t$ with $t = 1, ..., T$, the likelihood of the data can be written as a product of the conditional density of all observed values. The probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$ that maximize the likelihood of the

data set can be obtained by maximizing the non-linear function (Waltman et al., 2005b)

$$\begin{aligned} L(Y \mid X) &= \prod_{t=1}^{T} f(y_t \mid x_t) = \prod_{t=1}^{T} \sum_{q=1}^{Q} \beta_q(x_t) f(y_t \mid A_q) \\ &= \prod_{t=1}^{T} \sum_{q=1}^{Q} \beta_q(x_t) \sum_{j=1}^{J} \hat{Pr}(C_j \mid A_q) \frac{u_{C_j}(y_t)}{\int_{-\infty}^{\infty} u_{C_j}(y_t) dy_t} \end{aligned} \tag{3.65}$$

where $Y = \{y_1, ..., y_T\}$ and $X = \{x_1, ..., x_T\}$ and constraints (3.25)–(3.26) are satisfied. Note that since we are dealing with time series data in the empirical application, we index observations with $t = 1, \ldots, T$ for convenience. Similar to the case of least-mean squares estimation, this function can be maximized using a gradient search algorithm. Since (3.65) is concave the maximization problem is convex. Other search heuristics can be used to estimate the probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$, such as genetic algorithms (Almeida and Kaymak, 2009b).

Function approximation

The objective function in this case is the squared sums of prediction error. The minimization of this function can rely on minimizing the prediction error such that the difference between the real output and the estimated expected output of a probabilistic fuzzy system are minimized. For observations $t = 1, \ldots, T$ the problem can be defined as:

$$f(\mathbf{P}) = \frac{1}{T} \sum_{t=1}^{T} f_t(\mathbf{P}) = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{E}(y_t|x_t) - y_t \right)^2, \tag{3.66}$$

where the expected output $\hat{E}(y_t|x_t)$ of a probabilistic fuzzy system given by (3.20) can be written as:

$$\begin{aligned} \hat{E}(y_t|x_t) &= \sum_{q=1}^{Q} \beta_q(x_t) \sum_{j=1}^{J} Pr(C_j|A_q) z_j \\ &= \sum_{q=1}^{Q} \beta_q(x_t) \sum_{j=1}^{J-1} Pr(C_j|A_q) z_j + \sum_{q=1}^{Q} \beta_q(x_t) Pr(C_J|A_q) z_J \\ &= \sum_{q=1}^{Q} \beta_q(x_t) \left(\sum_{j=1}^{J-1} Pr(C_j|A_q) z_j + \left(1 - \sum_{j=1}^{J-1} Pr(C_j|A_q) \right) z_J \right) \end{aligned} \tag{3.67}$$

such that $\mathbf{P} = \{Pr(C_1|A_1), \ldots, Pr(C_{J-1}|A_Q)\}$ are the parameters to be optimized which conform to constraints (3.25) and (3.26). This is a concave problem as the following theorem shows.

Theorem 3.6 *The mean square error function* $f(\mathbf{P})$ *given by* (3.66) *is convex.*

Proof:

Since the sum of convex functions is convex, to prove the convexity of (3.66), it is sufficient to prove that $f_t(\mathbf{P})$ is a convex function for all t. From (3.66), the first and the second derivatives of $f_t(\mathbf{P})$ for $j = 1, \ldots, J-1$, $j' = 1, \ldots, J-1$ $q = 1, \ldots, Q$, $q' = 1, \ldots, Q$ are:

$$\frac{\partial f_t(\mathbf{P})}{\partial P(C_j|A_q)} = 2e_t\beta_q(x_t)\,(z_j - z_J) \tag{3.68}$$

$$\frac{\partial^2 f_t(\mathbf{P})}{\partial P(C_j|A_q)\partial P(C_{j'}|A_{q'})} = 2\beta_q(x_t)\beta_{q'}(x_t)\,(z_j - z_J)\,(z_{j'} - z_J) \tag{3.69}$$

where $e_t = \hat{E}(y_t|x_t) - y_t$.

For $q = q'$ and $j = j'$, the second derivative (3.69) is non-negative:

$$\frac{\partial^2 f_t(\mathbf{P})}{\partial P(C_j|A_q)^2} = 2\beta_q(x_t)^2\,(z_j - z_J)^2 \geq 0\,. \tag{3.70}$$

Hence, the diagonal elements of the Hessian matrix are non-negative.

Let Δ_M denote the determinant of the minors of the Hessian using the Laplace expansion, given by the $M \times M$ matrix of second derivatives:

$$\Delta_M = \begin{vmatrix} \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_1}|A_{q_1})\partial \hat{P}(C_{j_1}|A_{q_1})} & & \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_1}|A_{q_1})\partial \hat{P}(C_{j_M}|A_{q_M})} \\ & \ddots & \\ \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_M}|A_{q_M})\partial \hat{P}(C_{j_1}|A_{q_1})} & & \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_M}|A_{q_M})\partial \hat{P}(C_{j_M}|A_{q_M})} \end{vmatrix} \tag{3.71}$$

For the case of $M = 2$, using (3.69), the determinant of the 2×2 minor matrix is:

$$\begin{aligned} \Delta_2 &= \begin{vmatrix} \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_1}|A_{q_1})\partial \hat{P}(C_{j_2}|A_{q_2})} & \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_1}|A_{q_1})\partial \hat{P}(C_{j_4}|A_{q_3})} \\ \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_3}|A_{q_4})\partial \hat{P}(C_{j_2}|A_{q_2})} & \frac{\partial^2 f_t(\mathbf{P})}{\partial \hat{P}(C_{j_3}|A_{q_4})\partial \hat{P}(C_{j_4}|A_{q_3})} \end{vmatrix} \\ &= \beta_{q_1}(x_t)\beta_{q_2}(x_t)\,(z_{j_1} - z_J)\,(z_{j_2} - z_J)\,\beta_{q_4}(x_t)\beta_{q_3}(x_t)\,(z_{j_3} - z_J)\,(z_{j_4} - z_J) \\ &\quad -\beta_{q_1}(x_t)\beta_{q_3}(x_t)\,(z_{j_1} - z_J)\,(z_{j_4} - z_J)\,\beta_{q_4}(x_t)\beta_{q_2}(x_t)\,(z_{j_3} - z_J)\,(z_{j_2} - z_J) \\ &= 0\,, \end{aligned}$$

with $q_1, \ldots, q_4 \in \{1, \ldots, Q\}$ such that $q_m \neq q_{m'}$ for $m \neq m'$ and $j_1, \ldots, j_4 \in \{1, \ldots, J\}$ such that $j_m \neq j_{m'}$ for $m \neq m'$.

Applying the Laplace determinant expansion, (3.72) implies also that $\Delta_M = 0$ for any $M \geq 2$. From (3.70) and (3.72) we can conclude that the Hessian matrix is positive semidefinite.

Since $f_t(\mathbf{P})$ is a convex function for all t and the sum of convex functions is convex, the MSE given by (3.66) is therefore also convex.

■

In the nonlinear programming problem of finding estimates of the probability paramenters in a probabilistic fuzzy system, the functions in the constraints, given by (3.25)–(3.30), are linear, from which it follows that these functions are convex. Since the objective function is concave, the nonlinear programming problem is a convex programming problem. Convex programming problems have the convenient property that each local optimum is also a global optimum.

3.5 Examples: Synthetic data parameter estimation

In this section we illustrate the performance of the probabilistic fuzzy system and discuss the estimation issues using a known data generating processes to simulate data. Doing so, allows us to study the approximation capabilities of the probabilistic fuzzy system, *i.e.* perform function approximation and conditional density estimation. It will also serve to show the influence of the different parameters of a probabilistic fuzzy system and estimation procedures, as explained in Section 3.4.

In the following sections, the results displayed are only for one run. This is due to the fact that we are using simple simulated functions, which are kept constant throughout the experiments. Barring numerical problems, which were not detected during the experiments, the most likely source of differing results would be the clustering heuristic for the antecedent space. To solve this issue, all clustering algorithms optimization would stop when the error is less than 10^8 and the staibility of the solutions was checked with multiple runs. The obtained output membership functions and probability parameters solutions are unique as discussed in Section 3.4.3.

3.5.1 Function approximation

In this section, we consider a simulated dataset from a non-linear system. To facilitate visualisation, we choose a system of the form

$$y(x) = 0.01\sin(0.0007x^2)x + \epsilon_x,\ \epsilon_x \sim \mathrm{NID}(0, 0.04)\,, \tag{3.72}$$

with $x \in [0, 100]$. We used a training set $\mathcal{L} = \{(\mathbf{x}_p, y_p)\}$ with a uniform sample of size $P = 1981$. To identify the antecedent membership functions parameters of a PFS we used a fuzzy clustering heuristic as described in Section 3.4.1 on the regression hyperplane of y in $\mathbf{x}$. The obtained antecedent membership functions are of the triangular form given by (3.63). As

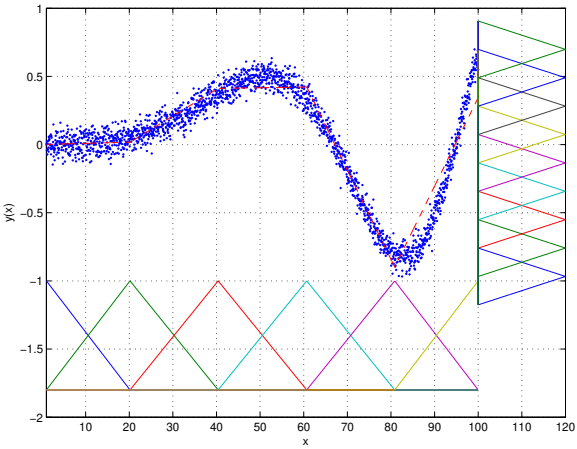


Figure 3.3: Function approximation: Diagram of a PFS with 6 input MF and 9 output MF.

discussed in Section 3.4.2, for function approximation, only the centroid of the output membership functions influences the expected output of a PFS. Thus, we settled on triangular membership functions, uniformly distributed between a minimum and maximum value. The probability parameters are obtained by minimizing the mean square error (3.66), as discussed in Section 3.4.3.

To evaluate the obtained results, we use the mean square error between the estimated crisp output of a PFS $\hat{\eta}_{\underline{y}|\mathbf{x}}$ and the observed y, defined as

$$\text{MSE} = \hat{\eta}_{\underline{y}|\mathbf{x}} - y \,. \tag{3.73}$$

Note that for a large number of observations, the MSE should converge to the variance of the error term ϵ_x since the errors are NID. In the first experiment we are interested in performing an empirical study of the influence of the number of input and output membership function of a PFS on the obtained approximation error. For the output membership function only the number varies. The overlap is the same irrespective of the number of membership functions such that the output space is well formed between the minimum and maximum values of sampled y, satisfying (3.35). The results obtained for a PFS derived with a FCM clustering heuristic are shown in Table 3.1 while the results obtained for a PFS derived with a GK clustering heuristic are shown in Table 3.2.

Table 3.1 and Table 3.2 show that as we increase the number or input membership functions and output membership functions the mean square error decreases. For the considered example, after a certain number of input and output membership functions, *e.g.* $Q = 10, J = 7$, a

Table 3.1: MSE for different numbers of input and output MF using PFS (FCM clustering heuristic) for function approximation of function (3.72).

		Number Output MF (J)										
		3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Number Input MF (Q)	**3**	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098
	4	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103
	5	0.045	0.039	0.036	0.035	0.035	0.035	0.034	0.039	0.038	0.037	0.036
	6	0.020	0.014	0.013	0.012	0.011	0.011	0.010	0.014	0.010	0.013	0.013
	7	0.019	0.012	0.010	0.009	0.008	0.008	0.007	0.007	0.007	0.007	0.010
	8	0.011	0.007	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.006	0.006
	9	0.008	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
	10	0.008	0.005	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	11	0.009	0.006	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	12	0.009	0.006	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	13	0.008	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004

Table 3.2: MSE for different numbers of input and output MF using PFS (GK clustering heuristic) for function approximation of function (3.72).

		Number Output MF (J)										
		3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Number Input MF (Q)	**3**	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
	4	0.020	0.013	0.011	0.010	0.010	0.010	0.010	0.010	0.010	0.013	0.011
	5	0.019	0.013	0.010	0.009	0.008	0.009	0.008	0.008	0.007	0.011	0.010
	6	0.019	0.012	0.010	0.009	0.008	0.008	0.008	0.008	0.008	0.008	0.008
	7	0.012	0.009	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
	8	0.008	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	9	0.009	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	10	0.009	0.005	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	11	0.009	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	12	0.008	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	13	0.008	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004

further increase of the number of input or output membership functions will not lead to a better approximation, but it is possible to encounter identification problems. This was not the case in this simple example. A diagram of a PFS system with 6 triangular input membership functions and 9 triangular output membership functions is shown in Fig. 3.3

For comparison purposes with Tang et al. (2012), we estimate the conditional probability parameters $\hat{P}(C_J|A_Q)$ using the definition of conditional probability of two fuzzy events A and B, given by (Zadeh, 1968)

$$\Pr(A|B) = \frac{\int_{-\infty}^{\infty} u_A(x)u_B(x)f(x)dx}{\int_{-\infty}^{\infty} u_B(x)f(x)dx}\,. \tag{3.74}$$

Table 3.3: MSE for different numbers of input and output MF using PFS (FCM clustering heuristic with $\hat{P}(C_J|A_Q)$ estimated using (3.74)) for function approximation of function (3.72).

		Number Output MF (J)										
		3	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Number Input MF (Q)	**3**	0.130	0.122	0.119	0.118	0.118	0.117	0.117	0.117	0.117	0.117	0.117
	4	0.125	0.116	0.114	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113
	5	0.087	0.073	0.068	0.066	0.064	0.064	0.063	0.063	0.063	0.063	0.063
	6	0.069	0.052	0.047	0.044	0.043	0.042	0.042	0.041	0.041	0.041	0.041
	7	0.060	0.043	0.037	0.035	0.033	0.033	0.032	0.032	0.032	0.032	0.032
	8	0.053	0.035	0.030	0.027	0.026	0.025	0.025	0.025	0.025	0.025	0.025
	9	0.047	0.029	0.024	0.022	0.021	0.020	0.020	0.019	0.019	0.019	0.019
	10	0.043	0.025	0.020	0.018	0.016	0.016	0.016	0.015	0.015	0.015	0.015
	11	0.040	0.022	0.017	0.015	0.014	0.013	0.013	0.013	0.012	0.012	0.012
	12	0.037	0.019	0.015	0.013	0.012	0.011	0.011	0.011	0.010	0.010	0.010
	13	0.035	0.017	0.013	0.011	0.010	0.009	0.009	0.009	0.009	0.009	0.009

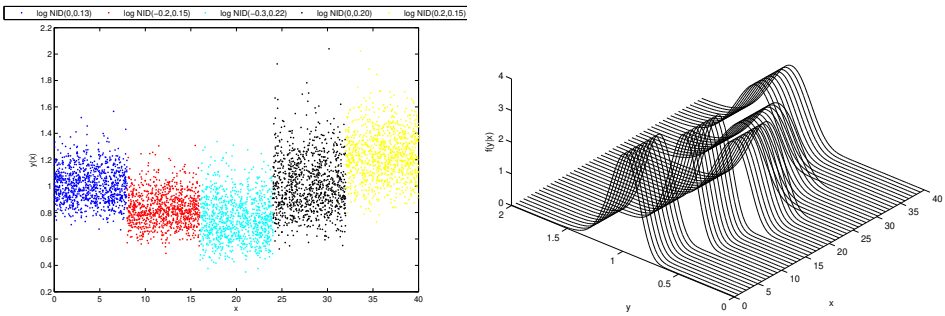
Although the method to obtain the input and output membership function presented in Tang et al. (2012) is different from the present work, we wish to highlight the influence of the estimation of probability parameters $\hat{P}(C_J|A_Q)$. The results obtained using a PFS derived with a FCM clustering heuristic are shown in Table 3.3. We note that the input and output membership functions are exactly the same as the ones used in Table 3.1.

Comparing Table 3.1 and Table 3.3 it is possible to observe that the former are always lower than the latter. This result is not unexpected since (3.74) has been show to be biased and inconsistent (Waltman et al., 2005b). Furthermore, by using (3.74), the PFS is dependent only on the number, location and type of input and output membership functions. A PFS designed using the methods discussed in Section 3.4.3 will also depend on the probability parameters.

For the same range and same number of output fuzzy sets, the amount of overlap between fuzzy sets will influence the output of a PFS. To analyze this effect, we partition the output space using membership functions of the same size, but varying the amount of overlap in terms of percentage of the support of each fuzzy set. We consider a PFS obtained using the FCM clustering heuristic. The results are shown in Table 3.4. As this table shows, for very low or very high percentages of overlap the accuracy of the system decreases, while it has a good approximation in the region of $40\% - 50\%$. One of the advantages of using fuzzy sets is that an observation can belong to more than one set, with a certain degree. By lowering the amount of overlap, the system becomes almost crisp. On the other hand, by increasing the amount of overlap too much, the each observation will belong to several sets at the same time, and the system becomes harder to identify.

Table 3.4: MSE for different amount of overlap between output MF using PFS (FCM clustering heuristic) with Q inputs MF and J output MF.

	% **Fuzzy set support overlap**									
MSE	(Q,J)	1%	5%	10%	20%	30%	40%	50%	60%	70%
	(5,5)	0.050	0.051	0.034	0.034	0.035	0.036	0.036	0.037	0.039
	(5,6)	0.039	0.040	0.040	0.041	0.034	0.035	0.035	0.035	0.036
	(5,8)	0.038	0.038	0.038	0.039	0.040	0.041	0.035	0.035	0.035
	(6,7)	0.015	0.016	0.016	0.017	0.010	0.011	0.011	0.012	0.012
	(8,5)	0.007	0.005	0.005	0.005	0.005	0.006	0.006	0.006	0.007



(a) Data points (b) Density distribution

Figure 3.4: Simulated dataset from a combination of log-normal distribution with sharp changing means and variances, following (3.75).

3.5.2 Conditional density approximation

In this section, we consider a simulated dataset from a combination of log-normal distribution with sharp changing means and variances in different crisp regions. The data are simulated from

$$y(x) = \begin{cases} \log \mathrm{NID}(0, 0.13), & \text{if } x \in [0, 8)\,, \\ \log \mathrm{NID}(-0.2, 0.15), & \text{if } x \in [8, 16)\,, \\ \log \mathrm{NID}(-0.3, 0.22), & \text{if } x \in [16, 24)\,, \\ \log \mathrm{NID}(0, 0.20), & \text{if } x \in [24, 32)\,, \\ \log \mathrm{NID}(0.2, 0.15), & \text{if } x \in [32, 40]\,, \end{cases} \tag{3.75}$$

using a training set $\mathcal{L} = \{(\mathbf{x}_p, y_p)\}$ with a uniform sample of size $P = 5000$. The obtained dataset is asymmetric and has fat-tails and changing mean as Fig. 3.4 shows.

Table 3.5: KLIC for different numbers of input and triangular output MF using PFS (FCM clustering heuristic) for density approximation of function (3.75).

Number Input MF (Q)	Number Output MF (J) 5	6	7	8	9	10	11	12	13
5	0.192	0.118	0.085	0.051	0.033	0.034	0.031	0.026	0.027
	(0.01, 0.97)	(0.02, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.96)	(0.04, 0.95)	(0.05, 0.95)
6	0.216	0.159	0.126	0.100	0.085	0.084	0.083	0.081	0.080
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.03, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)	(0.05, 0.95)
7	0.224	0.158	0.121	0.094	0.080	0.081	0.078	0.076	0.075
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.95)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)	(0.05, 0.95)
8	0.213	0.148	0.115	0.089	0.072	0.071	0.069	0.070	0.068
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.03, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)	(0.05, 0.95)
9	0.205	0.135	0.109	0.077	0.057	0.056	0.054	0.054	0.052
	(0.01, 0.97)	(0.02, 0.97)	(0.02, 0.97)	(0.03, 0.95)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
10	0.193	0.121	0.089	0.056	0.037	0.036	0.036	0.032	0.032
	(0.01, 0.97)	(0.02, 0.97)	(0.02, 0.97)	(0.03, 0.95)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
11	0.203	0.139	0.105	0.077	0.060	0.059	0.059	0.056	0.055
	(0.01, 0.97)	(0.02, 0.97)	(0.02, 0.96)	(0.03, 0.95)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
12	0.201	0.129	0.095	0.065	0.051	0.050	0.049	0.046	0.045
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.03, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
13	0.205	0.135	0.099	0.072	0.054	0.053	0.053	0.051	0.049
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.03, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)

To identify the antecedent membership functions parameters of a PFS we used a fuzzy clustering heuristic as described in Section 3.4.1 using FCM on the regression hyperplane of y in $\mathbf{x}$. The obtained antecedent membership functions are of the Gaussian form given by (3.60).

The Kullback-Leibler Information Criteria, (KLIC), also known as Kullback-Leibler distance, between the real density $f_{\underline{y}|\mathbf{x}}$ and an estimated density $\hat{f}_{\underline{y}|\mathbf{x}}$ is given by (Kullback and Leibler, 1951)

$$\text{KLIC} = E\left(\ln f_{\underline{y}|\mathbf{x}} - \ln \hat{f}_{\underline{y}|\mathbf{x}}\right). \tag{3.76}$$

The KLIC can be consistently estimated by the average Kullback-Leibler distance in the sample (Bao et al., 2007):

$$\text{KLIC} \approx \frac{1}{P}\sum_{t=1}^{P}\left(\ln f_{\underline{y}|\mathbf{x}}(y_t) - \ln \hat{f}_{\underline{y}|\mathbf{x}}(y_t)\right), \tag{3.77}$$

To evaluate the density estimation results, we compare the real density $f_{\underline{y}|\mathbf{x}}$ and an estimated density $\hat{f}_{\underline{y}|\mathbf{x}}$ using KLIC and by comparing the quantiles of the estimated density and the percentage of simulated data points corresponding to each quantile. From the estimated conditional probability distribution $\hat{f}(y|\mathbf{x})$, quantiles $\tau(c)$ can be calculated by solving $c = \int_{-\infty}^{\tau(c)} \hat{f}(y|\mathbf{x})dy$. For a good approximation of the output density, the quantiles of this estimated density should

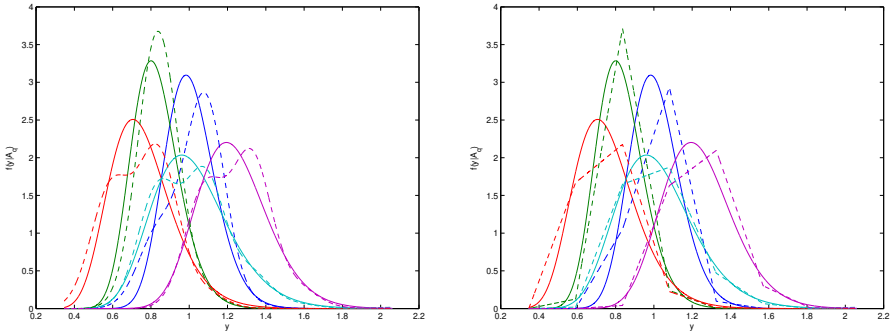
Table 3.6: KLIC for different numbers of input and Gaussian output MF using PFS (FCM clustering heuristic) for density approximation of function (3.75).

Number Input MF (Q)	Number Output MF (J) 5	6	7	8	9	10	11	12	13
5	0.188	0.117	0.086	0.051	0.034	0.032	0.03	0.027	0.027
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.96)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
6	0.212	0.157	0.13	0.099	0.085	0.084	0.082	0.081	0.08
	(0.01, 0.96)	(0.02, 0.97)	(0.04, 0.96)	(0.03, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)	(0.05, 0.95)
7	0.218	0.156	0.123	0.096	0.08	0.079	0.079	0.076	0.075
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
8	0.212	0.146	0.116	0.088	0.072	0.07	0.072	0.068	0.068
	(0.01, 0.96)	(0.02, 0.97)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
9	0.203	0.132	0.109	0.077	0.056	0.054	0.056	0.053	0.051
	(0.01, 0.96)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.95)	(0.05, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
10	0.188	0.115	0.091	0.054	0.036	0.036	0.035	0.031	0.031
	(0.01, 0.97)	(0.02, 0.97)	(0.03, 0.96)	(0.04, 0.95)	(0.05, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
11	0.199	0.136	0.107	0.075	0.06	0.058	0.057	0.055	0.054
	(0.01, 0.96)	(0.03, 0.97)	(0.03, 0.96)	(0.04, 0.95)	(0.05, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)	(0.05, 0.95)
12	0.195	0.123	0.097	0.065	0.05	0.048	0.048	0.044	0.044
	(0.01, 0.96)	(0.02, 0.97)	(0.04, 0.96)	(0.04, 0.95)	(0.05, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)
13	0.197	0.128	0.107	0.069	0.053	0.053	0.053	0.049	0.051
	(0.01, 0.97)	(0.02, 0.97)	(0.04, 0.96)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.04, 0.95)	(0.05, 0.95)

match with the quantiles of the data, *e.g.* 5% of the actual observations should fall in the 5% tail of the output density.

As in the previous section, we analyze the influence of the number of input and output membership function of a PFS on the obtained density approximation error. For the output membership function only the number varies. Following the previous section definition, the overlap is kept at 50% for all experiments, such that the output space is well formed between the minimum and maximum values of sampled y, satisfying (3.35). The KLIC and $\tau(c)$ for $c = 5\%, 95\%$ for a PFS derived with a FCM clustering heuristic and triangular output membership functions are shown in Table 3.5.

Table 3.5 shows that as we increase the number of input membership functions and output membership functions the KLIC decreases and the 5% and 95% quantiles are closer to the real one. The increase of input membership functions allows the system to better track changes in the input space. For the case of $Q = 5$, the results are very good, since the real data have 5 defined regions, while the increase of output membership functions allows the system to better approximate the output densities. The same effect is observed on classical crisp histograms. As the number of 'bins' increases so does the density approximation accuracy. As it was observed in the function approximation examples, it is possible that as this number increases, identification problems surface.



(a) Gaussian output MF (b) Triangular output MF

Figure 3.5: Comparison of fuzzy histograms $\hat{f}(y|A_q)$ for each rule of PFS system with 5 rules.

A PFS as defined in Section 3.2 smoothly changes from a rule output density into another rule density, and in certain regions of the space, combination of densities are possible. Crisp or very abrupt changes are smoothly approximated by the fuzzy input membership functions. Crisp changes can be modelled by replacing the fuzzy input membership functions with crisp ones.

For density approximation, the smoothness of the estimated density output depends directly on the smoothness of the consequent membership functions. For this reason, we compare the results obtained using triangular consequent membership functions with the ones obtained using Gaussian consequent membership functions, as defined in (3.59). For direct comparison with Table 3.5, the location and overlap of both systems are exactly the same. The results are shown in Table 3.6. As this table shows, the results are slightly better than for the case of output triangular membership functions. In Fig. 3.5 we compare each rule fuzzy histogram $\hat{f}(y|A_q)$ of PFS system with 5 rules with the 5 true densities as defined in (3.75). We can observe that the obtained densities using Gaussian output membership functions are smoother than the triangular counterpart. It is also possible to observe that for the obtained fuzzy histograms using Gaussian output membership functions appear to be bi-modal in certain cases. This is an artefact introduced when two or more membership functions have similar probabilities associated with them. This selection problem resembles that of finding adequate basis functions when applying radial basis functions networks (Bishop, 1995) for kernel regression or optimal bin width in kernel density estimators (Bashtannyk and Hyndman, 2001).

3.6 Application on US inflation data

Assessing the changes in prices, measured by inflation levels, is one of the central topics in economic analysis. Most central banks aim to keep inflation levels within a defined range through monetary policy instruments in order to stabilize price movements and to promote economic growth (Galí and Gertler, 1999). In this context, estimating and forecasting the complete inflation density is more adequate than performing point estimation since the former can be used to obtain an estimated range for inflation. Based on these predictions, a central bank can adjust the monetary policy instruments accurately. This interest in estimating and forecasting the complete inflation density has led several institutions, such as the Bank of England and the Norges Bank to report data on inflation density forecasts rather than point forecasts (Diebold et al., 1998).

The data set includes 209 observations for quarterly U.S. inflation over the period 1960 quarter II until 2012 quarter I. Inflation is defined as the growth rate of the implicit Gross Domestic Product (GDP) deflator as in Galí and Gertler (1999). The data over the period between 1960 quarter I and 2001 quarter IV are set as the estimation sample. The remaining data until 2012 quarter I are kept for one quarter ahead forecast evaluations.

An intuitive measure that influences inflation levels is the state of inflation expectations (Bernanke, 2007). People's expectations of future inflation is expected to change their consumption behavior, the overall price level, and hence inflation itself. A conventional measure for people's inflation expectations is the data set published by the University of Michigan Inflation Expectation (MICH) (Thomson Reuters/University of Michigan) which is summarized in Del Negro and Schorfheide (2013). In this survey, individuals are asked by how much they expect the Consumer Price Index (CPI) to change over the next 12 months. Note that the survey data are for CPI inflation expectations. The discrepancy between the CPI and GDP inflation is solved by subtracting the average difference between CPI and GDP inflation from the survey data as in Del Negro and Schorfheide (2013); Baştürk et al. (2013). Furthermore, since the survey data provide monthly four-steps-ahead (one-year) expectations, quarterly values are achieved by using the reported expectations at the beginning of each quarter and then dividing this data by four, assuming constant expectations over the year. Apart from the inflation expectations, we additionally include the inflation level in the last quarter as an antecedent in the PFS model for inflation. Past inflation is often used as an explanatory variable of current inflation since the inflation series is quite persistent (Stock and Watson, 2010).

Despite the growing interest in estimating and forecasting the inflation density, obtaining accurate results for these data is not straightforward since the data show different patterns over time. Inflation volatility changes substantially over time, with a clear decrease after the early 1980s, marking the period of Great Moderation (McConnell and Perez-Quiros, 2000; Stock and

Watson, 2007). Furthermore, during the recent economic crisis, a so-called distinct event of deflation was also observed. It is therefore argued that models for inflation forecasting should account for slowly changing patterns in inflation (Faust and Wright, 2012). Following these observations, assuming a constant inflation level or a constant volatility for inflation may be too restrictive. Hence the proposed PFS model, which can account for such complex time series behavior, is suitable for analyzing inflation.

An important issue in forecasting inflation is the forecast horizon for the inflation density. Relying on quarterly data, one can perform one quarter ahead forecasts for inflation. However, inflation forecasts for longer horizons, such as one year or years ahead inflation, are also of great importance since economic agents do not necessarily make their decisions on a quarterly basis. A model designed for one quarter ahead forecasts may not perform well when the focus is inflation forecasts at longer horizons. Hence the accuracy of the proposed model should also be assessed at longer estimation and forecast horizons.

A further issue in the inflation analysis is the data limitation. Maximum likelihood estimation, for which several analytical properties rely on large sample approximations, may not be appropriate for the considered inflation data. Estimating the parameters under less restrictive assumptions, such as minimizing the mean squared error in the sample, may therefore be more appropriate.

In relation to the aforementioned issues in inflation estimation and forecasting, we first note that the proposed model automatically provides density estimates for inflation together with point estimates. Hence the desired inflation metrics such as the mean, variance and deflation probabilities can be retrieved from the estimation results without additional computational burden. Regarding the estimation method, we consider the results from the proposed model using the maximum likelihood estimation method and also the results obtained by minimizing the mean squared error. We further report multiple horizon inflation estimates for 1 quarter ahead, 4 quarters ahead and 8 quarters ahead estimation. Finally, we show that the complex inflation behavior is well captured with the proposed model when one considers the 95% intervals for inflation both in the estimation sample and in the forecast sample.

The PFS model derived in this section is of the form (3.36). Let y denote inflation and $f_1(y)$, $f_4(y)$ and $f_8(y)$ denote the future 1, 4 and 8 quarters ahead inflation densities, respectively. Such a system is defined by a system of rules R_q, $q = 1, \ldots, Q$ of the form

$$\begin{aligned} R_q : \text{If } \mathrm{Inf}_t \text{ is } A_{q,1} \text{ and } \mathrm{Exp}_{t+1} \text{ is } A_{q,2} \text{ then } & f_1(y) \text{ is } f_1(y|A_{q,1}, A_{q,2}), \\ & f_4(y) \text{ is } f_4(y|A_{q,1}, A_{q,2}), \\ & f_8(y) \text{ is } f_8(y|A_{q,1}, A_{q,2}), \end{aligned}$$

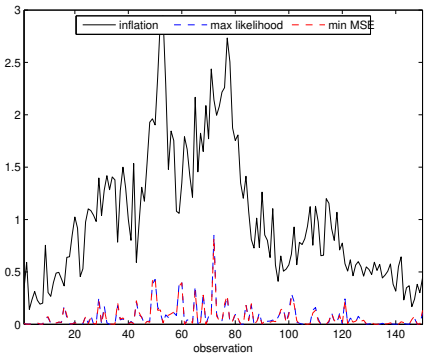


Figure 3.6: Observed inflation and squared errors from minimizing mean squared errors and maximum likelihood estimation.

where Inf_t is the inflation level at time t, and Exp_{t+1} is the inflation expectation for time $t+1$, i.e. a one quarter ahead inflation expectation. For the inflation data, we apply the PFS model in (3.78) with $Q = 6$ rules and 9 output membership functions. The antecedent fuzzy sets are obtained through a clustering heuristic using fuzzy-c means as described in Section 3.4.1 and the fuzzy consequents are obtained by distributing the membership functions uniformly over the universe of discourse.

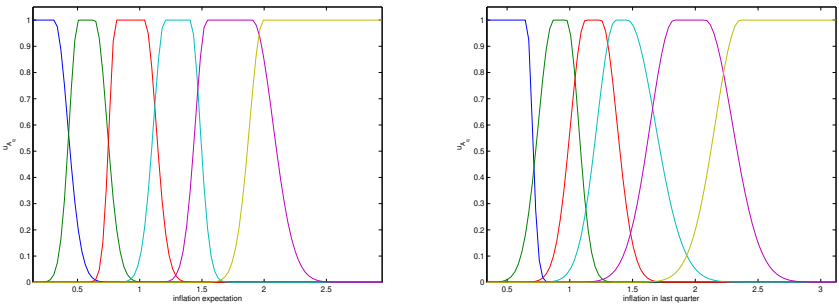
3.6.1 Comparison of different estimation methods for the inflation model

In this subsection we consider two alternative estimation methods for the probability parameters of the proposed model. The first alternative is the maximum likelihood estimation method to obtain parameter estimates, which rely heavily on the exact conditional density of inflation defined by the model. The second alternative is to obtain the parameter estimates by minimizing the mean squared error of the observations, which to a large extent refrains from the distributional assumptions in the model.

Figure 3.6 presents the data and the squared error for each observation in the estimation sample, where the output of PFS is taken as the conditional mean at each period. The two estimation methods do not differ substantially in terms of the mean squared error. Especially at the end of the sample period, squared errors from both estimation methods are low, indicating that the model captures inflation behavior accurately. An exceptional period in terms of the squared errors is the high inflationary period beginning at the end of 1970s and ending in mid-1980s, during which both models perform rather poorly. The relatively poor performance of both models in this period can be explained by the large variation in observed inflation compared

Table 3.7: Optimized probability parameters for inflation from two estimation methods.

	Maximum likelihood estimation										Minimizing the mean squared error								
	Consequents										Consequents								
Rule	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Rule**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	0.46	0.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1**	0.51	0.42	0.07	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.83	0.17	0.00	0.00	0.00	0.00	0.00	0.00	**2**	0.07	0.63	0.26	0.04	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.64	0.36	0.00	0.00	0.00	0.00	0.00	**3**	0.00	0.18	0.45	0.30	0.07	0.00	0.00	0.00	0.00
4	0.00	0.00	0.19	0.59	0.17	0.00	0.04	0.00	0.00	**4**	0.00	0.03	0.25	0.37	0.27	0.06	0.03	0.00	0.00
5	0.00	0.00	0.00	0.12	0.41	0.48	0.00	0.00	0.00	**5**	0.00	0.00	0.01	0.15	0.53	0.25	0.06	0.01	0.00
6	0.00	0.00	0.00	0.00	0.00	0.28	0.48	0.08	0.15	**6**	0.00	0.00	0.00	0.00	0.06	0.27	0.33	0.13	0.21



(a) inflation expectations (b) inflation in last quarter

Figure 3.7: Antecedent membership functions for 1 quarter ahead inflation.

to the remaining periods. The average squared error from minimizing the mean squared error and maximum likelihood estimation are 0.0653 and 0.0662, respectively. The obtained squared error is naturally smaller when the objective function is defined as the mean squared error.

Figure 3.7 shows the antecedent membership functions for inflation in the last quarter and expected inflation, which are common for the mean squared error minimization and maximum likelihood estimation methods. Table 3.7 on the other hand provides the obtained probability parameters for the PFS model for inflation obtained by minimizing the mean squared error and by maximizing the likelihood. The exact probability estimates differ between the two estimation methods. We relate this result to the extra assumptions made in the maximum likelihood estimation method and acknowledge that these assumptions may be restrictive given the small number of data points in this study. For this reason, results reported in the remaining analysis are based on minimizing the mean squared error rather than the maximum likelihood estimation method.

Table 3.8: Properties for the output density from each rule from MSE minimization for 1 quarter ahead inflation.

	Rule					
	1	**2**	**3**	**4**	**5**	**6**
mean	0.463	0.651	1.020	1.389	1.671	1.591
variance	0.105	0.192	0.300	0.387	0.318	0.289
skewness	0.381	1.788	1.110	0.573	-0.245	0.577
kurtosis	2.789	8.193	4.972	3.080	2.215	3.044

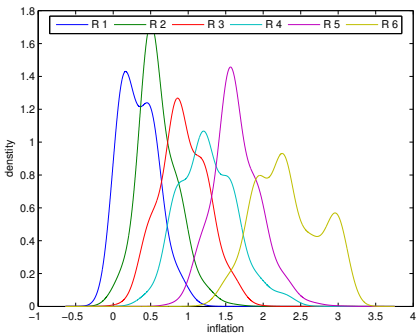


Figure 3.8: Density of rule outputs from MSE minimization for 1 quarter ahead inflation.

3.6.2 Inflation patterns according to the rule outputs

Table 3.8 presents the properties of the output density from each PFS rule, where we focus on the mean, variance, skewness and kurtosis of inflation. Obtained densities from each rule are shown in Fig. 3.8 in detail. Rule 1 and rule 2 clearly define low levels of inflation with relatively low variance compared to the remaining rules. These rules may capture the Great Moderation period where inflation levels and volatility are substantially lower compared to the remaining periods. These two rules may also capture deflationary pressures in the economy since inflation levels below point 0 have a positive probability mass according to these rules. In contrast, rule 5 and rule 6 clearly define a high level of inflation accompanied by high volatility in inflation. These rules may capture the high inflation and high volatility periods during 1980s. The combination of the rule outputs, on the other hand, may identify the transition between these periods with clear differences in inflation patterns.

Besides the observed variation of inflation levels and variances obtained from each rule, the skewness and the kurtosis of the obtained densities also differ substantially across rules. Most

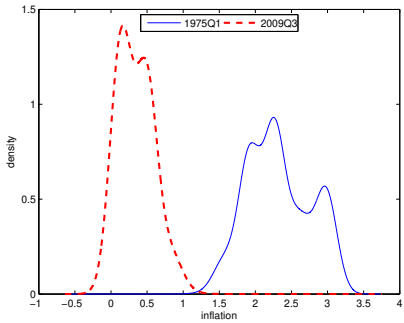
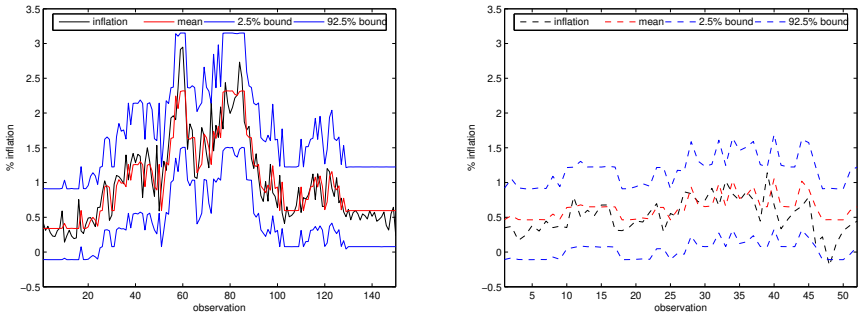


Figure 3.9: Conditional density of inflation for 1975 quarter I and 2009 quarter III from MSE minimization.

importantly, the skewness values are different from zero in all rules, and the kurtosis values are relatively far from 3, the normal density kurtosis value. Hence, assuming a normal distribution for inflation may be too restrictive. This problem is also considered in Ascari et al. (2012); Cúrdia et al. (2013) and the use of a student-t distribution instead of the conventional normal distribution is advocated. The obtained densities in Fig. 3.8, however, show that the inflation density is bimodal for some rules, particularly for rule 6. Furthermore, the combination of the 6 rules may lead to multiple modes in the inflation density even if the individual rule outputs provide uni-modal densities. Hence the student-t density with a single mode is still restrictive according to our results.

An important result in terms of inflation levels is the persistence in inflation, which can be assessed using the probabilities for each rule output on the right panel in Table 3.7. We find a positive relation between past inflation, inflation expectations and current inflation. Low values of past inflation and expected inflation, represented in rule 1 and rule 2 of the PFS are likely to lead to low inflation values since consequents 1, 2 and 3 are found to have high probability values. Similarly, high values of past inflation and expected inflation, represented by rule 5 and rule 6, are likely to lead to relatively high inflation values, captured in consequents 7, 8 and 9. Note that 'moderate' inflation levels, in consequents 4, 5 and 6, have a positive average probability for rules 5 and 6 compared to the zero average probability for rules 1 and 2. This difference in probability values show that inflation values are less persistent if past inflation and inflation expectations are high compared to the periods with a low level of past inflation and inflation expectations.



(a) estimation sample (b) forecast sample

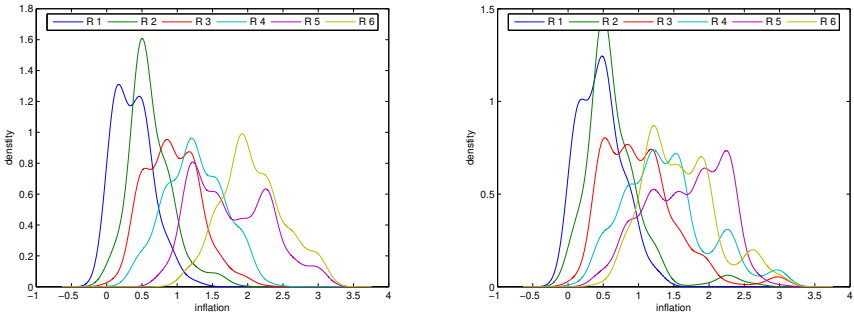
Figure 3.10: Mean and 95% interval for 1 quarter ahead inflation from MSE minimization for the estimation sample and the forecast sample.

3.6.3 Density estimates of inflation

Density estimates of inflation are obtained as a combination of the densities obtained from each rule output in the previous section. As noted earlier, even if the density obtained from each rule output is a well-behaved density, the combined density may have complex features such as non-zero skewness, fat tails or multiple modes. Two examples of such complex conditional densities obtained for inflation are presented in Fig. 3.9 for an observation in the estimation sample, inflation at 1975 quarter 1 and an observation in the forecast sample, inflation at 2009 quarter III. The obtained conditional densities are bimodal for both periods. Inflation density at 1975 quarter I clearly shows positive skewness as well.

Figure 3.10 presents the mean inflation levels and 95% interval estimates obtained from the PFS model for the estimation sample, for 1 period ahead inflation values. Mean inflation values obtained from the model track the observed inflation levels nicely. This result holds both for the estimation sample and the forecast sample. More importantly, the obtained density estimates are quite accurate. For the high inflationary period in 1980s, the interval estimates of inflation are only slightly wider than the remaining periods. Hence this abrupt change in the inflation pattern is captured accurately by the model. A similar result holds when we specifically consider the estimated inflation intervals for the forecast period. Inflation levels are again captured nicely, with a single observation outside the 95% interval. The estimated interval is not very wide, hence the interval estimates are not too conservative.

A unique aspect of this data is the occurrence of deflation during the recent crisis, on 2009 quarter 3. The estimated inflation intervals capture this possibility of deflation, since the 95% interval contains point 0, although the exact inflation value at this quarter is outside the esti-



(a) 4 quarter ahead inflation (b) 8 quarter ahead inflation

Figure 3.11: Density of rule outputs from MSE minimization for 4 and 8 quarter ahead inflation.

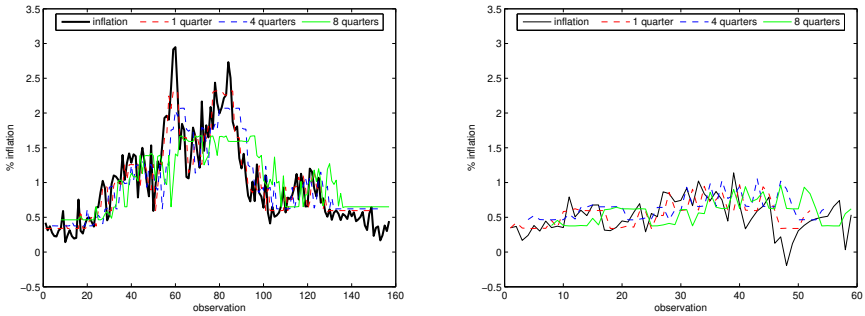
mated interval. Furthermore, the model signals deflationary pressure in the economy in the periods before deflation was actually observed. These periods where estimated inflation intervals included point 0 are the end of the training sample and at the beginning of the forecast period and they cover the recent crisis period.

Note that the obtained 95% inflation interval includes point 0 several observations in the forecast sample, while actual disinflation occurred very rarely during the considered period. Therefore the obtained inflation density may be considered 'too wide'. It is possible to overcome such lack of precision in inflation forecasts by including subjective experts' forecasts as mentioned in Faust and Wright (2012), but this topic is left for future research.

3.6.4 Inflation density estimates for multiple time periods

In this subsection we summarize the density estimates for 1 quarter ahead, 4 quarters (1 year) ahead and 8 quarters (2 years) ahead inflation. We first note that the proposed model is capable of incorporating these multiple period estimations or forecasts in a single model. Furthermore, the model does not require explicit assumptions for the individual distributions of inflation in different quarters, once past inflation and expected inflation are taken into account.

Conditional inflation densities obtained from each PFS rule are shown in Fig. 3.11 for 4 and 8 quarter ahead inflation levels. The density estimates are highly asymmetric and non-standard according to these results, especially compared to the conditional densities for 1 quarter ahead inflation shown in Fig. 3.8. This result is intuitive since the higher horizon inflation estimation contains more ambiguity compared to the 1 quarter ahead estimates even if past inflation and expected inflation are taken into account.



(a) estimation sample (b) forecast sample

Figure 3.12: Mean and 95% interval for multiple horizon forecasts from MSE minimization for the estimation sample and the forecast sample.

We further analyze the mean inflation levels at multiple horizons according to the PFS model where parameter estimates are obtained by minimizing the mean squared error. Figure 3.12 presents these inflation estimates for 1, 4 and 8 quarters ahead inflation for the estimation sample and for the forecast sample. Typically, mean inflation estimates are smoother for longer time horizons. Hence, the longer time horizon inflation estimates provide a smooth long-run inflation information rather than signalling sudden changes in inflation. This result follows from the information contained in the input variables. Given past inflation and inflation expectations, sudden changes in the long run, for example after 8 quarters, cannot be captured accurately. Despite this property of long run inflation estimates, the overall inflation levels follow the smooth changes in inflation patterns accurately.

3.7 Conclusions

In this work we present an analysis of the different aspects of probabilistic fuzzy systems in the context of function approximation and conditional density estimation. We analyse the relation of PFS with different types of systems with deterministic output that have universal approximation capability. We show that PFS is particularly similar to a Mamdami fuzzy system with weighted output or a fuzzy relational model. Hence PFS is suitable for problems involving function approximation.

Function approximation capabilities of PFS and quantitative measures of the shape of the obtained density, such as moments, have not been analyzed in detail in the literature. In this work we show that higher moments, such as skewness and kurtosis, of the conditional probabil-

ity density of the output can be derived from the PFS. Furthermore, we analyze the necessary conditions for a PFS, such that the estimated output density is a proper pdf and subsequent higher moments of this density exist. These conditions relax the previous assumption of a well-formed output space. They are not very restrictive and are not limited by a particular definition of a conditional probability of fuzzy systems. Obtaining these quantitative measures, such as higher moments, of the obtained density is particularly important in applications of PFS in risk management since traditional measures of risk, such as the variance, do not fully capture the distributions of most financial or macroeconomic data.

The performance of PFS in function approximation and conditional density estimation is illustrated using simulated data and real data on quarterly US inflation. Using simulated data, we show that PFS provides accurate density approximations and conditional density estimates in general. Furthermore, we analyse the influence of the PFS parameters, namely the number of input and output memberships, the choice of the PFS membership functions and the estimation method for the conditional probability parameters, on the performance and accuracy. Our application on the US inflation data shows that slowly changing patterns in inflation are accurately captured by the PFS model. The PFS model performs well in one period ahead and 1 year ahead forecasts of inflation. The model is also successful in capturing the deflationary pressure during the recent crisis.

Chapter 4

Probabilistic fuzzy systems in Value-at-Risk estimation[1]

4.1 Introduction

DUE to the volatile nature of the financial markets, risk management is an important activity for financial institutions that operate in these markets. As a result of risk management, activities are undertaken to reduce the possibility of failure to an acceptable range. These activities may include portfolio adjustment, hedging or insurance (Brealey and Myers, 2001; Hull, 2000). Nowadays, the financial sector operates under strict guidelines, which have been imposed through international agreements, partly due to various financial failures that have happened in 1990's. For example, due to the Basel Agreement, the financial institutes must have well documented procedures to manage the different kinds of risks that they are exposed to, such as the market risk, the credit risk and the operational risk (Jorion, 2006).

Managing risk is strongly dependent on the information available. When the amount of information grows beyond a specific level, there is a need for a concise representation of the risk a company or institution is facing. Due to the complex nature of financial markets in which many parties exchange information and interact through trading, the overall risk for a company is influenced by many internal and external factors. Nevertheless, it is customary for management to classify different types of risk and develop models for dealing with each type of risk in order to keep the risk management problem tractable. One of the different types of risk that a financial institution has to deal with is the market risk, which is the exposure to the uncertain market value of a portfolio (Holton, 2003). Value-at-risk (VaR) is a way to quantify the market risk. It is a single number for the senior management to express and summarise the

[1] Parts of this chapter have been published in Almeida and Kaymak (2009a,b).

total market risk of a portfolio with financial assets. Value at Risk measures the worst expected loss over a given horizon under normal market conditions at a given confidence level. Due to regulations, large banks must nowadays base their market risk capital requirements on the VaR estimate (Jorion, 2006). This drives the continued research into newer and better VaR models.

Simulation approaches or parametric approaches are usually used for VaR estimation. The simulation approach makes assumptions about the distribution of portfolio returns, and then applies Monte Carlo simulation to estimate the VaR. Because a large number of simulations is needed, this approach is very costly in terms of computational time. Furthermore, the quality of the results depends on the validity of the assumptions regarding the portfolio returns. In the parametric approaches, the risk is quantified in terms of volatility, which is expressed as the standard deviation σ of the portfolio. Normally, one measures the daily volatility that is estimated from historical data. In order to estimate the VaR for a given horizon, the daily volatility is scaled to multiple-day volatility (Hull, 2000). The simplest models of volatility assume that it does not vary over time. More advanced models acknowledge that volatility varies dynamically over time. The dynamic aspect of volatility could be modelled in various ways. For example, a multivariate switching regime approach to VaR estimation has been discussed in Billio and Pelizzon (2000). Another model where volatility changes dynamically in time is the GARCH (Generalised Auto Regressive Heteroscedasticity) model (Bollerslev, 1986). For the GARCH $(1, 1)$ model, which is used quite often in practice, the variance is estimated using a first-order autoregressive model of the squared returns.

The disadvantage of the parametric approach is that, due to the complexity of financial markets, the data usually do not follow the parametric distributions that are assumed for the data generating process. For example, the returns are typically non-Gaussian, they have fat tails and volatility clustering is often observed in financial markets (Cont, 2007, e.g.). Therefore, flexible modelling approaches such as non-parametric modelling or semi-parametric modelling are needed in which the models can adapt themselves into the underlying actual data distribution. In this context, neural network models for VaR estimation have been studied by various researchers (Taylor, 2000; Chapados and Bengio, 2001; Jiang et al., 2004, e.g.) as well as fuzzy set models (Zmeškal, 2005a; Cherubini and Della Lunga, 2001, e.g.).

A semi-parametric model in which the model structure and the model parameters can be adapted to the underlying data distribution is the probabilistic fuzzy system (PFS). In a probabilistic fuzzy system, a linguistic description of the system behaviour encoded by the fuzzy rules is combined with the statistical properties of data. The probabilistic fuzzy model allows the combination of both linguistic uncertainty and probabilistic uncertainty in the model. PFS are suitable for estimating probability distributions. Since accurate VaR estimation is enabled by estimating the probability distribution of the data, PFS could be used to estimate it.

In this chapter, we consider two approaches to designing a PFS and compare their performance in obtaining value-at-risk models. In the first approach, a Mamdani-type probabilistic fuzzy system (Kaymak et al., 2003) is used for estimating the VaR. The model parameters are obtained by a sequential approach in which the location of the antecedent membership functions is determined by using fuzzy clustering and maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The output membership functions are scaled by using a single scaling parameter. In the second approach, an alternative representation of a PFS as a fuzzy histogram is considered. This is an example of a conditional volatility model in which the future volatility (and hence the associated risk of the portfolio) is estimated by using a distribution function that is represented as a fuzzy histogram. In this case, the membership functions are fixed according to the mental model of the modeler, *i.e.* they are obtained from the modeler. The conditional probability parameters of the model are then estimated by minimising the test statistic of a back testing method by using a constrained evolutionary optimisation algorithm.

The proposed methodologies are applied to estimate the one-day VaR for six different stocks. The validity of obtained VaR models are evaluated by using the Kupiec test based on failure rates and compared to the performance of the GARCH models for VaR estimation. It is found that the statistical back testing always accepts PFS models after tuning, while GARCH models may be rejected.

The outline of this chapter is as follows. In Section 4.2, we give a brief introduction to VaR modelling and VaR models. We discuss the basics of probabilistic fuzzy systems and the concept of fuzzy histograms in Section 4.3. In Section 4.4, we introduce VaR modelling by using probabilistic fuzzy systems. The experimental setup for the empirical study using six different assets are given in Section 4.5, while the results are reported in Section 4.6. Finally, conclusions are given in Section 4.7.

4.2 Value-at-Risk models

Value-at-risk (VaR) is a single number for the senior management to express and summarise the market risk of a portfolio of financial assets. The VaR value of a portfolio is always calculated over a time horizon h at a significance level c. It indicates the maximum loss that a portfolio of assets will suffer over a horizon of h (days) with a confidence of c under normal market conditions. An overview of the mainstream value at risk estimation methods can be found in Duffie and Pan (1997). Several methods are also discussed in Wiener (1999). Various building blocks of VaR measurement, methods for model validation as well as the differences between the parametric and nonparametric estimation approaches are discussed in Jorion (2006).

4.2.1 Value-at-Risk

Assume that a portfolio has value W_t at time t. Let r denote the one period percentage return of the portfolio. If $f(r)$ is the probability density function of the returns, define r_v such that

$$1 - c = \int_{-\infty}^{r_v} f(r)dr. \tag{4.1}$$

The value at risk V_t of the portfolio at time t is then defined as

$$V_t = -r_v W_t. \tag{4.2}$$

Assuming that the returns are distributed normally, the key step in the value at risk estimation can be formulated as determining the variance σ^2 of the returns distribution. This is also called *volatility estimation*.

4.2.2 Volatility estimation

The simplest models of volatility assume that it does not vary over time. In that case, the variance could be estimated by using the observations until period t as

$$\sigma_t^2 = \frac{\sum_{i=1}^{T^*} r_{t-i}^2}{T^*}. \tag{4.3}$$

In (4.3), σ_t^2 denotes that the variance is re-estimated at every period and that there are T^* observations until period t. Usually, however, the standard deviation of the returns varies over time. For example, volatility clustering has been observed in the financial markets, which means that there are periods of high variability followed by low variability. GARCH models are used to capture the time varying behaviour of volatility. The general GARCH (p, p') model calculates the variance from the most recent p observations of returns and the most recent p' estimates of the variance rate. The most popular GARCH model used in practice is the GARCH $(1, 1)$ model in which the variance at period $t + 1$ depends on the variance and the realised returns at period t. It is assumed that the returns r_t at each period t are normally distributed with the same mean, but different variance (local volatility). At each period, the *local volatility* σ_t is assumed to move around the constant *global volatility* $\overline{\sigma}$, so in the long run, a GARCH model recognises that the local volatility reverts to the overall mean value. This property is known as 'mean reversion'. Each period, the local volatility estimate is updated by using

$$\sigma_{t+1}^2 = \gamma\overline{\sigma}^2 + \alpha r_t^2 + \beta\sigma_t^2, \tag{4.4}$$

where α, β and γ are positive constants that satisfy $\alpha + \beta + \gamma = 1$. The optimal values of these parameters can be determined from a data set by using maximum likelihood estimation.

4.2.3 Fuzzy VaR models

Usually, the assumptions of parametric models are not satisfied by real data. Therefore, semi-parametric models such as fuzzy models have been proposed to adapt the VaR estimation to the characteristics of the underlying data generation process. Fuzzy models have the additional benefit that they can be used to deal with non-probabilistic forms of uncertainty, such as linguistic uncertainty and vagueness.

In Zmeškal (2005a), a fuzzy stochastic approach is proposed to model value-at-risk. In this approach, the inputs to the VaR model are described as fuzzy sets. The computations of the model are done by representing the fuzzy sets as a collection of their α-cuts and propagating the fuzziness through the model. Eventually, a fuzzy VaR value is obtained, which the decision maker can use to assess the influence of non-probabilistic uncertainty on his/her decisions. An application of the same methodology for index portfolios is discussed in Zmeškal (2005b) and it is shown that this approach can be interpreted as a generalised sensitivity approach. Another approach to fuzzy VaR modelling has been proposed in Cherubini and Della Lunga (2001), where the authors use a fuzzy measure model for pricing options. In this way, they are able to deal with the cases where the distribution of the underlying asset is not known precisely, and they can account for changes in market liquidity. The authors apply their method to an option-based model of VaR and compute different VaR figures for long and short positions. In Bowden (2006) a different approach to option VaR modelling based on fuzzy set theory is described.

The fuzzy modelling approach proposed in this chapter differs from the above approaches in that we use a fuzzy system to explicitly approximate a probability density function. Hence, the output of the system is essentially a conventional distribution function. However, the working of the system can be described linguistically as a set of probabilistic fuzzy rules and it can be adapted to the linguistic framework of the modeler. Hence, the proposed model links the linguistic categories, which the modeler may define, to the numeric distributions that it estimates.

4.2.4 Model validation

Model validation for value-at-risk is the process of checking whether a VaR model performs adequately, and can be done in various ways. One method is statistical back testing. Back testing verifies within a statistical framework whether the projected losses are in line with the actual losses (Hull, 2000; Jorion, 2006). This entails comparing systematically the history of

VaR forecasts with the corresponding portfolio returns. For VaR users and risk managers, these checks are essential to examine whether their model is well calibrated.

In this chapter, we consider *exception based back testing*. In VaR modelling, an *exception* is said to occur when the actual loss in a period exceeds the VaR that the model predicts. In exception based back testing, the number of exceptions in a given sample is determined and it is tested statistically whether this number is within the range indicated by statistical tests, given a certain confidence interval. With too many exceptions, the model underestimates the volatility. With too few exceptions, the model is too conservative.

Kupiec (1995) has developed a statistical test for assessing the validity of a VaR model. Kupiec confidence regions are defined by the tail point of the log-likelihood ratio LR_{uc}

$$LR_c = -2\ln\left[c^{T-N}(1-c)^N\right] + 2\ln\left\{\left[1-\left(\frac{N}{T}\right)\right]^{T-N}\left(\frac{N}{T}\right)^N\right\}. \tag{4.5}$$

In (4.5), N is the number of exceptions and T is the total number of observations. This ratio is shown to be asymptotically χ^2-distributed, with 1 degree of freedom, under the null hypothesis that the VaR model is valid (Kupiec, 1995). Note that the Kupiec test statistic is two sided. Hence, the model is rejected both when there are too few exceptions as well as when there are too many exceptions. In this chapter, we apply the Kupiec test with 95%, 97.5% and 99% confidence to assess the validity of the VaR models.

4.3 Probabilistic fuzzy systems

A *probabilistic fuzzy system* (PFS) consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions. In this study, we consider Mamdani PFS in which the rules have the following form (Kaymak et al., 2003).

$$\begin{aligned}\text{Rule } R_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then}\\ & y \text{ is } C_{q1} \text{ with } \Pr(C_{q1}|A_q) \text{ and}\\ & y \text{ is } C_{q2} \text{ with } \Pr(C_{q2}|A_q) \text{ and} \ldots \text{ and}\\ & y \text{ is } C_{qN} \text{ with } \Pr(C_{qN}|A_q).\end{aligned} \tag{4.6}$$

Hence, a Mamdani PFS is a generalisation of a Mamdani fuzzy system in which the deterministic fuzzy rules are replaced with probabilistic fuzzy rules. These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The interpretation of the probabilistic fuzzy rules is as follows. Given the occurrence of a (multidimensional)

antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events C_j is likely to occur with probability $\Pr(C_j|A_q)$, $j = 1, 2, \ldots, N$. This applies for all the rules R_q, $q = 1, 2, \ldots, Q$. Note that two conditional probabilities $\Pr(C_j|A_q)$ and $\Pr(C_j|A_{q'})$ will be different, in general.

Let

$$\beta_q(\mathbf{x}) = \frac{u_{A_q}(\mathbf{x})}{\sum_{q'=1}^{Q} u_{A_{q'}}(\mathbf{x})} \tag{4.7}$$

be the normalised degree of fulfillment of rule R_q, where u_{A_q} is the degree of fulfillment of rule R_q. When $\mathbf{x}$ is n-dimensional, u_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.*,

$$u_{A_q}(\mathbf{x}) = u_{A_{q_1}}(x_1) \circ \cdots \circ u_{A_{q_n}}(x_n), \tag{4.8}$$

where x_n is the n-th components of $\mathbf{x}$ and $\circ$ denotes a t-norm. Then, it can be shown that the output of the above Mamdani PFS is a conditional probability density function if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents (van den Berg et al., 2004). The conditional probability of the output given an input vector $\mathbf{x}$ can be computed as

$$f(y|\mathbf{x}) = \sum_{j=1}^{N} \frac{\sum_{q=1}^{Q} \beta_q(\mathbf{x}) \Pr(C_j|A_q) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}, \tag{4.9}$$

assuming that the output space is well-formed, *i.e.* the output membership values satisfy

$$\sum_{j=1}^{N} u_{C_j}(y) = 1, \qquad \forall y \in Y. \tag{4.10}$$

It is also possible to compute the crisp output of the probabilistic fuzzy system by taking the conditional expectation of the output according to

$$E(y|\mathbf{x}) = \int_{-\infty}^{\infty} y f(y|\mathbf{x}) dy. \tag{4.11}$$

However, we do not consider the expected output of the system in this chapter, as we are primarily interested in the conditional distribution of the returns for computing the VaR value of a portfolio.

Assuming that the membership functions in the rule antecedents and the rule consequents have been defined, the optimal probability parameters $\Pr(C_j|A_q)$ can now be determined by

using maximum likelihood parameter estimation, in which the log-likelihood function

$$J = \sum_{k=1}^{K} \ln\left(\Pr(y_k|\mathbf{x}_k)\right) \tag{4.12}$$

is maximised where K is the number of samples in the data set (Waltman et al., 2005b). In (4.12), it is assumed that the samples in the data set are independent of one another. A suitable initialisation for iterative optimisation for maximum likelihood estimation is given by direct estimation from the data by using

$$\Pr(C_j|A_q) = \frac{\sum_{k=1}^{K} u_{C_j}(y_k) u_{A_q}(\mathbf{x}_k)}{\sum_{k=1}^{K} u_{A_q}(\mathbf{x}_k)}. \tag{4.13}$$

Note that the output of a PFS by using (4.9) can also be interpreted as a fuzzy histogram. The technique for estimating a probability density function (pdf) using (crisp) histograms is well-known. By appropriately partitioning the domain of the sample space Y into a set of N disjunct classes C_j, each "column" $f_j(y), (j = 1, 2, \ldots, N)$ of the histogram is defined by the functions

$$f_j(y) = \begin{cases} \frac{\Pr(C_j)}{c_j} & \text{if } y \in C_j \\ 0 & \text{if } y \notin C_j, \end{cases} \tag{4.14}$$

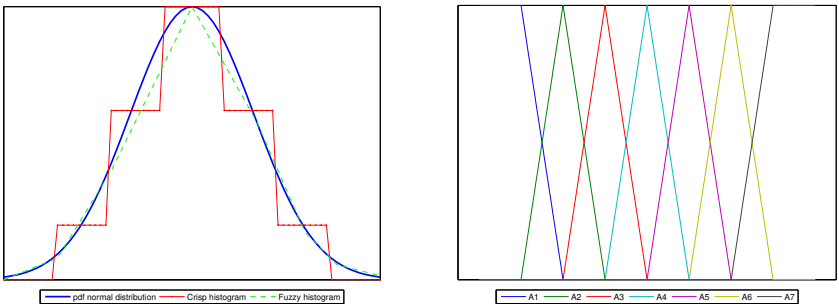
where the probability $\Pr(C_j)$ is estimated in the usual way (using the relative frequency of samples $y_k \in C_j$) and where the scaling scalar c_j equals the size of class C_j (which in the one-dimensional case, is equal to the length of the interval C_j). The probability density function $f(y)$ is approximated by a summation of the functions $f_j(y)$ according to

$$f(y) \approx f_{app}(y) = \sum_{j=1}^{N} f_j(y). \tag{4.15}$$

Probability density functions defined on a sample space Y that is fuzzily partitioned can also be estimated, this time by using a *fuzzy histogram*. To do so, we need a generalisation of the above-given crisp approach. Let Y be fuzzily partitioned in a set of N fuzzy classes C_j described by membership functions $u_{C_j}(y)$, then the (fuzzy) column $f_j(y)$ for fuzzy class C_j can be estimated according to

$$f_j(y) = \frac{\Pr(C_j) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}, \tag{4.16}$$

Equation (4.16) is a generalised version of (4.14). The numerator in (4.16) describes a probability weighted with membership function $u_{C_j}(y)$. The denominator of (4.16) is a scaling factor representing the fuzzified size of class C_j (which in the one-dimensional continuous case,



(a) Fuzzy histogram (b) Triangular membership functions

Figure 4.1: A fuzzy histogram approximates a pdf better than a crisp histogram (a), due to the overlapping membership functions (b).

equals the *fuzzy length* of the interval C_q). The complete pdf $f(y)$ is again approximated by a summation of the functions $f_j(y)$:

$$f(y) \approx f_{app}(y) = \sum_{j=1}^{N} f_j(y) = \sum_{j=1}^{N} \frac{\Pr(C_j) u_{C_j}(x)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}. \tag{4.17}$$

Due to the overlap of the fuzzy sets, fuzzy histograms approximate probability distributions better, in practice. In Figure 4.1 a representation of this phenomenon is shown, where a normal probability density function is approximated using both a crisp and a fuzzy histogram. In both cases, seven classes have been used.

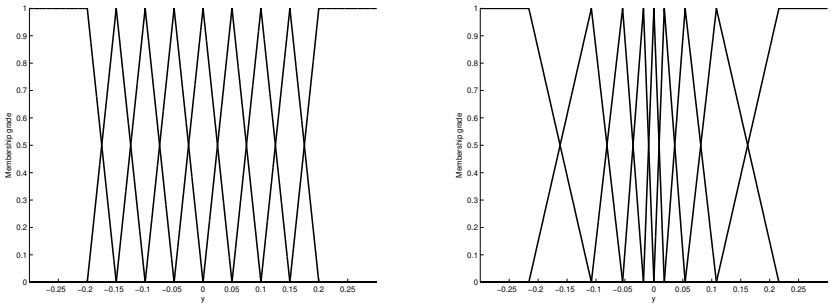
Note that (4.17) guarantees that the approximation $f_{app}(x)$ is properly defined in the sense that

$$\int_{-\infty}^{\infty} f_{app}(y) dy = 1. \tag{4.18}$$

The proof of this observation is obtained by using (4.17), so that

$$\begin{aligned} \int_{-\infty}^{\infty} f_{app}(y) dy &= \int_{-\infty}^{\infty} \sum_{j=1}^{N} \frac{\Pr(C_j) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y)} dy \\ &= \sum_{j=1}^{N} \Pr(C_j) \frac{\int_{-\infty}^{\infty} u_{C_j}(y) dy}{\int_{-\infty}^{\infty} u_{C_j}(y) dy} = \sum_{j=1}^{N} \Pr(C_j) = 1. \end{aligned} \tag{4.19}$$

By setting the multiplier in (4.17) as $\sum_{q=1}^{Q} \beta_q(\mathbf{x}) \Pr(C_j|A_q)$, (4.9) is obtained.



(a) Equally spaced. (b) More sets towards the origin.

Figure 4.2: Two possible partitions for the output space.

4.4 Probabilistic fuzzy models of value at risk

In this chapter, we consider probabilistic fuzzy systems that estimate probability distribution of returns given input data and fuzzy rules that describe the general system behaviour. The models estimate one-day ahead VaR of a portfolio. The same methodology could be applied for multiple-day estimates of VaR, too. The probabilistic fuzzy models that we consider use the returns r_t at period t to predict the distribution of the returns at period $t + 1$.

The model parameters are determined by two different approaches. For convenience we will name them PFS_1 and PFS_2. In the following sections we explain in detail how the parameters for both models are obtained.

4.4.1 Mamdani PFS

PFS_1 is a Mamdani probabilistic fuzzy system in which a data-driven sequential approach is used for determining the model parameters. The input and the output spaces are partitioned into nine fuzzy regions each. This implies that there are nine fuzzy rules in the model. The type, distribution and location of the membership functions can be determined in various ways. For example, it is possible to use triangular or Gaussian membership functions only, or combine them with shouldered membership functions at the edges of the domain. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying. Since the output membership functions must satisfy (4.10), it is convenient to use triangular membership functions for the output partition. Two possible distributions for the output membership functions in that case are shown in Fig. 4.2.

If necessary, the triangular membership functions are combined with shouldered membership functions at the edges of the domain, as shown in Fig. 4.2, to ensure that the domain is always covered by the fuzzy partition, no matter how extreme the returns may be on a particular day. Furthermore, financial returns data are usually distributed in such a way that there are many observations around the origin since usually the returns are either slightly positive or slightly negative. In order to capture the variability in the region with a large amount of data, more membership functions are placed around the origin. As one moves towards the edges of the universe of discourse, the membership functions become wider and the separation between them increases, as shown in Fig. 4.2(b). This is the partition that has been used in PFS_1. Note that the triangular membership functions in this partition are not symmetric.

The range in which the returns for different assets and different portfolios could vary differ a lot from one asset to the other. Hence, it is usually not possible to find a single partition that could be used for different data sets. In order to deal with this situation, we have introduced a scaling parameter z with which the parameters of the output membership functions are multiplied in order to adapt the distribution of the membership functions according to the properties of a given data set.

The input membership functions of PFS_1 are Gaussian. They are determined by using a fuzzy clustering heuristic. Given a pre-determined number of clusters Q, fuzzy c-means algorithm (Bezdek, 1981) is applied in the product space of the antecedent variables. Given the data $\mathbf{x}_k$, $k = 1, \ldots, K$, fuzzy c-means algorithm divides it into Q fuzzy groups by minimising the objective function

$$J(X; U, V) = \sum_{q=1}^{Q} \sum_{k=1}^{K} (u_{qk})^m \{\|\mathbf{x}_k - \mathbf{v}_q\|^2\}, \tag{4.20}$$

subject to the constraint

$$\sum_{q=1}^{Q} u_{qk} = 1. \tag{4.21}$$

We derive the antecedent membership functions from the clustering results. One rule is derived from each cluster. A multi-dimensional Gaussian membership function is placed at the location $\mathbf{v}_q$ of each cluster centre. The spreads of the membership functions are derived from the distribution of the data. The fuzzy covariance matrix can be used for this purpose. Let $\mathbf{F}_q$ be the fuzzy covariance matrix for cluster q, which is computed as

$$F_q = \frac{\sum_{k=1}^{K} (u_{qk})^m (\mathbf{x}_k - \mathbf{v}_q)(\mathbf{x}_k - \mathbf{v}_q)^T}{\sum_{k=1}^{K} (u_{qk})^m}. \tag{4.22}$$

Then, the composite membership function u_{A_q} for cluster q is given by

$$u_{A_q} = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{v}_q)^T \mathbf{F}_q^{-1}(\mathbf{x}-\mathbf{v}_q)}. \tag{4.23}$$

These membership functions are generally not oriented along the axes of the product space of antecedent variables. In fuzzy models, the membership functions are usually defined on each variable separately and then composed by using a t-norm. In that case, the multivariate Gaussian membership function can be projected onto the individual axes by taking the diagonal elements of $\mathbf{F}_q$ as

$$u_{A_q} = e^{-\frac{1}{2}(\mathbf{x}-\mathbf{v}_q)^T [\mathrm{diag}(\mathbf{F}_q)]^{-1}(\mathbf{x}-\mathbf{v}_q)}. \tag{4.24}$$

Note that for the PFS VaR models used in this chapter, (4.23) and (4.24) are equivalent, since we use a single dimensional antecedent space.

After determining the parameters of the antecedent and consequent membership functions, the probability parameters $\Pr(C_j|A_q)$ of the fuzzy system must be determined. As explained in Section 4.3, the optimal values of these parameters can be determined by using maximum likelihood estimation. Hence, we maximised the following log-likelihood function (Waltman et al., 2005b)

$$\begin{aligned} L &= \sum_{k=1}^{K} \ln \left(\sum_{j=1}^{N-1} \bar{u}_{C_j}(y_k) \sum_{q=1}^{Q} \beta_q(\mathbf{x}_k) \Pr(C_j|A_q) + \right. \\ &+ \left. \bar{u}_{C_N}(y_k) \sum_{q=1}^{Q} \beta_q(\mathbf{x}_k) \left[1 - \textstyle\sum_{j=1}^{N-1} \Pr(C_j|A_q) \right] \right), \end{aligned} \tag{4.25}$$

where

$$\bar{u}_{C_j}(y_k) = \frac{u_{C_j}(y_k)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}. \tag{4.26}$$

The optimisation algorithm has been initialised by using (4.13) to estimate the initial values for the probability parameters.

4.4.2 Fuzzy histogram model

In PFS_2, the probabilistic fuzzy system is interpreted as a fuzzy histogram. The membership functions of the model are selected based on the modeler's choice. We have fixed the distribution and type of the membership functions. The input and output spaces are both partitioned into nine fuzzy regions, using equally distributed triangular membership functions, combined with shouldered membership functions at the edges of the domain, as shown in Fig. 4.2(a).

Maximum likelihood estimation could again be used for determining the conditional probability parameters. Another alternative approach is based on using a method directly related to an exception based back testing, in this case the Kupiec statistical test. It is interesting to find the optimal parameters by minimising an objective function based on the test statistic since the acceptance of the models is based on it. In this approach, a suitable cost function is used and the optimal solution is determined by minimising the cost function using a general search algorithm. Since the cost function we use may have a finite number of discontinuities, we use a derivative-free optimisation algorithm. In particular, we use a real coded genetic algorithm (GA), where each element in the chromosome of an individual corresponds to a probability parameter.

A genetic algorithm repeatedly modifies a population of solutions. The solutions are represented as chromosomes which can be combined to produce offspring through crossover operations. Crossover operations use information from two or more parent chromosomes to generate new chromosomes. Members of the population can be altered between generations by applying local mutations to chromosomes. Mutation is a background operator, which produces spontaneous random changes in various chromosomes, to guarantee population diversity. Members of a current population are selected for crossover and mutation according to some determined random scheme, which takes into account a measure of the quality of the solution that they represent. This function is called the fitness function. Members with higher fitness are more likely to be selected than those with lower fitness and are therefore more likely to pass good solution information to the next generation. Over successive generations, the population evolves toward an optimal solution. The algorithm is well suited to problems that are complex and have a large search space, making them impossible to search exhaustively.

In this work we use the real coded GAs (Wright, 1991; Michalewicz, 1996; Herrera et al., 1998) to estimate the probability parameters $\Pr(C_j|A_q)$ of the fuzzy system. Since our models have nine antecedent membership functions and nine output membership functions, there are 81 parameters to optimise. Each solution in the real coded GA is represented as a vector of 81 values. Roulette wheel selection has been used to select individuals for reproduction. Weighted mean is used as the crossover operator, while Gaussian mutation was used as the mutation operator. The initial population was generated by random perturbations of the estimates obtained from (4.13) by adding zero-mean normally distributed noise with standard deviation 0.5. Afterwards, the disturbed parameters were normalised to make sure that the probabilities add up to 1.

At each iteration, the VaR model is computed with the corresponding probability parameters for each individual by using the returns at period t to predict an estimated distribution of the returns at period $t + 1$. With this estimated distribution, it is possible to obtain the number of

Table 4.1: Stocks used in the empirical study.

	Data Range
KPN	06/01/1999 – 27/12/2002
ABN	05/01/2000 – 29/12/2003
JiaLing	23/04/2002 – 17/07/2006
BaoShan	04/01/2001 – 04/04/2005
COSCO	26/09/2002 – 16/07/2007
Merchant Bank	10/05/2002 - 28/08/2006

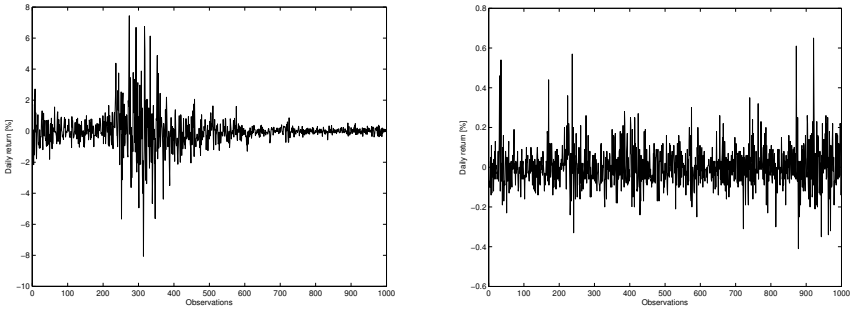
exceptions N for the candidate model. Note that it is possible to find the optimal number of exceptions N^* that minimises (4.5) since T and c are fixed in the experiments. Then, the goal of optimisation is to select those parameters that minimise the difference between N and N^*. We minimise the following cost function.

$$M = |N - N^*| + \frac{10 \text{ sign } (N - N^*) \, |P_{50} - \bar{x}|}{\max(x) - \min(x)}, \tag{4.27}$$

where P_{50} is the obtained r_v in (4.1) when c is 50%. Since $|N - N^*| \in N$ is a discontinuous function with flat regions, the second term of (4.27) enables the chromosomes to be distinguished from one another even for small changes of decision variables. The model built by this process, is by construction an optimal VaR model that minimises a cost function directly related to the Kupiec statistical test.

4.5 Experimental study

In an empirical study, we have studied the performance of the proposed probabilistic fuzzy systems (PFS_1 and PFS_2) to estimate VaR for different stocks. The performance of both PFS models has been compared with the performance of the GARCH models. We are interested in the one-period VaR value. Table 4.1 shows the six different stocks that have been considered in our study. Two of the companies (KPN and ABN AMRO) have originally been established in the Netherlands. Since 1990's, the stocks of these two companies have been traded on the Dow Jones Exchange. At the time of this study, both of them were one of the Fortune Global 500 corporations. Their stock prices have gone through several periods of positive and negative tides of the global economy. The other four companies that we studied are Chinese companies traded on the Shanghai Stock Exchange in China. These companies are the China Jialing Industry Company, China Merchant Bank, China Baoshan Steel Company and COSCO Group.



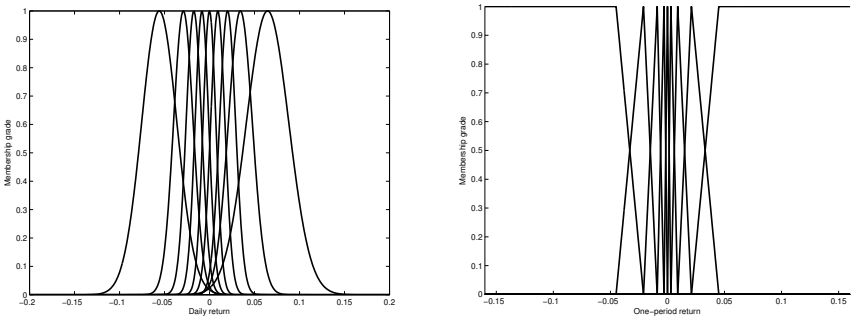
(a) KPN (b) China Merchant Bank

Figure 4.3: Daily returns of KPN and China Merchant Bank.

They all share some similar properties. First, each of them is a big state-owned company. The number of employees exceeds 5000. The value of their total assets reaches 100 billion Chinese dollar. Second, each of them has come into the stock market both in China and in Hong Kong. In addition, their stocks were first traded on the Chinese stock market since 1992. Third, these companies have already existed for more than 20 years and gone through times of huge economics innovation in China.

For all companies, we collected the company's daily closing stock price for 1000 trading days from either the Dow Jones Jones Exchange or the Shanghai Stock Exchange. The samples were collected from different periods selected arbitrarily in order to reduce the sensitivity to the global conjecture at a given period of time. The first 500 samples were used as the training set. The remaining 500 samples are used as the test set. Note that the stocks of the considered companies are not all similar. Figure 4.3 shows the daily returns of the KPN and the China Merchant Bank. As can be seen, there are differences in terms of volatility and the distribution of the returns.

For each of the data sets, probabilistic fuzzy value at risk models have been developed by using the two approaches outlined in Section 4.4. For PFS_1, models have been obtained for different values of the scaling parameter z for the output in order to determine a suitable value. We report the models obtained after finding a suitable value of z through simple search. PFS_2 models do not use a scaling parameter for the output. Furthermore, GARCH $(1, 1)$ models have been developed for each data set by using maximum likelihood estimation for the parameters. The performances of the PFS models and the GARCH models have been compared by using the test set. All the results reported in this section are related to the test sets. In order to reduce the sensitivity to algorithm initialisation, the experiments were run 30 times. We report the best



(a) Antecedent membership functions. (b) Consequent membership functions.

Figure 4.4: Antecedent and consequent membership functions for ABN AMRO.

models obtained. The validity of the models is assessed with the exception-based back testing method using the Kupiec statistical test.

In our models, we have used nine antecedent membership functions and nine consequent membership functions. Hence, the fuzzy system had nine rules. For model PFS_1 the FCM algorithm was run with nine clusters. In such a system, there are 81 probability parameters $\Pr(C_j|A_q)$ (nine for each rule). We now give some more details of the model for one of the stocks we have studied (ABN AMRO). The antecedent membership functions obtained for model PFS_1 after FCM clustering are shown in Fig. 4.4(a). Since FCM has the tendency to place more clusters in regions covered with more data, there are more antecedent membership functions in the centre, where more samples are available. The output membership functions are triangular, and they follow the pattern shown in Fig. 4.4(b).

Given the fuzzy membership functions whose parameters are determined as above, the conditional probability parameters for PFS_1 are determined by using maximum likelihood estimation. Given the conditional probability distribution of one period returns, the value at risk of the portfolio is obtained by using (4.1) and (4.2).

In the PFS_2 models, we have used nine antecedent membership functions and nine consequent membership functions, with 81 probability parameters $\Pr(C_j|A_q)$. As already stated these parameters are estimated using a real coded genetic algorithm to minimise (4.27). In our implementation of the GA, each individual from the population corresponds to a different set of probability parameters. A population with 20 individuals was used. The selection probabilities were calculated and successive pairs of individuals were drawn using the roulette wheel selection.

90% of the population at the next generation was created by a crossover function, that creates children that are the weighted arithmetic mean of two parents. Gaussian mutation was used, where a random number is added to a variable, taken from a Gaussian distribution. The termination condition was to stop the algorithm when the best individual did not improve in over 100 consecutive generations or if the cumulative change in the fitness function value was less than 1×10^{-6}. During our experiments all tests halted by the fitness function criterion. All models were implemented in Matlab.

The steps necessary for computing the one-period value-at-risk of a portfolio can now be summarised as follows for PFS_1 models.

1. Collect the price series regarding the portfolio and compute the one-period returns. Create training and validation data sets.

2. Determine antecedent membership functions: apply fuzzy c-means clustering to compute the locations of the membership functions and use cluster covariance (4.22) to obtain the spreads from (4.23).

3. Select the number of consequent membership functions and form a partition as shown in Fig. 4.2(b). Determine the value of the scaling factor z.

4. Given the definitions of the antecedent and the consequent membership functions, determine the optimal probability parameters of the PFS by maximising (4.25).

5. Using the test set, compute the estimated conditional probability distribution function for the one-period returns for each observation in the test set.

6. Given the conditional probability distribution functions, compute the VaR by using (4.1) and (4.2).

7. Validate the model by using exception based back-testing as explained in Section 4.2.

The steps necessary for computing the one-period value-at-risk of a portfolio for models PFS_2 can be summarised as follows.

1. Collect the price series regarding the portfolio and compute one-period returns. Create training and validation data sets.

2. Determine the antecedent and consequent partition over the universe of discourse, with nine equally spaced membership functions.

3. Given the antecedent and consequent membership functions, determine the 81 optimal probability parameter of the PFS, by minimising (4.27), for the training data set.

Table 4.2: Influence of the scaling factor on the failure rates for ABN AMRO for different VaR confidence.

	Scaling factor			
c	0.002	0.003	0.004	0.005
95%	0	0.068	0.068	0.092
97.5%	0	0.032	0.038	0.046
99%	0.006	0.014	0.026	0.022

4. Compute the estimated conditional probability distribution function for the one-period return for each observation in the validation data set and the corresponding VaR, using (4.1) and (4.2).

5. Validate the model by using the exception based back-testing as explained in Section 4.2.

4.6 Results

We start by considering the influence of the scaling factor z for PFS_1. Table 4.2 shows the influence of the scaling factor z on the observed failure rates of the model for different values of the VaR confidence level c for one of the data sets. It can be observed from the table that as the value of the scaling factor increases, the number of failure rates is also increasing, indicating an underestimation of the VaR value. Conversely, the number of failure rates may be reduced to zero when z becomes small, indicating an overestimation of the VaR value. Both cases are undesirable and hence an optimal value for z should be determined. This is done by minimising the average deviation in failure rate for different values of the VaR confidence c. Specifically, the absolute difference between the theoretical failure rate and the observed failure rate in the data is computed for 95%, 97.5% and 99% VaR estimation. The mean of these three numbers is taken as the index to be minimised. The optimal z value is the one that corresponds to the minimal value of this index. This procedure was repeated for the data of all companies. It can be seen that all probability variables are positive according to this estimation. Table 4.4 shows the optimal probability parameters obtained after maximum likelihood estimation for PFS_1. Note that some of the probability parameters are now zero. Furthermore, after an extreme return (*e.g.* first and ninth rows), the returns tend to be extreme as indicated by large values of probability in the first and ninth columns and low values in the fourth and fifth columns. Conversely, the returns tend to be average after an average event (*e.g.* fifth row) as indicated by zero probability in the first and ninth columns. This is an indication that there is volatility clustering in this data set (van den Berg et al., 2004).

Table 4.3 shows the initial probability parameters obtained with (4.13).

Table 4.3: Initial probability parameters for ABN AMRO model.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0.1003	0.1333	0.2066	0.0567	0.0441	0.1215	0.1012	0.1218	0.1143
2	0.0565	0.1351	0.1659	0.0792	0.0744	0.0972	0.1617	0.1489	0.0811
3	0.0459	0.1679	0.1495	0.1300	0.1024	0.0773	0.1463	0.1359	0.0448
4	0.0544	0.1683	0.1700	0.1105	0.0647	0.0802	0.1454	0.1579	0.0468
5	0.0516	0.1578	0.1800	0.1119	0.0692	0.0770	0.1547	0.1472	0.0506
6	0.0563	0.1648	0.1760	0.0956	0.0655	0.1206	0.1340	0.1227	0.0646
7	0.0700	0.1877	0.1659	0.0748	0.0405	0.0901	0.1147	0.2002	0.0562
8	0.0529	0.1625	0.1626	0.1122	0.0999	0.0930	0.1313	0.1286	0.0570
9	0.0539	0.1729	0.1624	0.1132	0.0746	0.0772	0.1476	0.1505	0.0476

Table 4.4: Probability parameters for PFS_1 ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0.1401	0.0850	0.3776	0	0	0.1586	0	0.0940	0.1446
2	0.0247	0.0762	0	0	0.1366	0.1239	0.3359	0.2380	0.0646
3	0	0.0308	0	0.3430	0.2831	0	0.3432	0	0
4	0.0017	0.1197	0.0895	0.1036	0.0348	0.1577	0.1609	0.3321	0
5	0	0.1893	0.4046	0.2158	0.0725	0	0.1178	0	0
6	0	0	0.2309	0.1297	0	0.3170	0.2131	0	0.1093
7	0	0.4398	0.1626	0.0046	0	0	0	0.3931	0
8	0.0455	0.3267	0.2359	0	0.0755	0.0511	0.0120	0.2105	0.0429
9	0.0595	0.3158	0.2772	0	0	0.0478	0.0980	0.1970	0.0047

Table 4.5: Probability parameters for PFS_2 ABN AMRO model after optimisation.

	Consequent								
Rule	1	2	3	4	5	6	7	8	9
1	0.0741	0.0850	0.1610	0.1680	0.1142	0.1041	0.0861	0.0998	0.1077
2	0.1056	0.0372	0.1052	0.1610	0.1737	0.0331	0.0640	0.1549	0.1654
3	0.0238	0.0380	0.1133	0.1286	0.1157	0.1744	0.1362	0.1032	0.1667
4	0.0000	0.0875	0.1181	0.0941	0.1516	0.1782	0.1751	0.1839	0.0113
5	0.0529	0.1130	0.1052	0.1298	0.1469	0.1806	0.1064	0.0873	0.0778
6	0.0540	0.0944	0.1351	0.1585	0.0739	0.1096	0.1076	0.1527	0.1140
7	0.0788	0.1554	0.1199	0.1678	0.0241	0.1673	0.1501	0.0219	0.1148
8	0.0869	0.0769	0.1450	0.1618	0.1263	0.1013	0.1158	0.0956	0.0904
9	0.1643	0.0178	0.0797	0.0951	0.1504	0.2045	0.1304	0.1284	0.0294

When considering the optimal parameters for the PFS_2 model (Table 4.5), we see that none of the probability parameters are zero. Hence, the genetic optimisation algorithm does not lead to maximum likelihood parameters, but as explained below the models are accepted. This indicates that the fitness landscape can be very flat, leading to different, but equally acceptable solutions.

Table 4.6 shows the obtained results of the exception-based back testing for the best GARCH and the probabilistic fuzzy models. This table shows the number of exceptions that have occurred in the validation data for different levels of the confidence parameter c. The bold face numbers indicate that the model is not rejected according to the test statistic. The non-rejection region for the Kupiec test statistic is also shown. The optimal number of exceptions, according to (4.5) is 25 for $c = 95\%$, 13 for $c = 97.5\%$ and 5 for $c = 99\%$.

As can be seen in Table 4.6, the GARCH models are rejected for some data sets, while the PFS models are accepted for all data sets. However, note that the estimation of the probability parameters is a problem with multiple minima. In different runs, different solutions for the probability parameters were obtained. Table 4.7 shows the mean (η_{PFS_2}) and standard deviation (σ_{PFS_2}) of the failure rate obtained during the 30 experiments made to reduce the sensitivity to algorithm initialisation in the PFS_2 model, as well as the percentage of tests (P_{PFS_2}) that are accepted by the exception-based back testing. As Table 4.7 shows, in some of the cases, the number of obtained exceptions was in the rejection region, and in the case of $c = 99\%$, the PFS_2 model is accepted most of the times. We conjecture that through the use of a different objective function that takes the differences in losses between the actual and predicted VaR, this problem can be solved.

It is also interesting to consider how the VaR values estimated by the PFS compare to the values estimated by the GARCH models. Table 4.8 shows the sum of the differences between the VaR estimated and the actual losses in the periods where the VaR estimation is smaller than the actual losses, i.e., when exceptions occur. The daily returns and VaR estimates for the PFS and GARCH models with $c = 97.5\%$ are shown in Fig. 4.5.

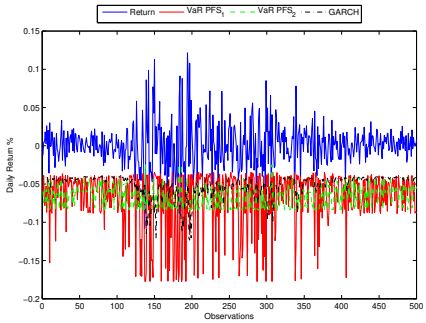
As can be seen in Table 4.8, the expected losses are in most cases smaller in the PFS models. In other cases where the GARCH model has smaller total expected losses than the PFS model, the GARCH model leads to a smaller number of exceptions (which could indicate a conservative model). Note, for instance, that in stock COSCO with $c = 95\%$ the GARCH model is too conservative and is not accepted according to the Kupiec test. The losses for the stocks Baoshan and ABN with $c = 95\%$ are much larger in the PFS_2 model than in the PFS_1 and the GARCH models. This indicates that the genetic optimisation suffers from multiple local minima. The model is accepted by the Kupiec test, which considers only the number of exceptions, but exhibits larger expected losses for a few cases. This is a weakness of the proposed training

Table 4.6: Failure rates for back testing.

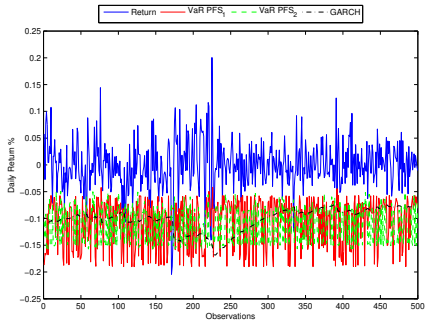
Asset	c	PFS_1	PFS_2	GARCH	Non-Rejection
ABN	95%	**34**	**25**	**19**	$16 < N < 36$
	97.5%	**16**	**12**	**13**	$6 < N < 20$
	99%	**7**	**6**	**9**	$1 < N < 10$
KPN	95%	**29**	**24**	11	$16 < N < 36$
	97.5%	**14**	**8**	**8**	$6 < N < 20$
	99%	**7**	**5**	**4**	$1 < N < 10$
JiaLing	95%	**31**	**25**	**22**	$16 < N < 36$
	97.5%	**19**	**13**	**14**	$6 < N < 20$
	99%	**8**	**4**	**6**	$1 < N < 10$
BaoShan	95%	**36**	**24**	12	$16 < N < 36$
	97.5%	**16**	**13**	**8**	$6 < N < 20$
	99%	**6**	**5**	**6**	$1 < N < 10$
COSCO	95%	**32**	**25**	14	$16 < N < 36$
	97.5%	**18**	**12**	**11**	$6 < N < 20$
	99%	**8**	**5**	**5**	$1 < N < 10$
Merchant	95%	**21**	**27**	10	$16 < N < 36$
	97.5%	**12**	**13**	5	$6 < N < 20$
	99%	**5**	**5**	**4**	$1 < N < 10$

Table 4.7: Average failure rates for PFS_2 model.

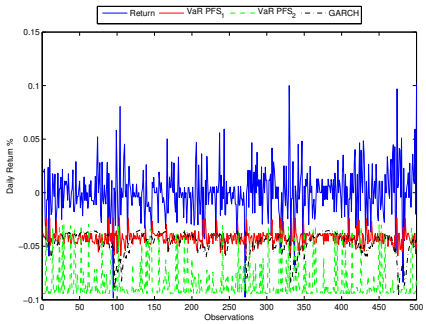
Asset	c	η_{PFS_2}	σ_{PFS_2}	P_{PFS_2}	Non-Rejection Region
ABN	95%	40.7	15.5	36.67%	$16 < N < 36$
	97.5%	24.0	13.4	33.33%	$6 < N < 20$
	99%	7.6	4.2	76.67%	$1 < N < 10$
KPN	95%	56.0	38.3	16.67%	$16 < N < 36$
	97.5%	25.7	15.3	23.33%	$6 < N < 20$
	99%	8.9	6.1	66.67%	$1 < N < 10$
JiaLing	95%	36.0	17.3	33.33%	$16 < N < 36$
	97.5%	25.7	18.8	23.33%	$6 < N < 20$
	99%	7.9	4.0	66.67%	$1 < N < 10$
BaoShan	95%	42.5	16.9	23.33%	$16 < N < 36$
	97.5%	25.9	12.1	33.33%	$6 < N < 20$
	99%	8.3	6.3	73.33%	$1 < N < 10$
COSCO	95%	42.9	20.9	30.00%	$16 < N < 36$
	97.5%	23.0	15.8	40.00%	$6 < N < 20$
	99%	7.2	4.5	66.67%	$1 < N < 10$
Merchant	95%	39.7	16.6	26.67%	$16 < N < 36$
	97.5%	25.1	14.9	33.33%	$6 < N < 20$
	99%	6.9	4.5	70.00%	$1 < N < 10$



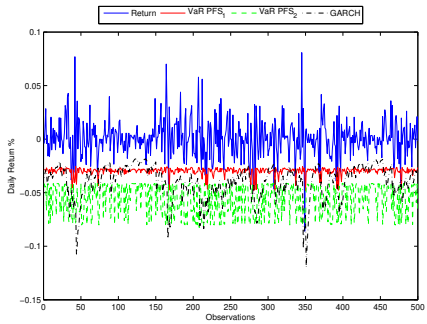
(a) ABN



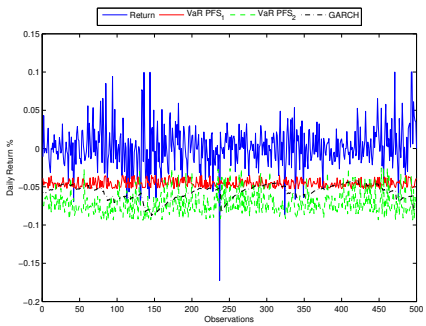
(b) KPN



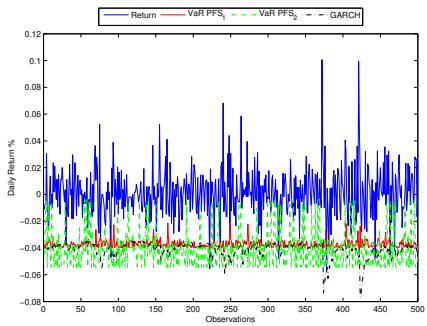
(a) JiaLing



(b) BaoShan



(a) COSCO



(b) Merchant

Figure 4.5: Daily Returns and VaR estimates with $c = 97.5\%$.

Table 4.8: VaR Exceptions.

Asset	c	PFS_1	PFS_2	GARCH
ABN	95%	0.5886	1.8622	0.3579
	97.5%	0.3185	0.2850	0.6878
	99%	0.1841	0.1196	0.8943
KPN	95%	0.8224	0.5209	0.9735
	97.5%	0.5166	0.4454	0.8070
	99%	0.3587	0.1465	0.4169
JiaLing	95%	0.5640	0.3344	0.8680
	97.5%	0.3288	0.6796	0.6314
	99%	0.2187	0.0827	0.3174
BaoShan	95%	0.3673	2.0010	0.3514
	97.5%	0.2025	0.2873	0.2864
	99%	0.0755	0.1978	0.1824
COSCO	95%	0.6517	0.9291	0.6776
	97.5%	0.3862	0.7746	0.6128
	99%	0.2072	0.2119	0.3054
Merchant	95%	0.2100	0.9166	0.3579
	97.5%	0.0900	0.3013	0.1864
	99%	0.0358	0.1963	0.1720

methods for PFS_2. Choosing another fitness function may help alleviate this problem. It is also interesting to note that the PFS_2 model in some cases exhibits a more conservative behaviour than the PFS_1 model and simultaneously has the number of exceptions equal, or very close, to the optimal number of exceptions, according to (4.5), for the stocks BaoShan, COSCO and Merchant, as can be seen in Fig. 4.5.

4.7 Conclusions

We have proposed applying probabilistic fuzzy systems for value at risk modelling of portfolios for quantifying the market risk. Probabilistic fuzzy systems combine an approximate rule based description of system behaviour with statistical properties of data. We have studied two approaches for determining the model parameters. In one, a Mamdani probabilistic fuzzy system is designed by following a sequential approach for determining the model parameters. The location of the antecedent membership functions is determined by using fuzzy clustering. Maximum likelihood parameter estimation is used for determining the probability parameters of the PFS. The output membership functions are scaled by using a scaling parameter. In another, we determine the membership parameters based on (expert) knowledge by fixing the distribution and type of the membership functions. The conditional probability parameters are estimated by minimising the test statistic of a back testing method by using a constrained evolutionary

optimisation method. It is found that the data-driven sequential approach leads to more stable models. With evolutionary optimisation, locally optimal solutions are often obtained.

The performance of the proposed models has been compared to the VaR estimation by using the popular GARCH (1,1) volatility estimation. Exception-based back testing is used for this purpose. It is found that PFS models are not rejected by back testing, while GARCH models are sometimes rejected. This shows the added flexibility that comes through the use of the probabilistic fuzzy models, enabling them to adapt to the properties of the data. In terms of the computed value at risk models, the proposed methods tends to be less conservative. However, this depends on the specific portfolio for which the VaR measure is being computed. Furthermore, with some portfolios, PFS models estimate larger VaR at 99% confidence. The reasons for this behaviour could be multi-fold, but we think it might be related to the degree of volatility clustering observed in the return series. As future work, we will investigate in more detail whether this difference in the behaviour can be attributed to the degree of volatility clustering.

Chapter 5

A multi-covariate multi-horizon conditional volatility model using probabilistic fuzzy systems[1]

5.1 Introduction

ANALYZING the prices and risk of financial products is a major concern for financial institutions as well as for macroeconomic policy makers such as central banks. Financial institutions undertake activities to reduce the possibility of failure to an acceptable range. These activities may include portfolio adjustment, hedging or insurance, to manage the different kinds of risk that they are exposed to, such as the market risk, the credit risk and the operational risk (Hull, 2000). Furthermore, macroeconomic policy makers such as central banks use prices of influential returns such as the S&P 500 returns as broad indicators of the country's economy. According to these indicators, monetary or fiscal policy can be adjusted (Rigobon and Sack, 2003).

The analysis of prices and risk of financial products is often based on their market risk. Market risk is the exposure to the uncertain market value of a portfolio (Holton, 2003). Uncertain market values of a portfolio or a financial product cannot be assessed if the focus is only the expected return of the portfolio. For this reason, estimating and forecasting the whole density of the portfolio, or the returns constituting the portfolio are of importance. Value at risk (VaR) is a conventional measure to quantify the market risk from the estimated conditional density of returns or a portfolio. It is a single number for the senior management to express and summarise the total market risk of a portfolio with financial assets. Value at risk measures the

[1]Parts of this chapter have been published in Almeida, Basturk, Kaymak, and Milea (2012a); van den Berg, Kaymak, and Almeida (2013); Almeida, Basturk, and Kaymak (2014a).

worst expected loss over a given horizon under normal market conditions at a given confidence level. Due to regulations, large banks must nowadays base their market risk capital requirements on the VaR estimate (Jorion, 2006). This implies that accurate VaR models are needed for allocating the capital more efficiently in order to cover possible losses.

Additionally, analyzing the risk of returns in longer horizons is important since the horizon of decision making relying on these financial products may vary. While some contracts allow the financial institutions to alter their portfolios in a very short time, other contacts are adjustable only in longer horizons and hence require a longer term risk analysis. Particularly the decision making by policy makers is not possible in short horizons. For example, the U.S. monetary or fiscal policy is not adjusted in a daily frequency. Hence analyzing prices and market risk of financial products at longer horizons is of interest both for financial institutions and for policy makers.

Value-at-Risk can be obtained from the negative lower quantile of the conditional distribution function. Simulation approaches or parametric approaches are usually used for conditional density estimation and for VaR estimation. The simplest models of volatility assume that it does not vary over time while more advanced models acknowledge that volatility varies dynamically over time. The disadvantage of the parametric approach is that the data usually do not follow the parametric distributions that are assumed to underlie the data generating process. Obtaining these density estimates or forecasts is not trivial since the considered density is typically dynamic, i.e. changes over time. These changes can be attributed to certain variables as well as seasonal factors, such as the day of the week (Berument and Kiymaz, 2001). Therefore, more flexible modelling approaches are needed.

A semi-parametric model in which the model structure and the model parameters can be adapted to the underlying data distribution is the probabilistic fuzzy system (PFS). In a probabilistic fuzzy system, a linguistic description of the system behaviour encoded by the fuzzy rules is combined with the statistical properties of data. PFS is suitable for approximating probability distributions. Since accurate VaR estimation is enabled by estimating the probability distribution of the data, PFS has been used to estimate it (Xu and Kaymak, 2008; Almeida and Kaymak, 2009b,a). These studies focused on single covariate models using a multi-step sequential approach for determining the model parameters. In Almeida and Kaymak (2009b) separate approaches are proposed to design a PFS for two different reasoning mechanisms.

In this work, we consider conditional density estimation and the related VaR estimation by using a multi-covariate and multi-output PFS model on daily S&P 500 returns data. We use a single approach to design a PFS for the two different reasoning mechanisms, which provide additional information, linguistic interpretation and process understanding. In this application we first analyze whether the conditional density of returns and the associated VaR can be suc-

cessfully estimated with a PFS multi-covariate model including information on past returns, as suggested in Villani et al. (2009). The performance of the model is illustrated using the quantiles of the estimated conditional density function for returns. Furthermore, the performance of the proposed model and a set of alternative models in terms of the VaR estimates are evaluated by using an unconditional coverage test (Kupiec test) and an independence test (Christoffersen Markov test). Comparing the proposed model and estimation method with the conventional GARCH models, it is found that the GARCH models are sometimes rejected, while the PFS models of VaR are never rejected. We next apply the PFS model with seasonal factors to the S&P 500 returns. Specifically, we analyze whether the conditional distribution of returns and associated VaR depend on the day of the week, as suggested in Berument and Kiymaz (2001). Furthermore, we analyze the performance of the PFS model in short and long term horizons focusing on the one day ahead and one month (21 trading days) ahead forecasts using the proposed multi-covariate multi-output model. The extended model captures seasonal patterns in S&P 500 returns in short and longer horizons as well as the increased risk factor in longer term forecasts of returns. The model is shown to perform well in VaR estimation according to the unconditional coverage and independence tests. Regarding the PFS model estimation, we present a two-step approach to estimate the model parameters, by dividing the parameter estimation into two groups: the input parameters and the output and reasoning parameters.

The outline of this chapter is as follows. In Section 5.2, we summarize the basics of PFS models, introduce the multi-covariate multi-output PFS model and provide a discussion on the additional information, linguistic interpretation and process understanding provided by two equivalent reasoning mechanisms that the PFS model provides. We discuss the VaR estimation using the PFS models and model validation techniques for VaR estimation in Section 5.3. The extended multi-covariate multi-output PFS model is introduced in Section 5.4. In Section 5.5 the multi-covariate PFS is applied to the S&P 500 returns for conditional density estimation using the two different reasoning mechanisms that PFS model provides. Value-at-Risk estimation for the S&P 500 returns including the seasonality analysis and multiple period ahead forecasts are provided in Section 5.6. Finally, conclusions are given in Section 5.7.

5.2 Probabilistic fuzzy systems

Probabilistic fuzzy systems (PFS) are based on the concept of the probability of a fuzzy event, as defined by Zadeh (1968). Probabilistic fuzzy systems combine two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. The PFS consists of a set of rules whose antecedents are fuzzy conditions and whose consequents are probability distributions.

Assuming that the input space is a subset of $\mathbb{R}^n$ and that the rule consequents are defined on a finite domain $Y \subseteq \mathbb{R}$, a probabilistic fuzzy system consists of a system of rules R_q, $q = 1, \ldots, Q$, of the type

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } f(y|A_q)\,, \tag{5.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an input vector, $A_q : X \longrightarrow [0,1]$ is a fuzzy set defined on X and $f(y|A_q)$ is the conditional pdf of the stochastic output variable $\underline{y}$ given the fuzzy event A_q.

For the purpose of this study, we consider two possible and equivalent reasoning mechanisms of PFS, namely the fuzzy histogram approach and the probabilistic fuzzy output approach (van den Berg et al., 2012). In both cases, we suppose that J fuzzy classes C_j form a fuzzy partition of the compact output space Y.

5.2.1 Fuzzy histogram model

In the fuzzy histogram approach, we replace in each rule of (5.1) the true pdf $f(y|A_q)$ by its fuzzy approximation (fuzzy histogram) $\hat{f}(y|A_q)$ yielding the rule set $\hat{R}_q$, $q = 1, \ldots, Q$ defined as

$$\hat{R}_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f(y) \text{ is } \hat{f}(y|A_q)\,, \tag{5.2}$$

where $\hat{f}(y|A_q)$ is a fuzzy histogram conform (van den Berg et al., 2001)

$$\hat{f}(y|A_q) = \sum_{j=1}^{J} \frac{\hat{\Pr}(C_j|A_q) u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}\,. \tag{5.3}$$

The numerator in (5.3) describes a superposition of fuzzy events described by their membership functions $u_{C_j}(y)$, weighted by the probability $\hat{\Pr}(C_j|A_q)$ of the fuzzy event. The denominator of (5.3) is a scaling factor representing the fuzzified size of class C_j. Because of overlapping membership functions, fuzzy histograms have a high level of statistical efficiency, compared to crisp ones. We show this in Fig. 5.1 where the probability density function (pdf) of the standard normal distribution is approximated by a classical and by a fuzzy histogram using in both cases a partitioning in seven classes. For more details we refer to van den Berg et al. (2004). Besides a high level of statistical efficiency, several classes of fuzzy histograms also have a high level of computational efficiency. An example of such type of fuzzy histogram is one that uses triangular membership functions (Waltman et al., 2005a).

The interpretation of this type of reasoning is as follows. Given the occurrence of a (multi-dimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined

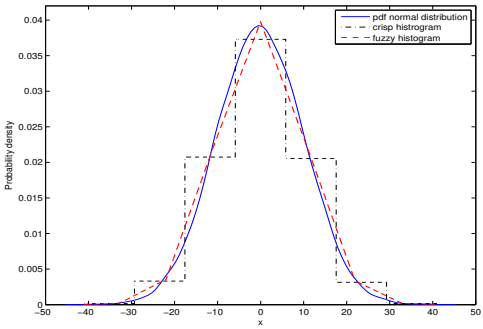


Figure 5.1: A fuzzy histogram approximates a pdf better than a crisp histogram.

on input variables, an estimate of the conditional probability density function based on a fuzzy histogram $\hat{f}(y|A_q)$ is calculated.

5.2.2 Probabilistic fuzzy output model

In the probabilistic fuzzy output approach, sometimes also referred to as Mamdani PFS (Kaymak et al., 2003; Xu and Kaymak, 2008; Almeida and Kaymak, 2009b), we decompose each rule (5.1) to provide a stochastic mapping between its fuzzy antecedents and its fuzzy consequents. The rules are written in the following form.

$$\begin{aligned} \text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } \underline{y} \text{ is } C_1 \text{ with } \hat{\Pr}(C_1|A_q) \text{ and} \\ \cdots \\ \underline{y} \text{ is } C_J \text{ with } \hat{\Pr}(C_J|A_q). \end{aligned} \tag{5.4}$$

These rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. The rules of a PFS also express linguistic information and they can be used to explain the model behaviour by a set of linguistic rules. This way, the system deals both with linguistic uncertainty as well as probabilistic uncertainty.

The interpretation for the probabilistic fuzzy output approach is as follows. Given the occurrence of a (multidimensional) antecedent fuzzy event A_q, which is a conjunction of the fuzzy conditions defined on input variables, each of the consequent fuzzy events C_j is likely to occur. The selection of consequent fuzzy events is done proportionally to the conditional probabilities $\hat{\Pr}(C_j|A_q)$, $j = 1, 2, \ldots, J$. This applies for all the rules R_q, $q = 1, 2, \ldots, Q$.

5.2.3 Equivalence of reasoning mechanisms

Although the fuzzy rule bases (5.2) and (5.4) are different, under certain conditions, the two corresponding probabilistic fuzzy systems implement the same crisp input-output mapping (van den Berg et al., 2012). Let $\beta_q(\mathbf{x}) = u_{A_q}(\mathbf{x}) / \sum_{q'=1}^{Q} u_{A_{q'}}(\mathbf{x})$ be the normalised degree of fulfillment of rule R_q, where u_{A_q} is the degree of fulfillment of rule R_q. When $\mathbf{x}$ is n-dimensional, u_{A_q} is determined as a conjunction of the individual memberships in the antecedents computed by a suitable t-norm, *i.e.,*

$$u_{A_q}(\mathbf{x}) = u_{A_{q_1}}(x_1) \circ \cdots \circ u_{A_{qn}}(x_n), \tag{5.5}$$

where $x_i, i = 1, \ldots, n$ is the i-th component of $\mathbf{x}$ and $\circ$ denotes a t-norm. Then, it can be shown that the output of the fuzzy rules (5.4) is a conditional probability density function, like in the rules (5.2), if an additive reasoning scheme is used with multiplicative aggregation of the rule antecedents (van den Berg et al., 2004). Assuming that the output space is well-formed, *i.e.* the output membership values satisfy

$$\sum_{j=1}^{J} u_{C_j}(y) = 1, \qquad \forall y \in Y\,, \tag{5.6}$$

the conditional probability of the output given an input vector $\mathbf{x}$ can be computed as

$$f(y|\mathbf{x}) = \sum_{j=1}^{J} \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{\mathrm{Pr}}(C_j|A_q) \frac{u_{C_j}(y)}{\int_{-\infty}^{\infty} u_{C_j}(y) dy}\,. \tag{5.7}$$

It is also possible to compute a crisp output of the probabilistic fuzzy system. As shown in van den Berg et al. (2012), using the estimated conditional probability density function, the expected conditional output of the probabilistic fuzzy output is given by the weighted average of the estimated expected output of each rule according to

$$\begin{aligned} \hat{\eta}_{\underline{y}|\mathbf{x}} = \hat{\mathrm{E}}(\underline{y}|\mathbf{x}) &= \int_{-\infty}^{\infty} y \hat{f}(y|\mathbf{x}) dy = \sum_{q=1}^{Q} \beta_q(\mathbf{x}) \hat{\mathrm{E}}(y|A_q) \\ &= \sum_{q=1}^{Q} \sum_{j=1}^{J} \beta_q(\mathbf{x}) \hat{\mathrm{Pr}}(C_j|A_q) z_{1,j}\,, \end{aligned} \tag{5.8}$$

and the estimated conditional variance is

$$\hat{\sigma}^2_{\underline{y}|\mathbf{x}} = \hat{\mathrm{E}}(\underline{y}^2|\mathbf{x}) - (\hat{\mathrm{E}}(\underline{y}|\mathbf{x}))^2 = \int_{-\infty}^{\infty} y^2 \hat{f}(y|\mathbf{x})dy - \hat{\eta}_{y|\mathbf{x}} = \sum_{q=1}^{Q} \beta_q(\mathbf{x})\hat{\mathrm{E}}(\underline{y}|\mathbf{x}) - \hat{\eta}^2_{\underline{y}|\mathbf{x}}$$
$$= \sum_{q=1}^{Q}\sum_{j=1}^{J} \beta_q(\mathbf{x})\hat{\mathrm{Pr}}(C_j|A_q)z_{2,j} - \hat{\eta}^2_{\underline{y}|\mathbf{x}}\,, \tag{5.9}$$

where $\hat{\mathrm{E}}(\underline{y}|A_q)$ is the estimated expected output of each rule,

$$z_{1,j} = \frac{\int_{-\infty}^{\infty} y u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,, \tag{5.10}$$

and $z_{2,j}$ is defined as

$$z_{2,j} = \frac{\int_{-\infty}^{\infty} y^2 u_{C_j}(y)dy}{\int_{-\infty}^{\infty} u_{C_j}(y)dy}\,. \tag{5.11}$$

In this work we are primarily interested in the output of the PFS as a fuzzy histogram, by using (5.3) as an approximation of the conditional distribution of the returns, for computing the VaR value of a portfolio. Note that the same type of fuzzy additive reasoning is applied in both schemes, which eventually yield the same crisp input-output mapping.

5.2.4 Probabilistic fuzzy systems with multiple outputs

The results presented in the previous sections can be extended for the case of multiple outputs, following the distinction between fuzzy input and conditional density output of (5.2). The basic idea is that each one of the c outputs will have an independent probability density function conditional on the same input variables $\mathbf{x}$, making the output of each PFS rule is defined by multiple densities. The fuzzy histogram model rules can be written as

$$R_q : \text{If } \mathbf{x} \text{ is } A_q \text{ then } f_1(y) \text{ is } f_1(y_1|A_q)\text{and}$$
$$\ldots \text{and} f_c(y) \text{ is } f_c(y_c|A_q)\,, \tag{5.12}$$

and the probabilistic fuzzy output rules are

$$\begin{aligned}\text{Rule } \hat{R}_q\text{: If } \mathbf{x} \text{ is } A_q \text{ then } & \underline{y_1} \text{ is } C_{1,1} \text{ with } \hat{\Pr}(C_{1,1}|A_q) \text{ and} \\ & \underline{y_c} \text{ is } C_{c,1} \text{ with } \hat{\Pr}(C_{c,1}|A_q) \text{ and} \\ & \quad \dots \\ & \underline{y_1} \text{ is } C_{1,J} \text{ with } \hat{\Pr}(C_{1,J}|A_q) \text{ and} \\ & \underline{y_c} \text{ is } C_{c,J} \text{ with } \hat{\Pr}(C_{c,J}|A_q). \end{aligned} \tag{5.13}$$

In this work we develop probabilistic fuzzy systems with multiple outputs. The outputs are multiple horizon conditional densities used to estimate value-at-risk.

5.3 Value-at-Risk estimation

In this section, we discuss value at risk estimation by using probabilistic fuzzy systems. Value at risk is a single number for the senior management to express and summarise the market risk of a portfolio of financial assets. The VaR value of a portfolio is always calculated over a time horizon h at a significance level c. It indicates the maximum loss that a portfolio of assets will suffer over a horizon of h (days) with a confidence of c. An overview of the mainstream value at risk estimation methods can be found in Duffie and Pan (1997). Several methods are also discussed in Wiener (1999). Various building blocks of VaR measurement, methods for model validation as well as the differences between the parametric and nonparametric estimation approaches are discussed in Jorion (2006).

5.3.1 Value-at-Risk

Assume that a portfolio has value W_t at time t. Let r denote the one period percentage return of the portfolio. If $f(r)$ is the probability density function of the returns, define $c \in (0, 1), r_v$ such that

$$1 - c = \int_{-\infty}^{r_v} f(r)dr. \tag{5.14}$$

The value at risk VaR_t of the portfolio at time t is then defined as

$$\text{VaR}_t(c) = -r_v W_t. \tag{5.15}$$

Ideally, $\Pr(r_t < -\text{VaR}_t(c)|\Omega_t) = c$, where Ω_t the information set at time t. Assuming that the returns are distributed normally, the key step in the value at risk estimation can be formulated as determining the variance σ^2 of the returns distribution. This is also called *volatility estimation*.

Other models estimate VaR without strict assumptions, such as Gaussian errors or zero mean, on the underlying data distribution. Mixture models (Villani et al., 2009) or nonparametric models (Zmeškal, 2005a) are examples of such models. These flexible models are proposed since the underlying data distribution or this distribution's properties are hard to assess. In this respect, the PFS model is a suitable and flexible model for VaR estimation, since the model does not rely on distributional assumptions of the data. Furthermore, the VaR of a return can be computed directly from the PFS output density.

5.3.2 Volatility estimation

Simulation approaches or parametric approaches are usually used for conditional density estimation and for VaR estimation. In order to estimate the VaR for a given horizon, the daily volatility is scaled to multiple-day volatility (Hull, 2000). The simplest models of volatility assume that it does not vary over time. In that case, the variance σ_t^2 could be estimated by using M observations as $\sigma_t^2 = \frac{\sum_{i=1}^{M} r_i^2}{M}, \forall t$. More advanced models acknowledge that volatility varies dynamically over time. For example, volatility clustering has been observed in the financial markets, which means that there are periods of high variability followed by low variability. GARCH models are used to capture the time varying behaviour of volatility. 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 $t = 1, \ldots, T$ observations is defined as:

$$y_t = \sqrt{h_t}\,\epsilon_t \sim NID(0, h_t) \tag{5.16a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{5.16b}$$

$$\epsilon_t \sim NID\,(0, 1) \tag{5.16c}$$

where y_t is the data with a conditional normal distribution and $NID(\mu, \sigma^2)$ denotes the normal and independently distribution with mean μ and variance σ^2. Scalars q and p are, respectively, the lag order for past returns and past conditional volatility in the GARCH model and $(\alpha_0, \alpha_i, \beta_j)$ for $i = 1, \ldots, q$ and $j = 1, \ldots, p$ are GARCH model parameters.

More advanced models acknowledge that volatility varies dynamically over time. The dynamic aspect of volatility could be modelled in various ways. For example, a multivariate regime switching approach to VaR estimation has been discussed in Billio and Pelizzon (2000). Another model where volatility changes dynamically in time is the GARCH (Generalised Auto

Regressive Heteroscedasticity) model (Bollerslev, 1986). For the GARCH $(1, 1)$ model, which is used quite often in practice, the variance is estimated using a first-order autoregressive model of the squared returns. The disadvantage of the parametric approach is that the data usually do not follow the parametric distributions that are assumed to underly the data generating process. Therefore, more flexible modelling approaches are needed. Flexible parametric models, based on regression density estimation using adaptive mixture of Gaussian (Villani et al., 2009) or student–t (Li et al., 2010) distributions, where the mixture probabilities of the components changes smoothly as a function of the covariates, have been use to analyze the distribution of daily returns. Semi-parametric approaches using a fuzzy stochastic approach (Zmeškal, 2005a), a fuzzy measure model for pricing options (Cherubini and Della Lunga, 2001) and using fuzzy set theory (Bowden, 2006) have also been used for VaR estimation. These flexible models are proposed since the underlying data distribution or this distribution's properties are hard to assess. In this respect, the PFS model is a suitable and flexible model for VaR estimation, since the model does not rely on distributional assumptions of the data. Furthermore, the VaR of a return can be computed directly from the PFS output density.

5.3.3 Model validation

Model validation is the process of checking whether a model performs adequately, and can be done in various ways. In this chapter, we consider a failure test of unconditional coverage test using the Kupiec test (Kupiec, 1995) and independence test using Christoffersen's Markov test (Christoffersen, 1998).

Kupiec has developed a statistical test for assessing the validity of a VaR model (Kupiec, 1995). Kupiec confidence regions are defined through the tail point of the log-likelihood ratio LR_c

$$LR_c = 2\ln\left(\left(\frac{1 - I/T}{c}\right)^{T-I}\left(\frac{I/T}{1-c}\right)^{I}\right) \tag{5.17}$$

In (5.17), I is the number of exceptions and T is the total number of observations. It is considered that an exception $I_t(c)$ has occurred when $r_{t+1} < \text{VaR}_t(c)$. This ratio is shown to be asymptotically χ^2-distributed, with 1 degree of freedom, under the null hypothesis that the VaR model is valid (Kupiec, 1995). Note that the Kupiec test statistic is two sided. Hence, the model is rejected both when there are too few exceptions, (the model is too conservative), as well as when there are too many exceptions, (the model underestimates the volatility).

Besides the tests of unconditional coverage to detect violations of an accurate VaR measure, a variety of tests have been developed which explicitly examine the independence property of the VaR estimation. Christoffersen developed a Markov test (Christoffersen, 1998) to examine

whether the likelihood of a VaR violation depends on the occurrence of a VaR violation on the previous day. If the VaR measure accurately reflects the underlying portfolio risk then the chance of violating today's VaR should be independent of whether or not yesterday's VaR was violated. The idea behind this test is that clustered violations represent a signal of risk model misspecification. Violation clustering is important as it implies repeated severe capital losses to the institution which together could result in bankruptcy.

The test is carried out by recording violations of the VaR on adjacent days, such that if I_t is a first-order Markov process the one-step ahead transition probabilities $\Pr(I_{t+1}|I_t)$ are given by

$$\Pr(I_{t+1}|I_t) = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix}$$

where π_{ij} is the transition $\Pr(I_{t+1} = j|I_t = i)$. Under the null hypothesis, the violations have a constant conditional mean which implies that $\pi_{01} = \pi_{11} = c$.

5.4 Multi-covariate multi-horizon probabilistic fuzzy model

In this chapter, we consider conditional density estimation and value at risk models for h-day ahead returns. In other words, the horizon over which the value at risk is computed is h days. The proposed model allows for multiple outputs, i.e. h_1-days ahead and h_2-days ahead returns. The probabilistic fuzzy models approximate the distribution of future returns, for example returns at time $t + h_1$ and time $t + h_2$, conditional on multi-covariates at time t. In the following sections we describe the covariates used as well as a description of how the models' parameters were obtained.

5.4.1 Model covariates

The covariates include the day-of-the-week to analyse the seasonality effects. The remaining covariates used in this work follows the choices of previous studies (Li et al., 2010; Villani et al., 2009), as these predictors appear to contain valuable information that improves the out-of-sample performance for VaR estimation. Two of the covariates were first used in Geweke and Keane (2007). The first covariate, *LastDay* is the percentage return r_t on day t:

$$\text{LastDay} = r_t = 100 \times \ln\left(p_t/p_{t-1}\right) \tag{5.18}$$

where p_t is the closing price at time t. The second covariate *CloseAbs*$_{95}$, a geometrically decaying average of past absolute returns, and is defined as

$$\text{CloseAbs}_\rho = (1-\rho)\sum_{s=0}^{\infty}\rho^s|r_{t-1-s}|, \tag{5.19}$$

where $\rho = 0.95$ is the discount factor. It is assumed that the mean of each component is constant. The remaining covariates were introduced in Villani et al. (2009). *LastWeek* and *LastMonth* are a moving average of the returns from the previous 5 and 20 trading days, respectively. The variable *CloseSqr*$_{95}$ is defined as

$$\text{CloseSqr}_\rho = \sqrt{(1-\rho)\sum_{s=0}^{\infty}\rho^s r_{t-1-s}^2}, \tag{5.20}$$

and the popular measure of volatility *MaxMin*$_{95}$

$$\text{MaxMin}_\rho = (1-\rho)\sum_{s=0}^{\infty}\rho^s\left(\ln p_{t-s}^{(h)} - \ln p_{t-s}^{(l)}\right), \tag{5.21}$$

where $p_t^{(h)}$ and $p_t^{(l)}$ are the highest and lowest values of the price at time t. This measure has been shown to carry more information on the volatility than changes in closing quotes (Alizadeh et al., 2002). By changing the value of $\rho = 0.80$ in (5.19), (5.20) and (5.21) we obtain, respectively, the covariates *CloseAbs*$_{80}$, *CloseSqr*$_{80}$ and *MaxMin*$_{80}$. In our study, the response variable is the percentage return at time t.

5.4.2 PFS model parameters

Since we use the same type of fuzzy additive reasoning in both the fuzzy histogram and the probabilistic fuzzy output schemes, the same crisp input-output mapping is obtained. Thus, to all intents of this work, the same parameter optimization can be used, yielding to two different interpretations, as it will become apparent in Section 5.6.

The parameters of the probabilistic fuzzy systems consist of the number of rules in the system, the parameters of the antecedent and consequent membership functions (*i.e.* number, type, location, etc.), and the probability parameters $\hat{\Pr}(C_j|A_q)$ of the stochastic mapping between the antecedent and the consequents. The estimation of all the parameters of the PFS simultaneously can be very time consuming and it suffers from the problem of multiple local minima. Thus, we use process knowledge to establish values of a subset of parameters. The other parameters are then optimized given the values of this subset of parameters. Following the distinction between

input and output present in the rule structure of (5.2), the optimization problem is divided in two parts. First we obtain the input membership parameters and then optimize simultaneously the output membership parameters and the probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$.

In this work we determine the parameters of the antecedent membership functions by using a fuzzy clustering heuristic, that uses the fuzzy c-means algorithm (Bezdek, 1981) on the product space of the antecedent variables, to obtain a fuzzy partition matrix $U = [u_{qp}]$ for $p = 1, \ldots, P$ samples. One dimensional fuzzy sets A_{qi}, where $i = 1, \ldots, n$ are obtained from the multidimensional fuzzy sets by projections onto the space of the input variables X. This is expressed by the point-wise projection operator of the form $u_{A_{qi}}(x_{ip}) = \mathrm{proj}_i(u_{qp})$ The point-wise defined fuzzy sets A_{qi} are then approximated by appropriate parametric functions. In this work we choose a combination of Gaussian membership functions of the form

$$u_{A_{qi}}(x_{ip}) = f(x_{ip}; \sigma^1_{qi}, c^1_{qi}, \sigma^2_{qi}, c^2_{qi}) = f^1(x_{ip}; \sigma^1_{qi}, c^1_{qi}) f^2(x_{ip}; \sigma^2_{qi}, c^2_{qi}) \tag{5.22}$$

where

$$f^1(x_{ip}; \sigma^1_{qi}, c^1_{qi}) = \begin{cases} \exp\left(\frac{-(x_{qi}-c^1_{qi})^2}{2(\sigma^1_{qi})^2}\right) & x_{qi} \leq c^1_{qi} \\ 1 & \text{otherwise} \end{cases} \tag{5.23}$$

$$f^2(x_{ip}; \sigma^2_{qi}, c^2_{qi}) = \begin{cases} \exp\left(\frac{-(x_{qi}-c^2_{qi})^2}{2(\sigma^2_{qi})^2}\right) & x_{qi} > c^2_{qi} \\ 1 & \text{otherwise} \end{cases} \tag{5.24}$$

The output membership functions are triangular, as this is a convenient manner to satisfy (5.6) and fuzzy histograms built with these type of membership functions exhibit a high level of computational efficiency (Waltman et al., 2005a). To satisfy (5.6) no matter how extreme the values may be and to ensure that the domain is always covered by the fuzzy partition, the membership functions at the edges of the domain are effectively a trapezoid, as depicted in Fig. 5.6. The distribution of the membership functions can be uniform over the universe of discourse, or it can be varying with more membership functions placed towards the origin (Xu and Kaymak, 2008) or towards the edges of the universe of discourse (Almeida and Kaymak, 2009a). This varying placement allows to better capture the variability in regions with more membership functions. We choose to use a uniform distribution over the universe of discourse as shown in Fig. 5.6. The location of all J fuzzy membership functions can be optimized as a function of the location of the membership functions at the edges, mfC_1 and mfC_J. A similar approach to optimize the output membership functions was done in Xu and Kaymak (2008). In this work, an optimal scaling parameter was found by performing an exhaustive search, separate from the optimization of the remaining parameters of PFS. Assuming that the membership

functions in the rule antecedents have been defined, and the type of consequent membership functions and their distribution are known, the optimal probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$ and location of the output membership functions can be determined by using maximum likelihood estimation, in which the log-likelihood function

$$L = \sum_{p=1}^{P} \ln\left(\hat{\mathrm{Pr}}(y_p|\mathbf{x}_p)\right) \tag{5.25}$$

is maximised when P samples $(\mathbf{x}_p, y_p)$ are available (Waltman et al., 2005b). In (5.25) it is assumed that the samples in the data set are independent of one another. The probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$ must satisfy $\hat{\mathrm{Pr}}(C_j|A_q) \geq 0$ and $\sum_{j=1}^{J} \hat{\mathrm{Pr}}(C_j|A_q) = 1$ for $q = 1, \ldots, Q$ and $j = 1, \ldots, J$. Using (5.7) and (5.25), the log-likelihood can be written as

$$L = \sum_{p=1}^{P} \ln\left(\sum_{j=1}^{J}\sum_{q=1}^{Q} \beta_q(\mathbf{x}_p)\hat{\mathrm{Pr}}(C_j|A_q) z_{1,j}(y_p)\right) \tag{5.26}$$

where

$$z_{1,j}(y_p) = \frac{u_{C_j}(y_p)}{\int_{-\infty}^{\infty} u_{C_j}(y)dy} \tag{5.27}$$

and u_{A_q} is calculated using (5.5). A suitable initialization of the probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$ and mfC_1 and mfC_J for iterative optimization for maximum likelihood estimation is given by direct estimation from the data by using, respectively

$$\hat{\mathrm{Pr}}(C_j|A_q) = \frac{\sum_{p=1}^{P} u_{C_j}(y_p) u_{A_q}(\mathbf{x}_p)}{\sum_{p=1}^{P} u_{A_q}(\mathbf{x}_p)}, \tag{5.28}$$

$$\mathrm{mfC}_1 = \min y, \ \mathrm{mfC}_J = \max y. \tag{5.29}$$

For the case of a PFS with multiple outputs, the likelihood function should take into account the multiple output densities defined by each rule, and combine these densities when deriving the likelihood function. If the multiple outputs of each rule output are assumed to be independent of each other, derivation of the likelihood is straightforward, *i.e.* one only has to multiply the conditional densities obtained in each rule output. This assumption is not very restrictive, since the independence only implies that the 'unexplained' part of the output is independent, given the relation with antecedent variables.

Given the probabilistic fuzzy system whose parameters are determined as above, the conditional probability distribution of h period aheads returns can be obtained from multi-covariates at the previous returns. The value at risk of the portfolio can then be obtained by using (5.14)

and (5.15). The steps for computing the h value at risk of a portfolio can now be summarized as follows.

1. Collect the price series regarding the portfolio and compute the nine quantitative covariates *LastDay*, *LastWeek*, *LastMonth*, $CloseAbs_{95}$, $CloseAbs_{80}$, $CloseSqr_{95}$, $CloseSqr_{80}$, $MaxMin_{95}$, $MaxMin_{80}$. If the purpose is to include the seasonality analysis, this covariate set additionally includes the seasonal covariate *WeekDay*. We use two trading years (2×255 days) for the out-of-sample data and use the remaining data for model estimation.

2. Determine antecedent membership functions: apply fuzzy c-means clustering to compute the membership partition matrix $U = [u_{qp}]$, obtain one dimensional fuzzy sets A_{qi} by projections onto the space of the input variables $\mathbf{x}$. Finally obtain the parametric input membership functions by approximating fuzzy sets A_{qi} to (5.22).

3. Select the number of triangular consequent membership functions satisfying (5.6).

4. Given the definitions of the antecedent and the consequent membership functions type and distribution, determine the optimal probability parameters of the PFS and the optimal location of the output membership functions by maximising (5.26).

5. Using the out-of-sample data set, compute the estimated conditional probability distribution function for the h-period returns for each observation in the out-of-sample data set.

6. Given the conditional probability distribution functions, compute the VaR by using (5.14) and (5.15).

7. Validate the model by using exception based back-testing and independence testing as explained in Section 5.3.3.

5.5 Conditional density estimation for the S&P 500 index

We first analyse whether the proposed multi-covariate model is able to approximate the conditional density of returns, since an accurate estimation of the left tail probability density is necessary for Value-at-Risk estimation.

In this work we use as an example the S&P 500 stock market index. Our data set contains 3718 daily returns from February 18, 1997 to November 23, 2011. The response variable is the percentage return, as calculated in (5.18). A time plot of the response variable r_t is given in Fig. 5.2. The differences in variance and distribution of the returns are clear from this figure.

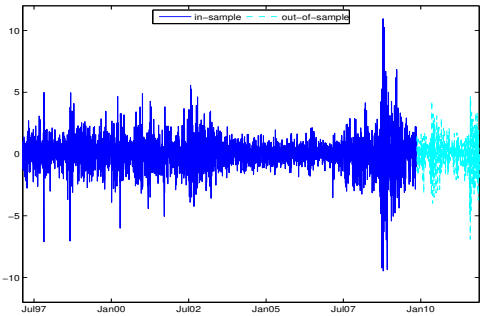


Figure 5.2: Daily returns for S&P 500 between February 18, 1997 to November 26, 2013.

Table 5.1: Estimated quantiles for the in-sample and out-of-sample.

Quantile	$\tau(1\%)$	$\tau(2.5\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$
In-sample	0.94%	2.37%	4.58%	9.57%	19.73%
Out-of-sample	1.37%	3.53%	6.47%	10.39%	19.02%

A probabilistic fuzzy value at risk model has been developed by using the multi-covariate model with nine covariates, *LastDay*, *LastWeek*, *LastMonth*, *CloseAbs*$_{95}$, *CloseAbs*$_{80}$, *CloseSqr*$_{95}$, *CloseSqr*$_{80}$, *MaxMin*$_{95}$ and *MaxMin*$_{80}$ as defined in Section 5.4. For the PFS model, we have used five antecedent membership functions and ten consequent membership functions. Hence, the fuzzy system has five rules, implying that the FCM algorithm was run with five clusters. In such a system, there are 50 probability parameters $\hat{\Pr}(C_j|A_q)$. An out-of-sample evaluation is conducted over a period of two trading years, assuming that a trading year is 255 days. Note that in Villani et al. (2009); Li et al. (2010), all variables except *LastDay*, *LastWeek* and *LastMonth* are used in logarithmic form. In this work we do not follow this approach due to the negative impact of the scaling effects and smoothing on the clustering algorithms.

Table 5.1 presents the estimated quantiles $\tau(c)$ for $c = 1\%, 2.5\%, 5\%, 10\%, 20\%$. The table shows that for example, for $c = 2.5\%$, 2.37% of the data points fall in the estimated quantile, in the in-sample estimation. For the out-of-sample data, the results are close to the corresponding quantiles except for $c = 5\%$, which is slightly higher. The quantiles for $c = 1\%, 5\%$ and the estimated expected value $\hat{\eta}_{\underline{y}|\mathbf{x}}$ are shown in Fig. 5.3. By visually inspecting Fig. 5.3, it is possible to verify that the multi-covariate PFS models show both singleton peaks around high volatility periods as well as stable periods of high volatility. The correct estimation of the conditional density model using PFS allows to provide a correct estimation of the probability distribution of the possible future values of that variable. This is an important issue for quantitative finance and risk management. In the following sections, the additional information and process un-

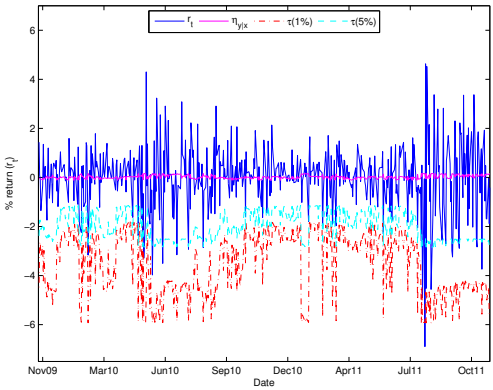


Figure 5.3: Out-of-sample returns r_t, expected value $\hat{\eta}_{\underline{y}|\mathbf{x}}$ and quantiles $\tau(c)$ for $c = 1\%, 5\%$.

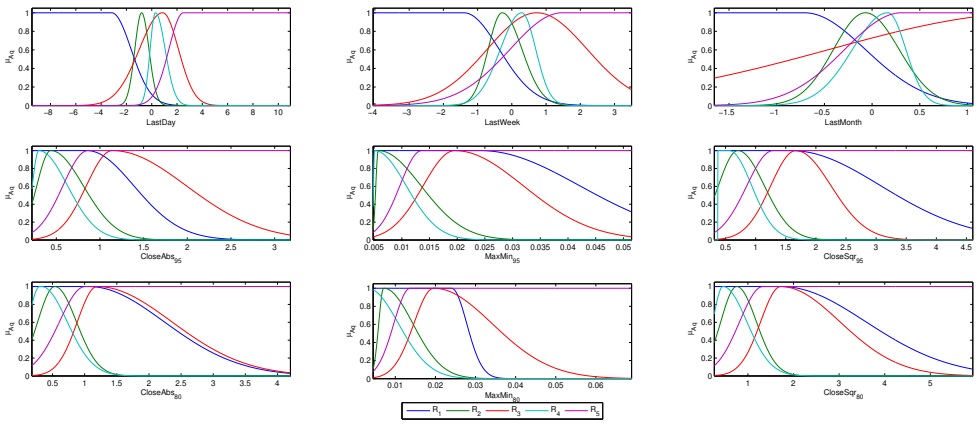


Figure 5.4: Antecedent membership functions.

derstanding provided by the different interpretations of the PFS models are illustrated. Other approaches, as the ones presented in Li et al. (2010); Villani et al. (2009), do not allow for this type of interpretations, as they only consider probabilistic uncertainty. Furthermore, we do not follow the conventional zero-mean assumption for the returns series. As shown in Fig. 5.3, the estimated expected returns $\hat{\eta}_{\underline{y}|\mathbf{x}}$ from (5.8) are not necessarily zero.

5.5.1 Fuzzy histogram model interpretation

Given the fuzzy histogram approach, the rules have the form (5.2). Figure 5.4 and Fig. 5.5 show, respectively, the rule's antecedents membership and each rule's output. Note that the rule

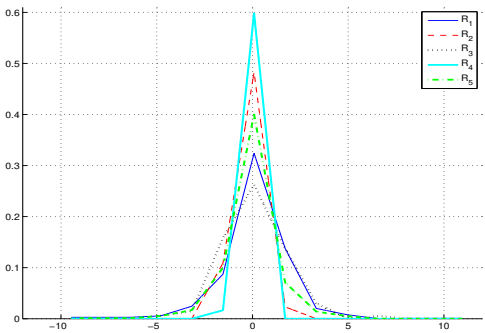


Figure 5.5: Fuzzy histogram rule output.

output for this PFS system is a probability density function estimated with fuzzy histograms, according to (5.3). As shown in Fig. 5.5, the return distribution for each rule is concentrated around 0, and each rule leads to an asymmetric output density for the rule output. Despite these similarities, the obtained return distributions vary substantially across rules as the variance of the output density, hence the volatility of the output is different for each rule. Specifically, the variance of the pdf's from rule 2 and rule 4 is small compared to the universe of discourse of the returns, while the fat tails found in the pdf's from rules 1, 3 and 5 may capture the sudden volatility movements in returns.

The PFS rules do not define clear-cut ranges for the returns, since all densities are concentrated around 0, but instead indicate differences in the volatility or risk in returns. The literature also documents that the exact values of returns are in general not predictable, while return volatility is predictable at least to some extent (Andersen et al., 2006). GARCH type of models are also adopted for these data since they model the return volatility explicitly instead of the level of returns (Bollerslev et al., 1986). In the later sections, however, it will be shown that a GARCH model performs worse than the proposed PFS model for the S&P 500 data considering the value at risk estimates.

Note that the underlying structure of the fuzzy histogram, as defined in (5.3), is based on the fuzzified size and distribution of the fuzzy sets C_j. By optimizing only the location of the membership functions at the edges, the distribution of the fuzzy sets over the universe of discourse is never changed. Thus, we may obtain several sets C_j concentrated on the central part of the output density, resulting in pdf's with small variance while fewer sets will be placed on the edges of the output distribution, resulting in fat tails. This is a caveat of this simple type of optimization. Nonetheless, optimizing only the extremes reduces the dimensionality problem, since the number of parameters to be optimized is reduced. Another interesting aspect is that the obtained densities are substantially different from each other. This can be explained with the

Table 5.2: Optimized probability parameters.

	Consequent									
Rule	1	2	3	4	5	6	7	8	9	10
1	0.0107	0.0091	0.0397	0.1417	0.5269	0.2235	0.0328	0.0125	0.0001	0.0030
2	0.0000	0.0000	0.0002	0.1771	0.7851	0.0376	0.0000	0.0000	0.0000	0.0000
3	0.0003	0.0098	0.0236	0.2573	0.4268	0.2236	0.0485	0.0015	0.0087	0.0000
4	0.0000	0.0000	0.0000	0.0268	0.9731	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0049	0.0087	0.0274	0.1622	0.6507	0.1150	0.0226	0.0084	0.0000	0.0000
Mean	0.0032	0.0055	0.0182	0.1530	0.6725	0.1199	0.0208	0.0045	0.0017	0.0006

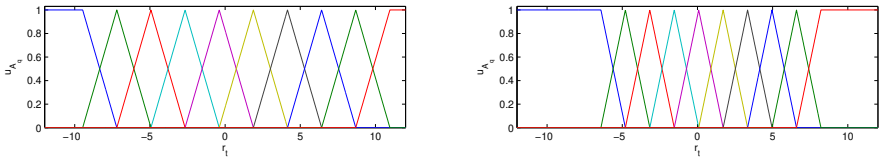
different combinations of covariates regarding previous time periods, which can be interpreted as a summary of information about last period returns, and variation of closing, minimum and maximum returns over the period.

5.5.2 Probabilistic fuzzy output model interpretation

In this approach the rules have the form (5.4). Figure 5.6 shows the output membership functions C_j before and after the optimization. Using this figure, a possible linguistic interpretation of consequents C_j, $j = 1, \ldots, 10$ can be obtained, ranging from very low returns to very high returns, through a 10 point scale. It is visible that the optimized membership functions are now more concentrated towards the center in the return region $[-5, 5]$, as most of the data lies in this region.

Table 5.2 presents the optimized probability parameters $\hat{\mathrm{Pr}}(C_j|A_q)$ of the rules base (5.4). The last row of this table corresponds to the unconditional probabilities for each consequent C_j. These probability values indicate that the overall return distribution is concentrated around point 0 as the highest mean probability is obtained for C_5 corresponding to returns around zero. The unconditional probabilities of very low returns (C_1 to C_3) are higher than those of very high returns (C_8 to C_{10}). Hence, despite being centered around value 0, the returns distribution is positively skewed with negative values occurring more often according to the mean probabilities. This finding is in line with several studies, see e.g. Harvey and Siddique (2000).

The individual rule outputs in Table 5.2 indicate the ability of each rule to capture specific output forms, such as extreme returns. Rule 2 and 4 signal highest probabilities for returns around 0, while rules 1, 3 and 5, capture very low negative returns. This means that the probability of a very low return (C_1) is not zero for these three rules. Note that only rule 1 captures very high returns (C_{10}) since the only positive probability for C_{10} is achieved by this rule. The



(a) Original (b) After optimization

Figure 5.6: Consequent membership functions C_j.

skewness of the distribution of the returns is reflected in the mass of probability given to negative returns (C_1 to C_4) compared to the positive returns (C_6 to C_{10}).

Note that the combination of multi-covariates can be interpreted as a summary of information about last period returns, and variation of closing, minimum and maximum returns over the last periods. Rule 2 and rule 4 are defined by moderate last period returns and moderate cumulated values for past variation, indicating stability in past returns. These rules lead to the most peaked return distributions compared to the remaining rules. The obtained density from these rules indicate 'mean reversion' in returns, i.e. slightly positive or negative returns with mediocre variation result in returns concentrated around 0 with relatively small volatility. Rule 1 and rule 5 on the other hand indicate a relatively high probability of extreme returns. These rules correspond to very low (high) past returns with very low (high) accumulated past variation, indicating that extreme return values or extreme volatility periods are followed by high uncertainty in returns for the next period. Finally rule 3, with antecedents corresponding to past returns around 0 and moderate past variation in returns still leads to a relatively wide return distribution. We conclude that sudden volatility movements are captured with the fat tails of rules 1, 3 and 5 rather than rules 2 and 4.

5.6 Value-at-Risk estimation for the S&P 500 index

In this section, we analyze VaR estimates of the proposed PFS models for S&P 500 data. The returns data and the forecast sample period used for the out-of-sample model evaluation are the same as those in the analysis in Section 5.5. We provide the results of the multi-covariate PFS model applied to the S&P 500 data for one day ahead returns and illustrate the gains from the proposed two-step estimation approach for the PFS model. Furthermore, we apply the proposed PFS models to analyze seasonality and multi-horizon forecasts for these data. These models' performances are compared with respect to alternative models in terms of the VaR estimates.

Table 5.3: Kupiec test. Bold face indicates rejection. Non-rejection regions 16<I(0.95)<36, 6<I(0.975)<21 and 1<I(0.99)<11.

c	GARCH(1,1)	PFS_1-Opt_2	PFS_9-Opt_1	PFS_9-Opt_2	PFS_{Dates}	$PFS_{t+1\&t+30}$ (1 day)	$PFS_{t+1\&t+30}$ (1 month)
$I(0.95)$	32	26	23	33	25	33	33
$I(0.975)$	21	16	13	18	15	**22**	**22**
$I(0.99)$	**12**	9	4	7	6	8	11

Table 5.4: Christoffersen Markov test.

c	π_{ij}	GARCH(1,1)	PFS_1-Opt_2	PFS_9-Opt_1	PFS_9-Opt_2	PFS_{Dates}	$PFS_{t+1\&t+30}$ (1 day)	$PFS_{t+1\&t+30}$ (1 month)
0.950	π_{01}	0.937	0.946	0.955	0.935	0.952	0.933	0.941
	π_{11}	0.938	1.000	0.957	0.939	0.920	1.000	0.903
0.975	π_{01}	0.959	0.968	0.974	0.963	0.972	0.955	0.955
	π_{11}	0.952	1.000	1.000	1.000	0.933	1.000	1.000
0.990	π_{01}	0.976	0.982	0.992	0.986	0.988	0.984	0.978
	π_{11}	1.000	1.000	1.000	1.000	1.000	1.000	1.000

5.6.1 Multi-covariate model results for the S&P 500 index

We consider the proposed multi-covariate model and the simultaneous estimation of the output membership functions edges, mfC_1 and mfC_J and probability parameters $\hat{\Pr}(C_j|A_q)$. The multi-covariate model includes nine covariates, *LastDay*, *LastWeek*, *LastMonth*, $CloseAbs_{95}$, $CloseAbs_{80}$, $CloseSqr_{95}$, $CloseSqr_{80}$, $MaxMin_{95}$ and $MaxMin_{80}$ as defined in Section 5.4. To check the effects of using several covariates, we also consider a PFS model with a single covariate r_t (Xu and Kaymak, 2008; Almeida and Kaymak, 2009a,b). For this model we optimize both the probability parameters and the output membership functions edges. The effects of optimizing the output membership functions edges are considered by considering a multi-covariate model with the optimization of only the probability parameters. Finally, all these models are compared against the popular GARCH $(1,1)$ model, as defined in Section 5.3.2. In summary, the following models are considered:

- PFS_1-Opt_2, PFS, single covariate, optimized $\hat{\Pr}(C_j|A_q)$, mfC_1 and mfC_J.
- PFS_9-Opt_1, PFS, multi-covariate, optimized $\hat{\Pr}(C_j|A_q)$.
- PFS_9-Opt_2, PFS, multi-covariate, optimized $\hat{\Pr}(C_j|A_q)$, mfC_1 and mfC_J.
- GARCH $(1,1)$, the GARCH $(1,1)$ model.

Table 5.3 and Table 5.4 show the results of the unconditional coverage and independence test, respectively, for the probabilistic fuzzy models and the GARCH model. Figure 5.7(a)–(d)

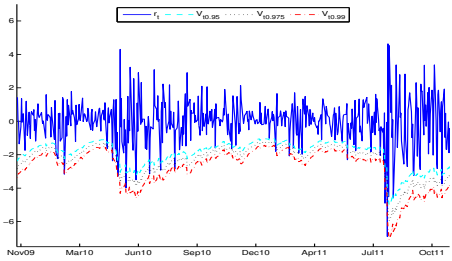
shows the out-of-sample returns r_t and VaR$_t(c)$ for c =0.95, 0.975, 0.99 for all models under consideration.

Table 5.3 shows that the validity of the model is only rejected for GARCH(1,1) at $c = 0.99$. Note that the non-rejection regions for the considered sample are 16<I(0.95)<36, 6<I(0.975)<21 and 1<I(0.99)<11. For GARCH(1,1) models and $c = 0.99$, Table 5.4 shows that the cluster violations are more accentuated than for the remaining models. Furthermore, this table shows that VaR violations do not depend greatly on the occurrence of a VaR violation on the previous day. By visually inspecting Fig. 5.7, it can be seen that the GARCH(1,1) model is slow to capture sudden changes in volatility. The PFS models under consideration, all react to these changes quickly.

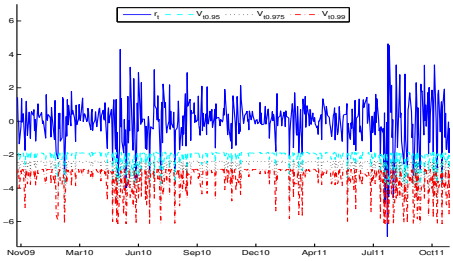
Fig. 5.7 also indicates the complexity involved in analyzing the return distribution in terms of the effects of the covariates. The PFS model with a single covariate of past returns r_t, PFS$_1$-Opt$_2$, cannot accurately capture the changing volatility particularly in high volatility periods, such as the period around June 2010. This model is rather conservative, which translates into a prudent risk measure, at the periods with low volatility, although it does not violate the independence or unconditional coverage tests. The calculated VaR levels are mostly constant, except for singleton peaks around high volatility periods.

Note that the single covariate r_t only provides instant information about volatility, therefore it is a restrictive measure of information on the past returns and past return distributions. In contrast to the model with a single covariate, PFS$_1$-Opt$_2$, the multi-covariate PFS models, PFS$_9$-Opt$_1$ and PFS$_9$-Opt$_2$, show both singleton and periods of high volatility, due to the increased information contained in the extra covariates. Hence, in order to obtain an accurate return distribution, information on past returns and on cumulated past values and volatility in returns should be taken into account. This finding is in line with studies finding long range dependence in stock returns (Andersen and Bollerslev, 1997b) since the added antecedents in these multi-covariate models, the geometrically decaying average of closing prices and the squares of closing prices in (5.19), (5.20) and (5.21), include return information from several past periods with high decay rates, 0.8 and 0.9.

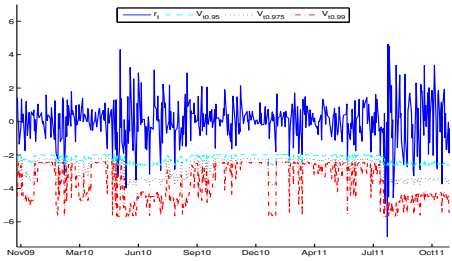
A further consideration is the optimization of the membership parameters in the multi-covariate models PFS$_9$-Opt$_1$ and PFS$_9$-Opt$_2$, where the membership parameters are only optimized in the latter model. The return volatility is accurately captured particularly when the membership parameters are optimized together with the probability parameters in PFS$_9$-Opt$_2$. In lower volatility periods, such as around Dec10, PFS$_9$-Opt$_1$ model is more conservative than PFS$_9$-Opt$_2$ model. This fact can also be seen by inspecting Table 5.3, where PFS$_9$-Opt$_2$ model has more exceptions. Therefore optimizing the membership parameters together with the probability parameters does increase model accuracy.



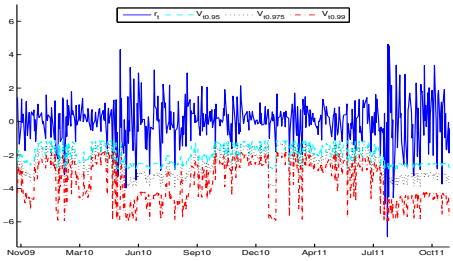
(a) GARCH(1,1)



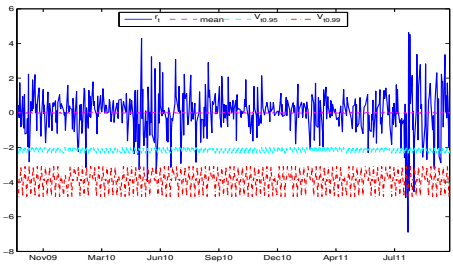
(b) PFS, single covariate r_t



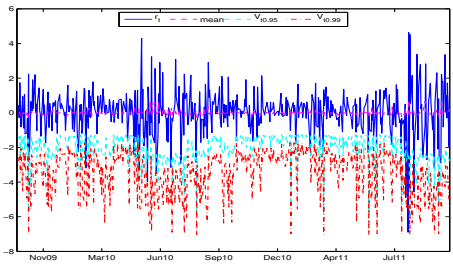
(c) PFS, multi-covariate, optimized $\hat{\mathrm{Pr}}(C_j|A_q)$



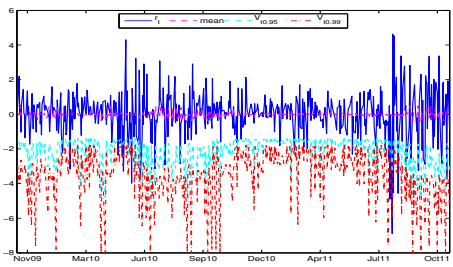
(d) PFS, multi-covariate, optimized $\hat{\mathrm{Pr}}(C_j|A_q)$, mfC_1 and mfC_J



(e) $\mathrm{PFS}_{\mathrm{Dates}}$ forecasts

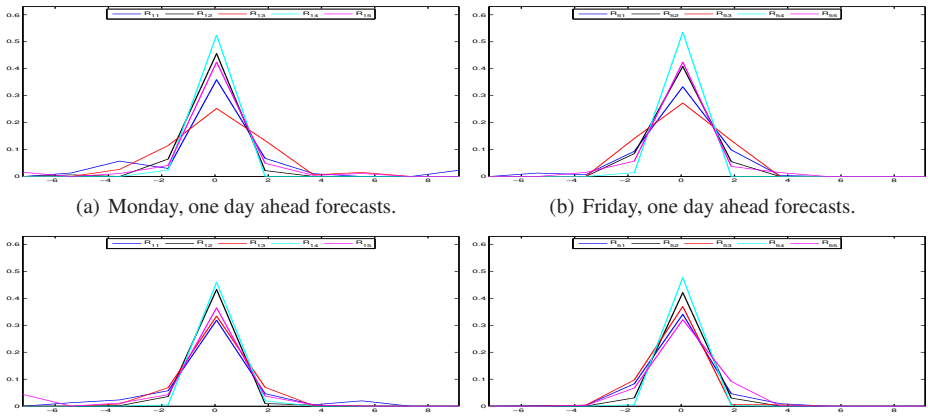


(f) $\mathrm{PFS}_{t+1\&t+30}$ multi-covariate PFS for one day ahead forecasts



(g) $\mathrm{PFS}_{t+1\&t+30}$ multi-covariate PFS for one month ahead forecasts

Figure 5.7: Out-of-sample returns r_t and $\mathrm{VaR}_t(c)$ for c =0.95, 0.99.



(a) Monday, one day ahead forecasts. (b) Friday, one day ahead forecasts.

(c) Monday, one month ahead forecasts. (d) Friday, one month ahead forecasts.

Figure 5.8: Rule outputs for Monday and Friday.

5.6.2 Seasonality analysis for the S&P 500 index

In this section we present the results of the seasonality analysis for the S&P 500 index. The seasonality analysis we perform is based on two PFS models. The first PFS model is developed to analyze possible difference of value-at-risk on different day of the week. This model denoted as $\text{PFS}_{\text{Dates}}$ has a single antecedent variable, day of the week, and a single output, days' returns. The second model we consider includes multiple covariates as explained in Section 5.4 and has two outputs, the one day ahead and one month ahead forecasts, denoted by $\text{PFS}_{t+1\&t+30}$. For each *WeekDay* we consider different partitions of the remaining 9 covariates. The antecedent membership functions for the qualitative covariates are obtained from the FCM algorithm with five clusters, while the seasonal covariate is modelled with a crisp characteristic function. Thus we obtain five rules for each *WeekDay* resulting in a model with 25 rules. The consequents are modelled using 10 triangular membership functions. We first consider the $\text{PFS}_{\text{Dates}}$ model of the form (5.2). In this model each rule corresponds to a different day of the week. Table 5.5 shows the estimated mean and variance for each day density forecast. As Table 5.5 shows, the mean and variances of the estimated probability density function are different for each day. This seems to indicate that seasonal patterns can be observed using a PFS model. Mondays and Tuesdays are the days with highest variation in returns, implying fatter tails in returns, compared to the rest of the days.

To further illustrate the seasonality effect we consider the multi-covariate multi-output model $\text{PFS}_{t+1\&t+30}$, estimated using ten covariates, including *WeekDay*. Figure 5.8 presents the obtained output probability density functions for one day and one month ahead forecasts for *Week-*

Table 5.5: Day of the week effect on returns mean and variances on 1-day ahead returns.

	Monday	Tuesday	Wednesday	Thursday	Friday
$\hat{\eta}_{y\|\mathbf{x}}$	-0.026	0.038	-0.015	-0.027	-0.002
$\hat{\sigma}_{y\|\mathbf{x}}$	2.209	1.998	1.673	1.809	1.494

Table 5.6: Optimized probability parameters for $\text{PFS}_{\text{Dates}}$.

	Consequent									
Rule	1	2	3	4	5	6	7	8	9	10
1	0.007	0.010	0.022	0.124	0.715	0.102	0.012	0.003	0.004	0.002
2	0.000	0.010	0.016	0.162	0.687	0.088	0.025	0.012	0.000	0.002
3	0.006	0.004	0.012	0.152	0.719	0.090	0.014	0.005	0.000	0.000
4	0.004	0.002	0.032	0.152	0.693	0.091	0.025	0.001	0.002	0.000
5	0.000	0.002	0.023	0.146	0.712	0.103	0.012	0.002	0.000	0.000

Day={Monday,Friday}. We recall that for each *WeekDay* the remaining nine covariates are partitioned in five fuzzy sets. Seasonal patterns in the returns distribution is apparent from this figure. The rule outputs for Monday are wider compared to those for Friday, indicating fat tails in the returns distribution on Monday. Such seasonal patterns in returns are also shown in Berument and Kiymaz (2001). This effect is also visible in the out-of-sample value-at-risk forecast presented in Fig. 5.7(e), specially for the $\text{VaR}_{t(0.99)}$ which displays periodic changes over time.

Apart from the obtained output densities for each day, the fuzzy output model in Section 5.2.2 can also be used to gain insight about seasonality. Table 5.6 presents the optimized conditional probability parameters $\hat{\Pr}(C_j|A_q)$ for the $\text{PFS}_{\text{Dates}}$ model, where rules 1 to 5 correspond to *WeekDay*, Monday to Friday. The corresponding optimized output membership parameters for this model are given in Figure 5.9. These membership functions range from very low returns to very high returns, through a 10 point scale, where consequent 5, C_5, corresponds to returns around 0. According to Table 5.6, returns in all days are concentrated around consequent 5. Hence most of the returns are around 0, regardless of the specific *WeekDay*. Mondays have the highest probability of very low returns (C_1), while Tuesdays and Fridays have an almost zero probability of very low returns. Furthermore, returns on Fridays are found to be relatively less volatile, since the probabilities of very low and very high returns (C_1 and C_{10}) are almost 0 for this day. Conditional probabilities of low returns (C_1 to C_3) are higher than those of high returns (C_8 to C_{10}) for all days. Therefore return distributions seem to be positively skewed regardless of the *WeekDay*. However, the probabilities of low and high returns, hence the degree of skewness, differ according to *WeekDay*.

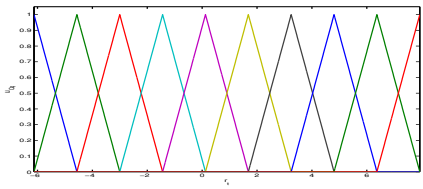


Figure 5.9: Consequent membership functions C_j for $\text{PFS}_{\text{Dates}}$.

5.6.3 Multiple horizon forecasts for the S&P 500 index

We next apply the proposed PFS model with multiple outputs, one day ahead and one month ahead forecast of VaR, to S&P 500 data. The multiple horizon model includes ten antecedents *WeekDay*, *LastDay*, *LastWeek*, *LastMonth*, $\textit{CloseAbs}_{95}$, $\textit{CloseAbs}_{80}$, $\textit{CloseSqr}_{95}$, $\textit{CloseSqr}_{80}$, $\textit{MaxMin}_{95}$ and $\textit{MaxMin}_{80}$ as defined in Section 5.4. An important aspect of such model is its ability to provide suitable long term forecasts. Table 5.3 and Table 5.4 show the results of the unconditional coverage and independence test, respectively, for all the probabilistic fuzzy models. For completeness we include the results for the $\text{PFS}_{\text{Dates}}$ model. Figure 5.7(f) and Fig. 5.7(g) show the out-of-sample returns r_t, estimated mean $\hat{\eta}_{y|\mathbf{x}}$ and value-at-risk $\text{VaR}_t(c)$ for $c = 0.95, 0.99$ for one period and one month period ahead forecasts.

Last three columns of Table 5.3 show that the validity of the model is only rejected for $\text{PFS}_{t+1\&t+30}$ 1 month and 1 day at $c = 0.975$. Note that the non-rejection regions for the considered sample are 16<I(0.95)<36, 6<I(0.975)<21 and 1<I(0.99)<11. Hence the model performs well in both forecast horizons when all antecedent variables are included. Regarding the independence test presented in Table 5.4, for the VaR levels of $c = 0.975, 0.99$ the one-step ahead transition probabilities π_{01} and π_{11} are close to the theoretical values for all model outputs. This indicated that even with only the *WeekDay* antecedent the VaR violations do not depend greatly on the occurrence of a VaR violation on the previous day. However, for the VaR level $c = 0.95$, the obtained transition probabilities are relatively far from the theoretical values. This result may stem from the low explanatory power of the included antecedents specially in longer horizons, which may be solved by including other type of antecedent variables in the PFS model. We leave this topic for future work.

By visually inspecting Fig. 5.7(e), it can be seen that the $\text{PFS}_{\text{Dates}}$ model considering only the *WeekDay* information is quite conservative. The estimated VaR levels are mostly constant, except for cyclic peaks. This covariate only provides information regarding seasonal effects, as discussed previously. In contrast, the multi-covariate PFS models for both horizons show both singleton and periods of high volatility, due to the increased information contained in the extra covariates. For the 1 month ahead forecast presented in Fig. 5.7(e) the model appears to

overestimate the risk, judging by the high $\mathrm{VaR}_t(0.99)$ values. This result is intuitive since the stock market is highly volatile and forecasting small changes in VaR levels is not trivial in long horizons (Baillie and DeGennaro, 1990). Furthermore, it appears that the set of covariates used contain relevant information for accurate 1 day ahead forecast of density returns. As Fig. 5.7(f) shows, the VaR levels follow the real variation of the returns. For both forecast horizons, the mean estimates change slightly over time, although the PFS model mainly captures the returns volatility. Estimated mean returns from the $\mathrm{PFS}_{\mathrm{Dates}}$ model shown in Fig. 5.7(e) have less variation over time compared to the mean returns from the multi-covariate PFS model shown in Fig. 5.7(f). This difference in mean returns indicates that additional antecedent variables such as *LastDay* and *CloseAbs* in the multi-covariate PFS model explain changes in mean returns.

5.7 Conclusions

We have proposed applying multi-covariate probabilistic fuzzy systems for conditional density estimation and value at risk modelling of the S&P 500 index. An extended multi-covariate and multi-output PFS model is introduced. This model provides the VaR estimates, as well as full density estimates, for multiple periods ahead returns data. The additional information, linguistic interpretation and process understanding provided by the different reasoning mechanisms that the PFS model provides are discussed. The proposed models are used to explain seasonal patterns and to obtain one-day and one-month ahead density forecasts of the S&P 500 index. The performance of the proposed models has been compared to the VaR estimation by using the popular GARCH (1,1) volatility estimation. Unconditional coverage test and an independence test are used for this purpose. It is found that the GARCH models are not always accepted, while the PFS models of VaR are never rejected. The proposed multi-covariate PFS models capture both instant volatility changes and periods of high volatility, due to the increased information contained in the extra covariates. The proposed parameter estimation leads to less conservative models that follow the volatility trends accurately. We find that the conditional distribution of returns and associated VaR depend on the day of the week. Furthermore, the extended model captures seasonal patterns in S&P 500 returns in short and longer horizons as well as the increased risk factor in longer term forecasts of returns. As future work, we will investigate in more detail different parameter estimation techniques.

Chapter 6

Estimation of flexible fuzzy GARCH models for conditional density estimation[1]

6.1 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 (Bollerslev, 1986). Models aiming at point forecasts (Araújo, 2010; Cheng et al., 2010) 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 (Jorion, 2006), 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 (Bollerslev, 1986). 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

[1]Parts of this chapter have been published in Almeida, Basturk, Kaymak, and Costa Sousa (2013a,b, 2014b).

data behavior, such as the GJR-GARCH (Glosten et al., 1993) models to capture skewness and Student–t GARCH models to capture fat tails (Bollerslev, 1987).

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 (González-Rivera, 1998) and regime-switching GARCH models (Haas et al., 2004) are proposed. Despite the generality of these model structures, estimation of these models is not trivial and it is impossible to apply standard maximum likelihood estimation, due to the recursive structure of conditional volatility (Bauwens et al., 2010). Artificial neural networks (Donaldson and Kamstra, 1997) and fuzzy systems have also been combined with GARCH models in different forms. In Popov and Bykhanov (2005); Hung and Hsu (2008); Hung (2009a,c,b, 2011b,a), 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 (Geng and Ma, 2008; Chang, 2006, 2008), and rough-set based neuro-fuzzy systems (Das et al., 2010), although in these cases the models are used to approximate either the return series (Chang, 2006, 2008) or realized volatility (Geng and Ma, 2008; Das et al., 2010), 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 [2]. A GARCH model with fuzzy coefficients is presented in Thavaneswaran et al. (2007) and Helin and Koivisto (2011b), where the error term is modelled using a set of fuzzy rules. These models combine fuzzy systems with a statistical model. Similarly, different types of uncertainty are combined in probabilistic fuzzy systems (van den Berg et al., 2013) which are also used to estimate conditional volatility of returns (Almeida and Kaymak, 2009b; Almeida et al., 2012a).

In previous work, we introduced key ideas for a new flexible fuzzy GARCH model for conditional density estimation (Almeida et al., 2013b), and illustrated model performance using simulated data from unimodal GARCH models. This model combines two different types of uncertainty, namely fuzziness or linguistic vagueness, and probabilistic uncertainty. However, a formal description and analysis of this type of systems still needs to be made. The properties,

[2]Forecasting return volatility can be seen as equivalent to forecasting realized volatility (Andersen and Bollerslev, 1997a; Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2002), under the assumptions that high-quality intraday return data are available (noisy intraday data has a detrimental impact on forecast accuracy (Andersen et al., 2011)) and the (log) return series is a continuous semimartingale process (Awartani and Corradi, 2005). For the case of discrete time process, such as GARCH, the assumption of homoskedasticity (all random variables have the same finite variance) must hold (Hansen and Lunde, 2005). If these assumptions are not fulfilled, the point estimation of return volatility will be biased.

estimation issues, model interpretation, differences with previous fuzzy GARCH models and a real data application on S&P 500 return series, which were not considered in Almeida et al. (2013b), are studied in detail in this chapter. In the S&P 500 return series application we show that the linguistic interpretation of this model can provide insight into stylized facts such as volatility clustering and leverage effects. 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 and variance, where the smooth changes are related to linguistic descriptors. Previous fuzzy GARCH models (Popov and Bykhanov, 2005; Hung and Hsu, 2008; Hung, 2009a,c,b, 2011b,a) 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 chapter. 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 (González-Rivera, 1998; Haas et al., 2004).

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 chaper is as follows. Section 6.2 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 6.3 and compared to other fuzzy and probabilistic GARCH mod-

els. 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 6.4, where we also show that the proposed model can successfully capture existing fuzzy GARCH models. In Section 6.5 we present a real world application of the new fuzzy GARCH model for conditional density estimation, and finally, Section 6.6 concludes the chapter.

6.2 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 $t = 1, \ldots, T$ observations is defined as (Bollerslev, 1986):

$$y_t = \sqrt{h_t}\,\epsilon_t \sim NID(0, h_t) \tag{6.1a}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q'} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p'} \beta_j h_{t-j} \tag{6.1b}$$

$$\epsilon_t \sim NID\,(0, 1) \tag{6.1c}$$

where y_t is the data with a conditional normal distribution and $NID(\mu, \sigma^2)$ denotes the normal and independently distribution with mean μ and variance σ^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 $i = 1, \ldots, q'$ and $j = 1, \ldots, p'$ are GARCH model parameters. At each period, the *conditional volatility*, h_t, is assumed to move around the constant *unconditional volatility* $\overline{h}$. 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 ϵ_t in (6.1) 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 (Engle, 1982) 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:

$$\overline{h} = \alpha_0 / \left(1 - \sum_{i=1}^{q'} \alpha_i - \sum_{j=1}^{p'} \beta_j\right) . \tag{6.2}$$

Sufficient conditions for positive variance h_t at every period are:

$$\alpha_0 > 0,\ \alpha_i \geq 0,\ \beta_j \geq 0,\ \sum_{i=1}^{q'} \alpha_i + \sum_{j=1}^{p'} \beta_j < 1, i = 1, \ldots, q',\ j = 1, \ldots, p', \tag{6.3}$$

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: $t^\star = \max(p', q') + 1$. Initial observations $y_1, \ldots, y_{q'}$ can be obtained from the data or set as the unconditional mean of the data. The initial unobserved variances $h_1, \ldots, h_{p'}$ can be set as the unconditional variance of the data. Conditional on these initial values, the likelihood of a single observation is:

$$\ell(y_t \mid I_{t-1}) = \ell(y_t \mid h_t) = \phi(y_t; 0, h_t) \tag{6.4}$$

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

Using the independence assumption in (6.1), the likelihood of the whole sample is obtained by multiplying (6.4) for all t:

$$\ell(y) = \prod_{t=t^\star}^{T} \ell(y_t \mid I_{t-1}) = \prod_{t=t^\star}^{T} \ell(y_t \mid h_t) = \prod_{t=t^\star}^{T} \phi(y_t; 0, h_t), \tag{6.5}$$

where $y = \{y_{t^\star}, \ldots, y_T\}$ and the variance term h_t is obtained recursively using the equality in (6.1b).

In order to obtain the parameter estimates, the likelihood in (6.5) 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 (Zivot, 2009). 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 (6.1) 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 (Zivot, 2009).

Two commonly used extensions are the Student-t GARCH model (Bollerslev, 1987) and the GJR GARCH model (Glosten et al., 1993), 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 (James Chu et al., 1996), compared to the linear variance model assumed by the GARCH model (Haas et al., 2004). 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. These models propose K separate GARCH models (Klaassen, 2002):

$$\begin{aligned} y_t &= \sqrt{h_t}\epsilon_t \sim NID(0, h_t), \\ h_t &= h_{t,k}, \text{ if } s_t = k, \ k = 1, \ldots, K, \\ h_{t,k} &= \alpha_{0,k} + \textstyle\sum_{i=1}^{q'} \alpha_{i,k} y_{t-i}^2 + \sum_{j=1}^{p'} \beta_{j,k} h_{t-j}, \text{ for } k = 1, \ldots, K, \\ \epsilon_t &\sim NID\left(0, 1\right), \end{aligned} \tag{6.6}$$

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. Our 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 Popov and Bykhanov (2005); Hung (2009a,c, 2011b,a). This type of models consists 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 (Popov and Bykhanov, 2005; Hung, 2009a,c, 2011b,a):

$$R_l : \text{If } \mathbf{x} \text{ is } F_l \text{ then } h_{t,l} = \alpha_{0,l} + \sum_{i=1}^{q'} \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^{p'} \beta_{j,l} h_{t-j}\,, \tag{6.7}$$

where $\mathbf{x} \in \mathbb{R}^n$ is an input vector, $F_l : X \longrightarrow [0, 1]$ is a multidimensional fuzzy set defined on a continuous sample space X. The output of this fuzzy model is presented as

$$y_t = \sqrt{h_t}\epsilon_t \,, \tag{6.8a}$$

$$h_t = \sum_{l=1}^{L} g_{t,l} h_{t,l} \,, \tag{6.8b}$$

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 $l = 1, \ldots, L$, $\sum_{l=1}^{L} u_{t,l} > 0$, and by definition $g_{t,l} \geq 0$ and $\sum_{l=1}^{L} g_{t,l} = 1$. The inference used for the output (6.8a) and (6.8b) is similar to the inference of a Takagi-Sugeno fuzzy model (Takagi and Sugeno, 1985). Although not clear in Hung (2009a,c, 2011b,a), we assume that like in Popov and Bykhanov (2005), 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.,*

$$u_{t,l}(\mathbf{x}) = u_{F_{l_1}}(x_1) \circ \cdots \circ u_{F_{l_n}}(x_n), \tag{6.9}$$

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 (6.7) provides the *unobserved* conditional variance h_t. The density of output y_t is based on h_t:

$$y_t \mid h_t, x_t \sim NID(\mu, h_t) \,. \tag{6.10}$$

In Popov and Bykhanov (2005), the parameters of the model in (6.7) 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 (6.7) a GARCH(3,3) model or by constraint of $\beta_j = 0$ using a GARCH(0,5) model. For the real data example (Popov and Bykhanov, 2005), 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

$$R_l : \text{If } y_{t-1} \text{ is } F_l \text{ then } h_{t,l} = h_l. \tag{6.11}$$

In Hung (2009a,c), the parameters of the fuzzy GARCH model are obtained using a genetic algorithm, while in Hung (2011b) 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

$$E_1 = \sum_{t=1}^{T} (y_t^* - y_t)^2 \,. \tag{6.12}$$

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 Hung (2009b), although the rule base model is different. Furthermore, in Hung (2011a) a similar objective function, based on the squared returns, is defined:

$$E_2 = \sum_{t=1}^{T} (y_t^{*2} - y_t^2)^2 \,, \tag{6.13}$$

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 (Andersen and Bollerslev, 1998). More accurate results can be obtained using the sum of squared intraday results (Andersen and Bollerslev, 1997a; Andersen et al., 2001; Barndorff-Nielsen and Shephard, 2002; Andersen et al., 2011).

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 (6.10) 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 6.3 which can be estimated using a maximum likelihood approach.

6.3 Proposed fuzzy GARCH model

In this chapter we present a new flexible fuzzy GARCH model (Almeida et al., 2013b), named FGARCH(L,p',q') where the output y_t and conditional variance h_t are defined by each of l-th fuzzy rule

$$R_l : \text{If } \mathbf{x} \text{ is } F_l \text{ then } y_{t,l} \mid x_t, h_{t,l} \sim NID(\mu_l, h_{t,l}) \,, \tag{6.14a}$$

$$\text{with } h_{t,l} = \alpha_{0,l} + \sum_{i=1}^{q'} \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^{p'} \beta_{j,l} h_{t-j} \,, \tag{6.14b}$$

where h_t is given by (6.8b) 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 μ_l and variance $h_{t,l}$, with parameters given by:

$$\begin{aligned} y_{t,l} &= \sqrt{h_{t,l}}\,\epsilon_t\,, \\ h_{t,l} &= \alpha_{0,l} + \sum_{i=1}^{q'} \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^{p'} \beta_{j,l} h_{t-j}\,, \\ h_{t-j} &= \sum_{l=1}^{L} g_{t-j,l} h_{t-j,l}\,, \text{ for } j = 1\ldots,p'\,. \end{aligned} \tag{6.15}$$

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

The output of this FGARCH model is:

$$y_t \mid h_t, x_t \sim \sum_{l=1}^{L} g_{t,l} NID(\mu_l, h_{t,l})\,, \tag{6.16}$$

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 (6.16) and the output (6.8), it is clear that the outputs follow different inference mechanisms. In (6.8) the output is a probabilistic normal distribution with zero mean and the variance modelled with a fuzzy system. Conversely, the output (6.16) 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 membership values satisfy

$$g_{t,l} \geq 0, \forall l, t \tag{6.17a}$$

$$\sum_{l=1}^{L} g_{t,l} = 1, \forall t\,. \tag{6.17b}$$

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

(6.14a) are not defined otherwise. Inserting (6.14b) in (6.8b), variance h_t at time t is:

$$h_t = \sum_{l=1}^{L} g_{t,l} \left(\alpha_{0,l} + \sum_{i=1}^{q} \alpha_{i,l} y_{t-i}^2 + \sum_{j=1}^{p} \beta_{j,l} h_{t-j} \right) \tag{6.18a}$$

$$= \sum_{l=1}^{L} g_{t,l} \alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} y_{t-i}^2 + \sum_{l=1}^{L} g_{t,l} \sum_{j=1}^{p} \beta_{j,l} h_{t-j} \,. \tag{6.18b}$$

A sufficient condition to ensure a positive and finite expected variance in (6.18) is to incorporate standard GARCH model conditions for each rule $l = 1, \ldots, L$ in the model:

$$\alpha_{0,l} > 0,\ \alpha_{i,l} \geq 0,\ \beta_{j,l} \geq 0,\ \sum_{i=1}^{q'} \alpha_{i,l} + \sum_{j=1}^{p'} \beta_{j,l} < 1, i = 1, \ldots, q',\ j = 1, \ldots, p' \,, \tag{6.19}$$

as the following theorems show.

Theorem 6.1 *For the FGARCH(L,p′,q′) model defined by* (6.14)*, the variance* h_t *given by* (6.18) *is positive for each* t *if* (6.19) *is satisfied,* $g_{t,l} \geq 0$*,* $\forall t, l$*,* $\sum_{l=1}^{L} g_{t,l} = 1$*,* $\forall t$ *and initial variances* $h_1, \ldots h_{t^\star - 1}$ *are set as positive values according to the definition of variance.*

Proof: Given that $g_{t,l} \geq 0$, $\forall t, l$, $\sum_{l=1}^{L} g_{t,l} = 1$, $\forall t$ and initial variances $h_1, \ldots h_{t^\star - 1}$ are set as positive values we obtain

$$\begin{aligned} h_t &= \sum_{l=1}^{L} g_{t,l} \alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} y_{t-i}^2 + \sum_{l=1}^{L} g_{t,l} \sum_{j=1}^{p} \beta_{j,l} h_{t-j} \\ &\geq \sum_{l=1}^{L} g_{t,l} \min_{l}(\alpha_{0,l}) + \sum_{l=1}^{L} g_{t,l} \min_{i,l}(\alpha_{i,l}) y_{t-i}^2 + \sum_{l=1}^{L} g_{t,l} \min_{j,l}(\beta_{j,l}) h_{t-j} \,. \end{aligned} \tag{6.20}$$

Note that the positivity of initial variances $h_1, \ldots h_{t^\star - 1}$ ensures that $h_{t-j} > 0$ for $j = 1, \ldots, p$.

Furthermore, since $\min_l(\alpha_{0,l}) > 0$ and $\min_{i,l}(\alpha_{i,l}) = \min_{j,l}(\beta_{j,l}) = 0$ we obtain

$$\begin{aligned} h_t &\geq \sum_{l=1}^{L} g_{t,l} \min_{l}(\alpha_{0,l}) + \sum_{l=1}^{L} g_{t,l} \min_{i,l}(\alpha_{i,l}) y_{t-i}^2 + \sum_{l=1}^{L} g_{t,l} \min_{j,l}(\beta_{j,l}) h_{t-j} \\ &> \sum_{l=1}^{L} g_{t,l} \min_{i,l}(\alpha_{i,l}) y_{t-i}^2 + \sum_{l=1}^{L} g_{t,l} \min_{j,l}(\beta_{j,l}) h_{t-j} = 0 \,. \end{aligned} \tag{6.21}$$

Given (6.20) and (6.21), the variance term at the initial time $t^\star$ is positive, and the consequent variances are also positive.

■

Theorem 6.2 *For the FGARCH(L,p′,q′) model defined by* (6.14)*, the expectation of the variance* h_t *given by* (6.18) *is finite.*

Proof: Defining the unconditional expectation of the variance term as $\bar{h} = E(h_t)$, $\forall t$, using the definition $E(h_{t-i}) = E\left(y_{t-i}^2\right) - E\left(\mu_{t-i}^2\right)$ and (6.18b), we obtain

$$\begin{aligned}\bar{h} = E(h_t) &= \sum_{l=1}^{L} g_{t,l}\alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} E\left(y_{t-i}^2\right) + \sum_{l=1}^{L} g_{t,l} \sum_{j=1}^{p} \beta_{j,l} E(h_{t-j}) \\ &= \sum_{l=1}^{L} g_{t,l}\alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} \left(\bar{h} + E\left(\mu_{t-i}^2\right)\right) + \sum_{l=1}^{L} g_{t,l} \sum_{j=1}^{p} \beta_{j,l}\bar{h} \\ &= \frac{\sum_{l=1}^{L} g_{t,l}\alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} E\left(\mu_{t-i}^2\right)}{1 - \sum_{l=1}^{L} g_{t,l}\left(\sum_{i=1}^{q} \alpha_{i,l} + \sum_{j=1}^{p} \beta_{j,l}\right)} . \end{aligned} \tag{6.22}$$

As a result of theorem 6.1, $h_t > 0, \forall t \Rightarrow \bar{h} > 0$.

The numerator of (6.22) is finite since

$$E\left(\mu_{t-i}^2\right) < \infty, \forall i, t \tag{6.23}$$

and model parameters are constrained by definition (6.19)

$$\begin{aligned}\sum_{l=1}^{L} g_{t,l}\alpha_{0,l} + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \alpha_{i,l} E\left(\mu_{t-i}^2\right) &< \sum_{l=1}^{L} g_{t,l} \max_{l}(\alpha_{0,l}) + \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} \max_{i,l}(\alpha_{i,l}) E\left(\mu_{t-i}^2\right) \\ &= \max_{l}(\alpha_{0,l}) + \max_{i,l}(\alpha_{i,l}) \sum_{i=1}^{q} E\left(\mu_{t-i}^2\right) < \max_{l}(\alpha_{0,l}) + E\left(\mu_{t-i}^2\right) < \infty . \end{aligned} \tag{6.24}$$

Due to (6.19) and $\sum_{l=1}^{L} g_{t,l} = 1, \forall t$ the denominator of (6.22) is also finite

$$1 - \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} (\alpha_{i,l} + \beta_{j,l}) \leq 1 - \sum_{l=1}^{L} g_{t,l} \min_{l} \left(\sum_{i=1}^{q} \alpha_{i,l} + \sum_{j=1}^{p} \beta_{j,l}\right) \tag{6.25}$$

$$= 1 - \min_{l} \left(\sum_{i=1}^{q} \alpha_{i,l} + \sum_{j=1}^{p} \beta_{j,l}\right) < \infty , \tag{6.26}$$

and non-zero

$$1 - \sum_{l=1}^{L} g_{t,l} \sum_{i=1}^{q} (\alpha_{i,l} + \beta_{j,l}) \geq 1 - \sum_{l=1}^{L} g_{t,l} \max_{l} \left(\sum_{i=1}^{q} \alpha_{i,l} + \sum_{j=1}^{p} \beta_{j,l} \right)$$
$$= 1 - \max_{l} \left(\sum_{i=1}^{q} \alpha_{i,l} + \sum_{j=1}^{p} \beta_{j,l} \right) > 0 \,. \tag{6.27}$$

According to (6.23), (6.24), (6.25) and (6.27), the unconditional expected variance in (6.22) is positive and finite. ■

The output density function of FGARCH(L,p',q') model defined by (6.14) is stationary since, according to theorem 6.1 and theorem 6.2, the variance h_t is positive and the expectation of the variance is finite.

Note that these conditions should also hold for the fuzzy GARCH models proposed in the literature (Popov and Bykhanov, 2005; Hung, 2009a,c, 2011b,a), although they have not been explicitly considered.

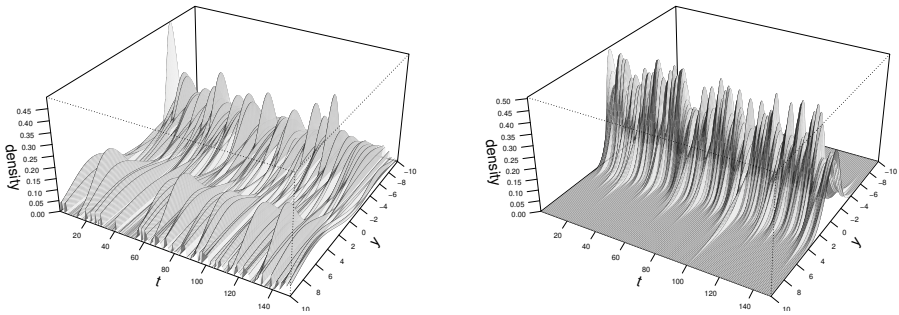
6.3.1 Interpretation of the model

Intuitively, the conditional distribution of the proposed model in (6.16) 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 (6.14a) and (6.14b), 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 (6.10) 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:

1. If $\mu_l = \mu_{l^\star}$ for all $l, l^\star \in \{1, \ldots, L\}$, 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 (6.10).

2. If mean parameters μ_l are relatively different and $h_{t,l}$ are relatively small and similar across $l = \{1, \ldots, L\}$, output distribution is likely to be multimodal.



(a) FGARCH(2,1,1) model

(b) Model proposed in Popov and Bykhanov (2005); Hung (2009c).

Figure 6.1: Conditional distributions of simulated data from fuzzy GARCH models.

3. If mean parameters μ_l are relatively close to each other and $h_{t,l}$ are relatively different across $l = \{1, \ldots, L\}$, output distribution is likely to be skewed.

We illustrate the difference between the fuzzy GARCH model defined in Popov and Bykhanov (2005); Hung (2009c) and the proposed FGARCH(L,p',q'), using simulated data. Figure 6.1 shows the conditional density of output y for simulated data from the proposed model (6.14a) and model (6.7). In this example, both models have two rules with Gaussian membership parameters $\{c_{1,1}, s^2_{1,1}, c_{1,2}, s^2_{1,2}\} = \{-2.3, 2.5, 1, 1\}$ defined in (6.29) 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 (Popov and Bykhanov, 2005; Hung, 2009c) $\mu = 0$, while in the FGARCH(2,1,1) model $(\mu_1, \mu_2) = (-6, 6)$.

The model in (6.7) 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 (Zadeh, 1975), 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 (Hachicha et al., 2011), with imprecise descriptions of phenomena in a similar way humans do it. Following the concept of granularity (Zadeh, 1979, 2008), 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 (Zadeh, 1997), where each linguistic label has a degree of valid-

ity. The probabilistic uncertainty can be captured by the GARCH model. In this extensively studied and good performing model (Hansen and Lunde, 2005), 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 (Geweke and Keane, 2007). 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).

6.3.2 Parameter estimation

It is possible to estimate the model in (6.14a) 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 $y = \{y_{t^\star}, \ldots, y_T\}$ is:

$$\ln \ell(y \mid I_{t-1}) = \ln \prod_{t=t^\star}^{T} \ell(y_t \mid x_t, h_t) = \sum_{t=t^\star}^{T} \ln \left(\sum_{l=1}^{L} g_{t,l} \phi(y_t; \mu_l, h_{t,l}) \right), \tag{6.28}$$

where h_t is calculated from (6.14b), $t^\star = \max(p', q') + 1$ and initial variances $\{h_1, \ldots, h_{t^\star-1}\}$ are assumed to be known. In practice, $\{h_1, \ldots, h_{t^\star-1}\}$ can be set as the unconditional data variance.

In order to calculate the likelihood in (6.28) it is necessary to specify suitable membership functions, that satisfy conditions (6.17a) and (6.17b). In this work, the FGARCH models considered use Gaussian membership functions of the form (Jang and Sun, 1997):

$$u_{t,l} = u_{t,l}(x_t) = \prod_{k=1}^{n} \exp\left(-\frac{1}{2} \frac{(x_{kt} - c_{k,l})^2}{s_{k,l}^2} \right). \tag{6.29}$$

These membership functions were chosen because they naturally satisfy conditions (6.17a) and (6.17b) since $g_{t,l} = u_{t,l} / \sum_{l=1}^{L} u_{t,l}$, $u_{t,l} \geq 0$ for $l = 1, \ldots, L$, $\sum_{l=1}^{L} u_{t,l} > 0$. This re-

duces the need for additional parameter constraints in the gradient search optimization of the maximum log-likelihood estimation (6.30).

The parameter estimates can be obtained by maximizing the log-likelihood in (6.28), using gradient search methods. We maximize the log-likelihood with respect to the GARCH parameters $\theta_{g,l} = \{\alpha_{0,l},\ \alpha_{1,l},\ \ldots,\ \alpha_{q',l},\ \beta_{1,l},\ \ldots,\ \beta_{p',l}\}$, the output mean for each rule μ_l and the membership function parameters $\theta_{u,l}$, simultaneously. The optimization problem can be defined as:

$$\begin{aligned}
\underset{\mu_l,\theta_{g,l},\theta_{u,l}}{\text{minimize}} \quad & -\log \ell(y \mid I_{t-1}) = -\sum_{t=t^\star}^{T} \ln\left(\sum_{l=1}^{L} g_{t,l}\phi(y_t;\mu_l,h_{t,l})\right) \\
\text{subject to} \quad & \alpha_{0,l} > 0,\ \alpha_{i,l} \geq 0,\ \beta_{j,l} \geq 0,\ i = 1,\ldots,q',\ j = 1,\ldots,p', \\
& \sum_{i=1}^{q'} \alpha_{i,l} + \sum_{j=1}^{p'} \beta_{j,l} < 1,\ i = 1,\ldots,q',\ j = 1,\ldots,p', \\
& c_l \leq c_{l+1},\ l = 1,\ldots,L-1,
\end{aligned} \tag{6.30}$$

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 chapter. 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 (Hansen and Lunde, 2005). One exception to this is the asymmetric GARCH models (Awartani and Corradi, 2005) which are naturally considered in a FGARCH(L,1,1) model.

6.4 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 6.3.2, 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 (6.7) and (6.8), used *e.g.* in Hung (2011b), which restricts the output density to a normal conditional distribution with zero mean. The obtained data distribution is symmetric with fat tails. We show that the FGARCH model proposed in this chapter can correctly capture the properties of this data. Second, we consider simulated data from the proposed FGARCH model defined by (6.14a) and (6.14b). The obtained data distribution is asymmetric, multimodal and has fat tails. 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 $\theta_{g,l} = \{\alpha_{0,l},\ \alpha_{1,l},\ \ldots,\ \alpha_{q',l},\ \beta_{1,l},\ \ldots,\ \beta_{p',l}\}$, the output mean for each rule μ_l, and the Gaussian membership parameters $\theta_{u,l} = \{c_l, s_l^2\}$ for $l = 1, \ldots, L$, simultaneously, as defined in (6.30).

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 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.

6.4.1 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 (6.7) and (6.8). The obtained data distribution is symmetric with fat tails, similar to the conditional distribution presented in Fig. 6.1(b). We estimate the parameters of the proposed FGARCH model defined by (6.14a) and (6.14b), 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

Table 6.1: 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)

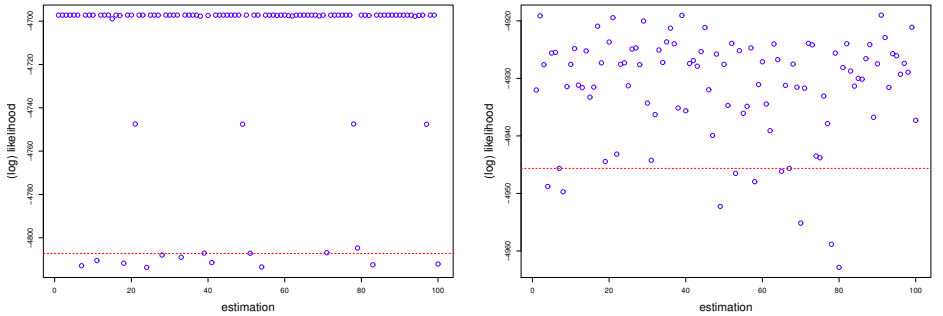
Table 6.2: Parameter estimates and true values for simulated data from a FGARCH(L,1,1) models restricted to a normal conditional distribution with zero mean. 90% intervals from 100 random initializations are given in parentheses.

FGARCH(2,1,1)				
		$l = 1$		$l = 2$
	value	estimate	value	estimate
μ	0.00	0.03 (-0.02, 0.04)	0.00	-0.02 (-0.03, 0.02)
α_0	0.50	0.47 (0.42, 0.66)	1.00	1.00 (0.62, 1.01)
α_1	0.25	0.29 (0.26, 0.53)	0.50	0.58 (0.41, 0.65)
β_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
μ	0.00	-0.02 (-1.09, 0.80)	0.00	-0.09 (-0.82, 0.63)	0.00	0.03 (-0.46, 0.53)
α_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)
α_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)
β_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)

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 6.1, for FGARCH(2,1,1) and FGARCH(3,1,1) models.

Table 6.1 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 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.



(a) FGARCH(2,1,1) (b) FGARCH(3,1,1)

Figure 6.2: FGARCH model with 0 mean: Log-likelihood values form different starting values, and 90% lower band for obtained log-likelihood.

Table 6.2 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 6.2 shows, the effect of the initial points on parameter estimation is not negligible. Estimates of the GARCH parameters $(\alpha_0, \alpha_1, \beta_1)$ and rule output means μ_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 6.1.

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 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 suffer 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

Table 6.3: 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)

parameters lead to similar output density approximations. Hence the optimization of the fuzzy membership parameters is more sensitive to the initial points for optimization.

Figure 6.2 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 maximum 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, 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.

6.4.2 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. The obtained data distribution is asymmetric, multimodal and has fat tails, similar to the conditional distribution presented in Fig. 6.1(a). Table 6.3 shows the mean estimates and 90% intervals of the quantiles $\hat{\tau}(c)$ and Table 6.4 presents the true parameter values, together with mean estimates and 90% interval of parameter estimates (in parentheses).

Similar to the results in Section 6.4.1, we observe that the estimated output densities capture the tails of the distribution, as Table 6.3 shows. Furthermore, as it is presented in Table 6.4, 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

Table 6.4: 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
μ	-2.00	-0.51 (-1.65, 1.71)	2.00	0.67 (-1.86, 1.71)
α_0	0.50	0.70 (0.60, 1.05)	1.00	0.89 (0.61, 0.86)
α_1	0.25	0.28 (0.19, 0.45)	0.50	0.37 (0.15, 0.45)
β_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
μ	-2.00	-0.19 (-1.18, 1.14)	0.00	-0.05 (-0.82 0.97)	2.00	0.72 (-0.45 1.64)
α_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)
α_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)
β_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)

in the obtained 90% intervals for the parameter estimates than obtained in Section 6.4.1. 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 model 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 6.4 indicate.

The problem of local optima can be a more severe problem when the 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 chapter achieves good approximation properties, as Table 6.1 and Table 6.3 show, even using a small number of rules and the simple GARCH(1,1) model, as the models described in these experiments.

6.5 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

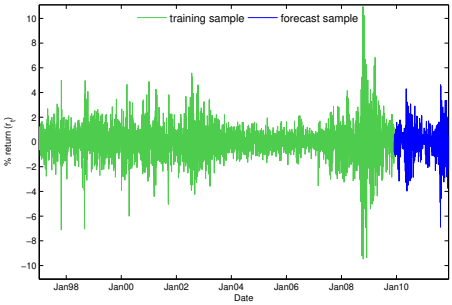


Figure 6.3: S&P 500 returns in percentages.

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 (Huang, 2007). 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 (Jorion, 2006).

In this chapter, 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. 6.3. 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 (Mandelbrot, 1963) 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 (Engle and Ng, 1993) refers to the asymmetric relation between lagged unexpected returns and

Table 6.5: 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.

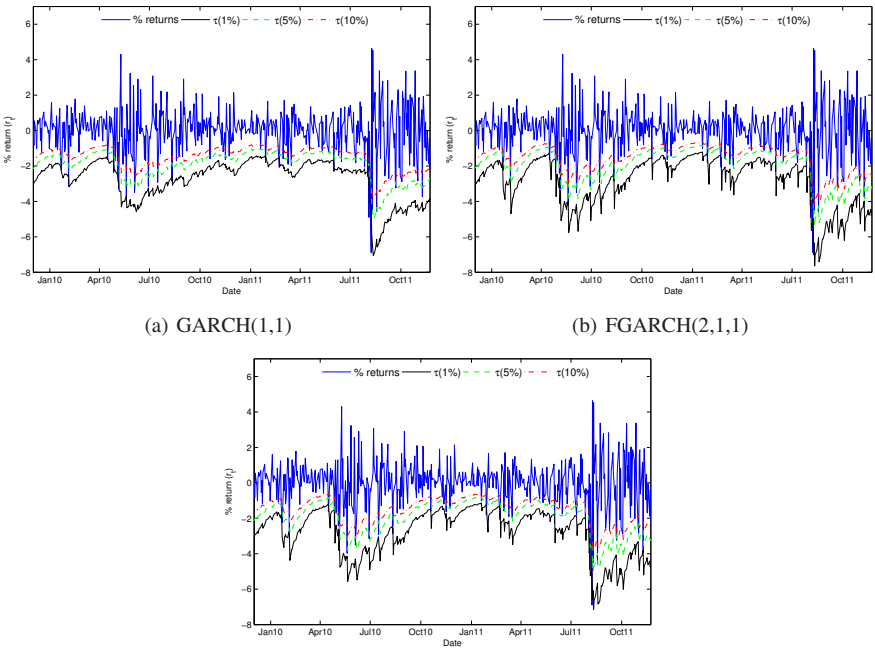
Standard GARCH(1,1)*					
sample	$\tau(1\%)$	$\tau(5\%)$	$\tau(10\%)$	$\tau(20\%)$	$\tau(40\%)$
training	1.6	5.0	9.6	18.3	36.3
	(1.6, 1.6)	(5.0, 5.0)	(9.6, 9.6)	(18.3, 18.3)	(36.3, 36.3)
forecast	2.4	6.4	9.8	17.0	32.4
	(2.4, 2.4)	(6.4, 6.4)	(9.8, 9.8)	(17.0, 17.0)	(32.4, 32.4)
FGARCH(2,1,1)					
sample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

* The differences in estimated values are negligible, with the reported digits.
^{b} Best estimation providing the highest log-likelihood value.

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 6.5 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 6.4 shows the forecast sample and the estimated conditional density quantiles from the GARCH and FGARCH models.

Table 6.5 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 differ-



(a) GARCH(1,1)

(b) FGARCH(2,1,1)

(c) FGARCH(3,1,1)

Figure 6.4: 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.

ent 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 6.4, 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 (Villani et al., 2009; Hachicha et al., 2011; Almeida et al., 2012a; Zheng and Chen, 2013). This topic is left for future research. Despite this overestimation in the tails, by visually inspecting Fig. 6.4, 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 van den Berg et al. (2013); Almeida et al. (2012a), who show that the 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 (6.2). The rule-based of the FGARCH(2,1,1) is given by

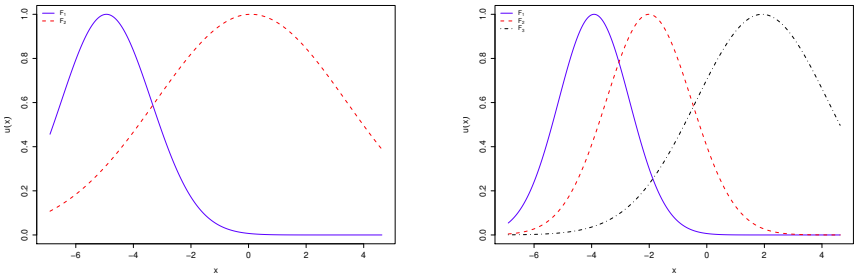
$$\begin{aligned}
R_1 \quad &: \text{If } \mathbf{y}_{t-1} \text{ is } F_1 \text{ then } y_{t,1} \mid y_{t-1}, h_{t,1} \sim N(-0.255, h_{t,1}) \\
&\quad \text{with } h_{t,1} = 3.247 + 0.029 y_{t-1}^2 + 0.970 h_{t-1}\,,\ \sqrt{\bar{h}_1} = 54.2645\,, \\
&\quad \text{and } c_1 = -4.936, s_1^2 = 1.565, \\
R_2 \quad &: \text{If } \mathbf{y}_{t-1} \text{ is } F_2 \text{ then } y_{t,2} \mid y_{t-1}, h_{t,2} \sim N(0.016, h_{t,2}) \\
&\quad \text{with } h_{t,2} = 0.000 + 0.000 y_{t-1}^2 + 0.922 h_{t-1}\,,\ \sqrt{\bar{h}_2} = 0.0007 \\
&\quad \text{and } c_2 = 0.083, s_2^2 = 3.306.
\end{aligned} \tag{6.31}$$

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

$$\begin{aligned}
R_1 \quad &: \text{If } \mathbf{y}_{t-1} \text{ is } F_1 \text{ then } y_{t,1} \mid y_{t-1}, h_{t,1} \sim N(-0.102, h_{t,1})\,, \\
&\quad \text{with } h_{t,1} = 2.768 + 0.019 y_{t-1}^2 + 0.981 h_{t-1}\,,\ \sqrt{\bar{h}_1} = 199.9176 \\
&\quad \text{and } c_1 = -3.916, s_1^2 = 1.234, \\
R_2 \quad &: \text{If } \mathbf{y}_{t-1} \text{ is } F_2 \text{ then } y_{t,2} \mid y_{t-1}, h_{t,2} \sim N(0.104, h_{t,2})\,, \\
&\quad \text{with } h_{t,2} = 0.002 + 0.000 y_{t-1}^2 + 1.000 h_{t-1}\,,\ \sqrt{\bar{h}_2} = 6.9384\,, \\
&\quad \text{and } c_2 = -2.010, s_2^2 = 1.489, \\
R_3 \quad &: \text{If } \mathbf{y}_{t-1} \text{ is } F_3 \text{ then } y_{t,3} \mid y_{t-1}, h_{t,3} \sim N(-0.034, h_{t,3}) \\
&\quad \text{with } h_{t,3} = 0.002 + 0.000 y_{t-1}^2 + 0.887 h_{t-1}\,,\ \sqrt{\bar{h}_3} = 0.1421 \\
&\quad \text{and } c_3 = 1.920, s_3^2 = 2.293.
\end{aligned} \tag{6.32}$$

The estimated membership values for both models are presented in Fig. 6.5. For comparison purposes, the GARCH(1,1) is given by

$$\begin{aligned}
& y_t \mid y_{t-1}, h_t \sim N(0.005, h_t)\,, \\
\text{with} \quad & h_t = 0.013 + 0.077 y_{t-2}^2 + 0.917 h_{t-1}\,, \text{ and } \sqrt{\bar{h}} = 1.433\,.
\end{aligned} \tag{6.33}$$



(a) FGARCH(2,1,1) (b) FGARCH(3,1,1)

Figure 6.5: 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 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 (6.33) 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 each rule's 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 (6.31), 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. 6.5 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 α_1 is lower than in the GARCH(1,1) model presented in (6.33), since the effect of the past returns in variance is already modeled by the fuzzy antecedents in the rule-base model.

An indication of the existence of volatility clustering and volatility persistence (Andersen and Bollerslev, 1997a) 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 (6.33), except for rule 3 of FGARCH(3,1,1). This rule indicates that the effect of conditional volatility is lower when past returns are above 2%. Rule 1 of FGARCH(3,1,1) model (6.32) captures extreme negative events followed by very high volatility the next day. Mean returns in these volatile days are 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. 6.5, this rule is always combined with the other two rules. Rule 1 of the FGARCH(2,1,1) model presented in (6.31), 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 will vary between the unconditional volatility of each rule.

6.6 Conclusion

This chapter 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 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.

Chapter 7

Summary and Conclusions

This thesis shows that models combining fuzzy and probabilistic representations of uncertainty are useful in approximating complex conditional densities while providing a parsimonious and linguistic description of the dynamic behaviour of the system. The models proposed in this thesis, namely the probabilistic fuzzy system (PFS) and fuzzy GARCH models, require very few assumptions regarding the functional form of the estimated density or changes across the space of covariates. These models possess good approximation capabilities and provide a simple interpretation essential for process understanding. It is shown that the estimation of these models can be performed by adopting the standard tools for model estimation, such as the maximum likelihood estimation and least mean squares. Particular attention is given to the interpretation of the models such that they can be useful in many fields such as macroeconomic analysis, quantitative finance and risk management. In this thesis, these models are shown to be useful to model non-linear relations without strict assumptions where regression density estimation is the goal of the analysis.

7.1 General findings

We provide a formal description of probabilistic fuzzy systems. These systems take probabilistic nature of uncertainty into account and also the fuzzy uncertainty through their fuzzy partitioning of input and output spaces. Two possible and equivalent reasoning mechanisms are presented, which lead to two different interpretations of this type of systems. We discuss an additive reasoning scheme for probabilistic fuzzy systems that leads to the estimation of conditional probability densities. We analyse the necessary conditions for a PFS, such that the estimated output density is a proper probability density function and subsequent higher moments derived from this density exist. We consider the relation of probabilistic fuzzy system with different types of deterministic systems that have universal approximation capability. This

relation indicates that a PFS is also suitable for problems of function approximation. Furthermore, we show that PFS can be used to estimate conditional densities of multiple outputs.

In this thesis we propose 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. Furthermore, the use of GARCH models in rule outputs, allows the system to capture time dependency in the conditional distributions in a flexible way. 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 illustrate the model capabilities using synthetic datasets exhibiting different data properties and also to real work problems. We apply the PFS and fuzzy GARCH model on S&P 500 returns. We show that the proposed models are suitable for analysis of the returns distribution, as they capture the properties of the underlying data distribution in all cases we consider. Particular relevance is given to the interpretation of these models and its use in the study of stylized facts, such as seasonality and volatility clustering. The PFS is also applied to the US inflation data. The system shows that slowly changing patterns in inflation are accurately captured by the PFS model. Application of PFS in multi-horizon estimation of quarterly U.S. inflation, which provides point estimates as well as the density estimates of inflation, is relevant for a comprehensive analysis of inflation, particularly for policy making. The PFS model performs well in one period ahead and 1 year ahead forecasts of inflation. The model is also successful in capturing the deflationary pressure during the recent crisis.

7.2 Conclusions

Based on the results presented in this thesis, we can conclude that probabilistic fuzzy systems and fuzzy GARCH models can successfully approximate conditional probability density functions, using a linguistic link between variables. As such, they deal explicitly with both the fuzziness in the linguistic descriptions and the probabilistic uncertainty in the output density. Furthermore, probabilistic fuzzy systems can also be successfully applied to problems of function approximation since they are functionally equivalent to well-known universal approximators.

These flexible models are successfully applied to analyse financial or macroeconomic data which may possess non-trivial statistical properties, such as fat tails, asymmetric distributions and changing variation over time. A probabilistic fuzzy system can be used to forecast inflation and to analyse stock market data, for multiple horizons in a single model.

Density forecasts of inflation show periods with multimodality, hence standard distributional assumptions, such as normality, may not hold for inflation. Inflation is a process with a varying persistence over time, indicated by the different conditional distributions for each probabilistic fuzzy system rule. Each rule is associated with a different combination of linguistic descriptions of inflation and inflation expectation levels.

Probabilistic fuzzy systems and fuzzy GARCH models allow to analyse seasonality, volatility clustering and leverage effects for stock market data, using a careful construction of the system. The probabilistic fuzzy systems considered in this work uses a simple dynamic structure and includes additional information through proxy variables on system's dynamics. The fuzzy GARCH model includes information on the system's dynamic structure through a well performing econometric model, which is combined with fuzzy linguistic descriptors.

7.3 Future Research

There are several pertinent aspects of the proposed models that were not considered in this work. Due to the particular probabilistic and fuzzy nature of the rule-base systems presented, the interpretation of the fuzzy rules (Dubois and Prade, 1996) in these models should be subject to study. An essential aspect is the definition of conditional probability of fuzzy events in probabilistic fuzzy system. Although there are definitions of conditional probabilities of fuzzy events that satisfy the classical axioms of conditional probabilities as given by de Finetti (1949) and Popper (1959), such as Baldwin et al. (1996); Coletti and Scozzafava (2006), it is necessary to consider how these definitions can be interpreted in the context of probabilistic fuzzy systems. This study will also serve to clarify possible mathematical similarities which may lead to misunderstandings regarding probabilistic fuzzy systems. The relation between PFS and stochastic models used in econometrics, such as Markov chain models, should also be subject to a detailed study. Furthermore, the relation of the models to other representations of uncertainty such as the Dempster-Shafer (Shafer, 1976; Smets and Kennes, 1994; Smets, 1998), which also have been used in rule base form in Almeida and Kaymak (2010); Almeida et al. (2012b), should be studied. Finally, methods to handle missing data, as in Almeida et al. (2010); Pereira et al. (2011), should be developed for the proposed models.

In terms of the econometric approach, this thesis leaves more detailed model comparison and model selection analyses for future research. In this thesis, model comparison is based on the selected properties of the estimated conditional distribution. An appropriate theoretical framework should be developed to address the models' performance adequately, as in Vuong (1989); Burnham and Anderson (2002). Another interesting econometric extension for the proposed models is the estimation of multivariate conditional densities where the joint conditional

distribution is non-standard with dependencies between output variables. Such a theoretical extension will be useful to analyse co-movements in a set of financial data, such as stock returns of different companies, as discussed in Bauwens et al. (2006).

In this thesis, the proposed models are applied to finance and density estimation problems. Other applications of fuzzy models to financial data have been considered in Milea et al. (2010, 2011, 2012), and these applications can be extended in future work. Some design aspects of the proposed models for applications, such as selecting appropriate input variables or the number of fuzzy rules are left for future research. Given the complex structure of financial data and several macroeconomic and financial events that have a potential effect on financial data, developing a method for feature selection from a large set of potential variables deserves special attention for future work. Such a study may improve the prediction and forecasting power of the proposed models substantially.

References

Abonyi, J. and Szeifert, F. Supervised fuzzy clustering for the identification of fuzzy classifiers. *Pattern Recognition Letters*, 24:2195–2207, 2001.

Adam, T. and Goyal, V. K. The investment opportunity set and its proxy variables. *Journal of Financial Research*, 31(1):41–63, 2008. ISSN 1475-6803. doi: 10.1111/j.1475-6803.2008.00231.x.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57(3):1047–1091, 06 2002.

Almeida, R. J. and Kaymak, U. Tail point density estimation using probabilistic fuzzy systems. In *IFSA World Congress / EUSFLAT Conference (IFSA-EUSFLAT 2009)*, pages 1809–1814, Lisbon, Portugal, July 2009a.

Almeida, R. J. and Kaymak, U. Probabilistic fuzzy systems in Value-at-Risk estimation. *Intelligent Systems in Accounting, Finance & Management*, 16(1–2):49–70, 2009b.

Almeida, R. J. and Kaymak, U. TS-models from evidential clustering. In Hüllermeier, E., Kruse, R., and Hoffmann, F., editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*, volume 80 of *Communications in Computer and Information Science*, pages 228–237. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-14055-6.

Almeida, R. J., Kaymak, U., and Sousa, J. A new approach to dealing with missing values in data-driven fuzzy modeling. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–7, july 2010. doi: 10.1109/FUZZY.2010.5584894.

Almeida, R. J., Basturk, N., Kaymak, U., and Milea, V. 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, march 2012a.

Almeida, R. J., Denoeux, T., and Kaymak, U. Constructing rule-based models using the belief functions framework. In Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., and Yager, R. R., editors, *Advances in Computational Intelligence*, volume 299 of *Communications in Computer and Information Science*, pages 554–563. Springer Berlin Heidelberg, 2012b. ISBN 978-3-642-31717-0.

Almeida, R. J., Basturk, N., Kaymak, U., and Costa Sousa, J. Estimation of flexible fuzzy GARCH models for conditional density estimation. ERIM Report Series ERS–2013–013–LIS, Erasmus Research Institute of Management ERIM, July 2013a.

Almeida, R. J., Basturk, N., Kaymak, U., and Costa Sousa, J. M. Conditional density estimation using fuzzy GARCH models. In Kruse, R., Berthold, M. R., Moewes, C., Gil, M. A., Grzegorzewski, P. l. a., and Hryniewicz, O., editors, *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, volume 190 of *Advances in Intelligent Systems and Computing*, pages 173–181. Springer Berlin Heidelberg, 2013b.

Almeida, R. J., Basturk, N., and Kaymak, U. Conditional density estimation using probabilistic fuzzy systems. Accepted for publication in 2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), 2014a.

Almeida, R. J., Basturk, N., Kaymak, U., and Costa Sousa, J. Estimation of flexible fuzzy GARCH models for conditional density estimation. *Information Sciences*, 267(0):252–266, 2014b. ISSN 0020-0255. doi: 10.1016/j.ins.2014.01.021.

Almeida, R. J., Verbeek, N., Kaymak, U., and Costa Sousa, J. Probabilistic fuzzy systems as additive fuzzy systems. Accepted for publication in Communications in Computer and Information Science (CCIS) by Springer, 2014c.

Andersen, T. G. and Bollerslev, T. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2–3):115–158, June 1997a.

Andersen, T. G. and Bollerslev, T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance*, 52 (3):975–1005, 1997b.

Andersen, T. G. and Bollerslev, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, November 1998.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. Volatility and correlation forecasting. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 15, pages 777–878. Elsevier, Amsterdam, North-Holland, 2006.

Andersen, T. G., Bollerslev, T., and Meddahi, N. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234, 2011.

Araújo, R. d. A. Swarm–based translation-invariant morphological prediction method for financial time series forecasting. *Information Sciences*, 180(24):4784–4805, 2010. ISSN 0020–0255. doi: 10.1016/j.ins.2010.08.037.

Ascari, G., Fagiolo, G., and Roventini, A. Fat-tail distributions and business-cycle models. LEM Papers Series 2012/02, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, Jan. 2012.

Awartani, B. M. and Corradi, V. Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, 21(1):167–183, 2005.

Babuška, R. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, Dordrecht, London, 1998.

Baillie, R. T. and DeGennaro, R. P. Stock returns and volatility. *Journal of Financial and Quantitative Analysis*, 25(02):203–214, June 1990.

Baldwin, J., Lawry, J., and Martin, T. A mass assignment theory of the probability of fuzzy events. *Fuzzy Sets and Systems*, 83(3):353–367, 1996. ISSN 0165–0114. doi: 10.1016/0165-0114(95)00297-9.

Bao, Y., Lee, T.-H., and Saltoğlu, B. Comparing density forecast models. *Journal of Forecasting*, 26(3):203–225, 2007. ISSN 1099-131X. doi: 10.1002/for.1023.

Barndorff-Nielsen, O. E. and Shephard, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002. ISSN 1467–9868.

Bashtannyk, D. M. and Hyndman, R. J. Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298, 2001.

Baştürk, N., Cakmaklı, C., Ceyhan, P., and Van Dijk, H. K. Posterior-predictive evidence on US inflation using Phillips Curve models with non-filtered time series. Technical Report 13–011/3, Tinbergen Institute, 2013.

Bauwens, L., Laurent, S., and Rombouts, J. V. Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

Bauwens, L., Preminger, A., and Rombouts, J. Theory and inference for a Markov switching GARCH model. *The Econometrics Journal*, 13(2):218–244, 2010.

Becker, G. S., Philipson, T. J., and Soares, R. R. The quantity and quality of life and the evolution of world inequality. *American Economic Review*, 95(1):277–291, 2005. doi: 10.1257/0002828053828563.

Bernanke, B. S. Inflation expectations and inflation forecasting. In *Remarks at the Monetary Economics Workshop of the National Bureau of Economic Research Summer Institute, Cambridge, Massachusetts, available on the Federal Reserve Board web site*, 2007.

Bertoluzza, C., Gil, M. A., and Ralescu, D. A., editors. *Statistical Modeling, Analysis and Management of Fuzzy Data*. Studies in Fuzziness and Soft Computing. Physica Verlag, Heidelberg, 2002.

Berument, H. and Kiymaz, H. The day of the week effect on stock market volatility. *Journal of Economics and Finance*, 25(2):181–193, 2001.

Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

Billio, M. and Pelizzon, L. Value-at-risk: a multivariate regime switching approach. *Journal of Empirical Finance*, 7:531–554, 2000.

Bishop, C. M. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer–Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

Bollerslev, T. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547, 1987.

Bollerslev, T., Engle, R. F., and Nelson, D. B. ARCH models. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 49, pages 2959–3038. Elsevier, 1 edition, 1986.

Bowden, R. J. Option value at risk and the value of the firm: does it pay to hedge? Working paper, Victoria University of Wellington, Wellington, New Zealand, May 2006.

Brealey, R. A. and Myers, S. C. *Principles of Corporate Finance*. McGraw Hill, New York, 7th edition edition, 2001.

Broomhead, D. and Lowe, D. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

Bugmann, G. Normalized Gaussian radial basis function networks. *Neurocomputing*, 20(1–3): 97–110, 1998. ISSN 0925–2312. doi: 10.1016/S0925-2312(98)00027-7.

Burnham, K. P. and Anderson, D. R. *Model selection and multi-model inference: a practical information-theoretic approach*. Springer, 2002.

Chang, B. R. Applying nonlinear generalized autoregressive conditional heteroscedasticity to compensate anfis outputs tuned by adaptive support vector regression. *Fuzzy Sets and Systems*, 157(13):1832–1850, 2006. ISSN 0165–0114.

Chang, B. R. Resolving the forecasting problems of overshoot and volatility clustering using anfis coupling nonlinear heteroscedasticity with quantum tuning. *Fuzzy Sets and Systems*, 159(23):3183–3200, 2008. ISSN 0165–0114.

Chapados, N. and Bengio, Y. Cost functions and model combination for VaR-based asset allocation using neural networks. *Neural Networks, IEEE Transactions on*, 12(4):890–906, July 2001.

Cheng, C.-H., Chen, T.-L., and Wei, L.-Y. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*, 180(9):1610–1629, 2010. ISSN 0020–0255. doi: 10.1016/j.ins.2010.01.014.

Cherubini, U. and Della Lunga, G. Fuzzy value-at-risk: accounting for market liquidity. *Economic Notes by Banca Monte dei Paschi di Siena SpA*, 30(2):293–312, 2001.

Christoffersen, P. F. Evaluating interval forecasts. *International Economic Review*, 39(4):841–62, November 1998.

CoIL2000. Soundbytes from CoIL 2000. *Synergy*, 3:10–11, autumn 2000. Newsletter of CoIL, the Computational Intelligence and Learning Cluster.

Coletti, G. and Scozzafava, R. Conditional probability and fuzzy information. *Computational Statistics & Data Analysis*, 51(1):115 – 132, 2006. ISSN 0167-9473. doi: 10.1016/j.csda.2006.04.028. The Fuzzy Approach to Statistical Analysis.

Colubi, A., Fernandez-Garcia, C., and Gil, M. Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. *Fuzzy Systems, IEEE Transactions on*, 10(3):384–390, june 2002. ISSN 1063-6706. doi: 10.1109/TFUZZ.2002.1006441.

Cont, R. Volatility clustering in financial markets: empirical facts and agent-based models. In Teyssiére, G. and Kirman, A. P., editors, *Long Memory in Economics*, pages 289–310. Springer, Berlin, 2007.

Coppi, R., Gil, M. A., and Kiers, H. A. The fuzzy approach to statistical analysis. *Computational Statistics & Data Analysis*, 51(1):1–14, 2006. ISSN 0167-9473. doi: 10.1016/j.csda.2006.05.012.

Cristianini, N. and Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 1 edition, 2000. ISBN 0521780195.

Cúrdia, V., Del Negro, M., and Greenwald, D. L. Rare shocks, great recessions. Working paper series, Federal Reserve Bank of San Francisco, 2013.

da Costa Sousa, J. M. and Kaymak, U. *Fuzzy Decision Making in Modeling and Control*, volume 27 of *World Scientific Series in Robotics and Intelligent Systems*. World Scientific, New Jersey, 2002.

Das, R., Ang, K. K., and Quek, C. A synergy of econometrics and computational methods (GARCH-RNFS) for volatility forecasting. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–8, july 2010.

de Finetti, B. Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Univ. Trieste*, 19:3–55, 1949. (English translation: Probability, Induction, Statistics. Wiley, London, 1972 (Chapter 5).

de Finetti, B. *Probability, Induction and Statistics: The Art of Guessing*. Wiley series in probability and mathematical statistics. John Wiley and Son, London, 1972.

de Finetti, B. *Theory of Probability*, volume 1. John Wiley and Sons, New York, 1974.

Del Negro, M. and Schorfheide, F. DSGE model-based forecasting. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 57–140. Elsevier, 2013.

Diebold, F. X., Gunther, T. A., and Tay, A. S. Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–83, November 1998.

Donaldson, R. and Kamstra, M. An artificial neural network-GARCH model for international stock return volatility. *Journal of Empirical Finance*, 4(1):17–46, 1997. ISSN 0927–5398.

Dubois, D. and Prade, H. Fuzzy sets and probability: misunderstandings, bridges and gaps. In *Fuzzy Systems, 1993., Second IEEE International Conference on*, volume 2, pages 1059–1068, 1993. doi: 10.1109/FUZZY.1993.327367.

Dubois, D. and Prade, H. What are fuzzy rules and how to use them. *Fuzzy Sets and Systems*, 84(2):169–185, 1996. ISSN 0165-0114. doi: 10.1016/0165-0114(96)00066-8.

Dubois, D. and Prade, H. The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90(2):141 – 150, 1997. ISSN 0165-0114. doi: 10.1016/S0165-0114(97)00080-8. Fuzzy Sets: Where Do We Stand? Where Do We Go?

Dubois, D. and Prade, H. Quantitative possibility theory and its probabilistic connections. In Grzegorzewski, P., Hryniewicz, O., and Gil, M. A., editors, *Soft Methods in Probability, Statistics and Data Analysis*, Advances in Soft Computing, pages 3–26. Physica Verlag, Heidelberg, 2002.

Dubois, D. and Prade, H. Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems*, 192(0):3 – 24, 2012. ISSN 0165-0114. doi: 10.1016/j.fss.2010.11.007. Fuzzy Set Theory Where Do We Stand and Where Do We Go?

Dubois, D., Esteva, F., Godo, L., and Prade, H. An information-based discussion of vagueness: six scenarios leading to vagueness. In Cohen, H. and Lefebvre, C., editors, *Handbook of Categorization in Cognitive Science*, chapter 40, pages 891–909. Elsevier, Amsterdam, 2005.

Duffie, D. and Pan, J. An overview of value at risk. *The Journal of Derivatives*, 4(3):7–49, Spring 1997.

Engle, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

Engle, R. F. and Ng, V. K. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1748–1778, December 1993.

Fan, J. A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, 20:317–337, 2005.

Fan, J., Yao, Q., and Tong, H. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189–206, 1996.

Faust, J. and Wright, J. Forecasting inflation. *manuscript, Johns Hopkins University*, 2012.

Figueiredo, M. A. T. On Gaussian radial basis function approximations: interpretation, extensions, and learning strategies. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 618–621 vol.2, 2000. doi: 10.1109/ICPR.2000.906151.

Friedman, J. H. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1–67, 1991. ISSN 00905364. doi: 10.2307/2241837.

Galí, J. and Gertler, M. Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2):195–222, 1999.

Geng, L. and Ma, J. 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, oct. 2008.

Geweke, J. and Keane, M. Smoothly mixing regressions. *Journal of Econometrics*, 138(1): 252–290, May 2007.

Glosten, L., Jagannathan, R., and Runkle, D. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801, 1993.

González-Rivera, G. Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):61–78, 1998.

Goodman, I. R. and Nguyen, H. T. Fuzziness and randomness. In Bertoluzza, C., Gil, M. A., and Ralescu, D. A., editors, *Statistical Modeling, Analysis and Management of Fuzzy Data*, Studies in Fuzziness and Soft Computing, pages 3–21. Physica Verlag, Heidelberg, 2002.

Grzegorzewski, P., Hryniewicz, O., and Gil, M. A., editors. *Soft Methods in Probability, Statistics and Data Analysis*. Advances in Soft Computing. Physica Verlag, Heidelberg, 2002.

Gustafson, D. and Kessel, W. Fuzzy clustering with a fuzzy covariance matrix. In *Proceedings of IEEE Conference on Decision and Control, CDC*, pages 761–766, San Diego, CA, USA, 1979.

Haas, M., Mittnik, S., and Paolella, M. S. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530, 2004.

Hachicha, N., Jarboui, B., and Siarry, P. A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics. *Information Sciences*, 181(1): 79–91, 2011. ISSN 0020–0255.

Hacking, I. *The Emergence of Probability: a Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference*. Cambridge University Press, London, 1975. ISBN 0-521-20460-7.

Hansen, P. R. and Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7):873–889, 2005. ISSN 1099–1255.

Harvey, C. R. and Siddique, A. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55(3):1263–1295, 2000.

Heimes, F. and van Heuveln, B. The normalized radial basis function neural network. In *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, volume 2, pages 1609–1614, 1998. doi: 10.1109/ICSMC.1998.728118.

Helin, T. and Koivisto, H. The GARCH-fuzzy density method for density forecasting. *Applied Soft Computing*, 11(6):4212–4225, 2011a. ISSN 1568-4946. doi: DOI:10.1016/j.asoc.2011.03.015.

Helin, T. and Koivisto, H. The GARCH-fuzzy density method for density forecasting. *Applied Soft Computing*, 11(6):4212–4225, Sept. 2011b. ISSN 1568–4946.

Hengjie, S., Miao, C., Shen, Z., Roel, W., D'Hondt, M., and Francky, C. A probabilistic fuzzy approach to modeling nonlinear systems. *Neurocomputing*, 74:1008–1025, February 2011. ISSN 0925-2312. doi: 10.1016/j.neucom.2010.11.014.

Herrera, F., Lozano, M., and Verdegay, J. L. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial intelligence review*, 12(4):265–319, 1998.

Herzig, A. Action representation and partially observable planning using epistemic logic. In *International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1067–1072. Morgan Kaufmann, San Francisco, 2003.

Hesse, M. Bayesian methods and the initial probabilities of theories. In Maxwell, G. and Anderson, R., editors, *Induction, Probability, and Confirmation*, Minnesota Studies in the Philosophy of Sciences. University of Minnesota Press, Minneapolis, 1975.

Hinojosa, W., Nefti, S., and Kaymak, U. Systems control with generalized probabilistic fuzzy-reinforcement learning. *Fuzzy Systems, IEEE Transactions on*, 19(1):51–64, feb. 2011. ISSN 1063-6706. doi: 10.1109/TFUZZ.2010.2081994.

Holton, G. A. *Value at Risk: thoery and practice*. Academic Press, Amsterdam, The Netherlands, 2003.

Hong, S., Lee, H., and Kim, E. A new probabilistic fuzzy model: Fuzzification–maximization (FM) approach. *International Journal of Approximate Reasoning*, 50:1129–1147, July 2009. ISSN 0888-613X. doi: 10.1016/j.ijar.2009.05.004.

Huang, X. A new perspective for optimal portfolio selection with random fuzzy returns. *Information Sciences*, 177(23):5404–5414, 2007. ISSN 0020–0255.

Hull, J. *Options, Futures and Other Derivatives*. Prentice Hall, Upper Saddle River, 4 edition, 2000.

Hung, J.-C. Forecasting volatility of stock market using adaptive fuzzy-GARCH model. In *Computer Sciences and Convergence Information Technology, 2009. ICCIT '09. Fourth International Conference on*, pages 583–587, nov. 2009a.

Hung, J.-C. A fuzzy asymmetric GARCH model applied to stock markets. *Information Sciences*, 179(22):3930–3943, 2009b.

Hung, J.-C. A fuzzy GARCH model applied to stock market scenario using a genetic algorithm. *Expert Systems with Applications*, 36(9):11710–11717, 2009c.

Hung, J.-C. Applying a combined fuzzy systems and GARCH model to adaptively forecast stock market volatility. *Applied Soft Computing*, 11(5):3938–3945, 2011a. ISSN 1568–4946.

Hung, J.-C. Adaptive fuzzy-GARCH model applied to forecasting the volatility of stock markets using particle swarm optimization. *Information Sciences*, 181(20):4673–4683, 2011b.

Hung, J.-C. and Hsu, F.-Y. Adaptive fuzzy asymmetric GARCH model applied to stock market. In *Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on*, volume 1, pages 991–996, nov. 2008.

Hunt, K., Haas, R., and Murray-Smith, R. Extending the functional equivalence of radial basis function networks and fuzzy inference systems. *Neural Networks, IEEE Transactions on*, 7 (3):776–781, 1996. ISSN 1045-9227. doi: 10.1109/72.501735.

James Chu, C.-S., Santoni, G. J., and Liu, T. Stock market volatility and regime shifts in returns. *Information Sciences*, 94(1–4):179–190, 1996. ISSN 0020–0255.

Jang, J.-S. R. and Sun, C.-T. *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1 edition, 1997. ISBN 0-13-261066-3.

Jiang, M., Meng, Z., and Hu, Q. A neural network model on solving multiobjective conditional value-at-risk. In Yin, F., Wang, J., and Guo, C., editors, *Advances in Neural Networks*, number 3174 in Lecture Notes in Computer Science, pages 1000–1006. Springer, Berlin, 2004.

John, G. H. and Langley, P. Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.

Jorion, P. *Value at Risk: the new benchmark for managing financial risk*. McGraw–Hill, New York, 3 edition, 2006.

Karnik, N., Mendel, J., and Liang, Q. Type-2 fuzzy logic systems. *Fuzzy Systems, IEEE Transactions on*, 7(6):643–658, dec 1999. ISSN 1063-6706. doi: 10.1109/91.811231.

Kaymak, U., van den Bergh, W.-M., and van den Berg, J. A fuzzy additive reasoning scheme for probabilistic Mamdani fuzzy systems. In *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems*, volume 1, pages 331–336, St. Louis, USA, May 2003.

Kecman, V. *Learning and Soft Computing*. MIT Press, Cambridge, MA, 2001.

Khrennikov, A. *Interpretations of Probability: Mathematics*. VSP, 1999.

Klaassen, F. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2):363–394, 2002.

Klir, G. On the alleged superiority of probabilistic representation of uncertainty. *Fuzzy Systems, IEEE Transactions on*, 2(1):27–31, 1994. ISSN 1063-6706. doi: 10.1109/91.273119.

Klir, G. J. and Yuan, B. *Fuzzy Sets and Fuzzy Logic: theory and applications*. Prentice Hall, Upper Saddle River, 1995.

Koenker, R. *Quantile regression*. Number 38. Cambridge university press, 2005.

Koenker, R. and Hallock, K. F. Quantile regression. *Journal of Economic Perspectives*, 15(4): 143–156, Fall 2001.

Kosko, B. Fuzzy systems as universal approximators. *IEEE Transactions on Computers*, 43: 1329–1333, 1994.

Kosko, B. *Fuzzy Engineering*. Prentice Hall, Upper Saddle River, New Jersey, 1997.

Kosko, B. Global stability of generalized additive fuzzy systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 28(3):441–452, 1998. ISSN 1094–6977. doi: 10.1109/5326.704584.

Kreinovich, V. and Nguyen, H. Towards intuitive understanding of the cauchy deviate method for processing interval and fuzzy uncertainty. In *IFSA World Congress / EUSFLAT Conference (IFSA-EUSFLAT 2009)*, pages 2336–2341, July 2009. doi: 10.1109/ICSMC.2010.5641997.

Kreinovich, V., Nguyen, H., and Sriboonchitta, S. Symmetries: A general approach to integrated uncertainty management. In Huynh, V.-N., Nakamori, Y., Lawry, J., and Inuiguchi, M., editors, *Integrated Uncertainty Management and Applications*, volume 68 of *Advances in Intelligent and Soft Computing*, pages 141–152. Springer Berlin / Heidelberg, 2010. ISBN 978-3-642-11959-0.

Kruse, R. The strong law of large numbers for fuzzy random variables. *Information Sciences*, 28:233–241, 1982.

Kullback, S. and Leibler, R. A. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Kupiec, P. H. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 2:73–84, 1995.

Lee, H.-E., Park, K.-H., and Bien, Z. Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data. *Fuzzy Systems, IEEE Transactions on*, 16(1):263–277, feb. 2008. ISSN 1063-6706. doi: 10.1109/TFUZZ.2007.903314.

Li, F., Villani, M., and Kohn, R. Flexible modeling of conditional distributions using smooth mixtures of asymmetric student t densities. *Journal of Statistical Planning and Inference*, 140(12):3638–3654, 2010. ISSN 0378–3758.

Li, H.-X. and Liu, Z. A probabilistic neural-fuzzy learning system for stochastic modeling. *IEEE Transactions on Fuzzy Systems*, 16(4):898–908, 2008. ISSN 1063-6706. doi: 10.1109/TFUZZ.2008.917302.

Liu, Z. and Li, H.-X. A probabilistic fuzzy logic system for modeling and control. *IEEE Transactions on Fuzzy Systems*, 13(6):848–859, 2005. doi: 10.1109/TFUZZ.2005.859326.

Liu, Z. and Li, H.-X. Probabilistic fuzzy logic system: a tool to process stochastic and imprecise information. In *Proceedings of the 18th international conference on Fuzzy Systems*, FUZZ-IEEE'09, pages 848–853, 2009. ISBN 978-1-4244-3596-8.

Mamdani, E. H. and Gaines, B. R., editors. *Fuzzy Reasoning and its Applications*. Academic Press, London, 1981.

Mandelbrot, B. The variation of certain speculative prices. *The Journal of Business*, 36:394–419, 1963.

McConnell, M. M. and Perez-Quiros, G. Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review*, 90(5):1464–1476, December 2000.

Meghdadi, A. H. and Akbarzadeh-T., M.-R. Probabilistic fuzzy logic and probabilistic fuzzy systems. In *Proceedings of the Tenth IEEE International Conference on Fuzzy Systems*, pages 1127–1130, Melbourne, Australia, Dec. 2001.

Michalewicz, Z. *Genetic algorithms+ data structures= evolution programs*. Springer, 1996.

Milea, V., Almeida, R. J., Kaymak, U., and Frasincar, F. A fuzzy model of the MSCI EURO index based on content analysis of european central bank statements. In *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*, pages 1–7, july 2010. doi: 10.1109/FUZZY.2010.5584815.

Milea, V., Almeida, R. J., Kaymak, U., and Frasincar, F. A fuzzy model of a European index based on automatically extracted content information. In *Computational Intelligence for Financial Engineering and Economics (CIFEr), 2011 IEEE Symposium on*, pages 1–8, april 2011. doi: 10.1109/CIFER.2011.5953571.

Milea, V., Almeida, R. J., Sharef, N. M., Kaymak, U., and Frasincar, F. Computational content analysis of european central bank statements. *International Journal of Computer Information Systems and Industrial Management Applications*, 4:628–640, 2012.

Montgomery, M. R., Gragnolati, M., Burke, K. A., and Paredes, E. Measuring living standards with proxy variables. *Demography*, 37(2):155–174, 2000. ISSN 0070-3370. doi: 10.2307/2648118.

Moody, J. and Darken, C. J. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1(2):281–294, June 1989. ISSN 0899-7667. doi: 10.1162/neco.1989.1.2.281.

Nau, R. F. De finetti was right: Probability does not exist. *Theory and Decision*, 51(2–4): 89–124, 2001. ISSN 0040-5833.

Pedrycz, W. Applications of fuzzy relational equations for methods of reasoning in presence of fuzzy data. *Fuzzy Sets and Systems*, 16(2):163–175, July 1985. ISSN 0165-0114. doi: 10.1016/S0165-0114(85)80016-6.

Pereira, R., Almeida, R. J., Kaymak, U., Vieira, S., Sousa, J., Reti, S., Howell, M., and Finkelstein, S. Predicting septic shock outcomes in a database with missing data using fuzzy modeling: Influence of pre-processing techniques on real-world data-based classification. In *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*, pages 2507–2512, june 2011. doi: 10.1109/FUZZY.2011.6007606.

Popov, A. A. and Bykhanov, K. V. Modeling volatility of time series using fuzzy GARCH models. In *9th Russian–Korean International Symposium on Science and Technology, KORUS2005*, pages 687–692, 2005.

Popper, K. *The Logic of Scientific Discovery*. Routledge, London, 1959.

Rigobon, R. and Sack, B. Measuring the reaction of monetary policy to the stock market. *The Quarterly Journal of Economics*, 118(2):639–669, May 2003.

Rudra, A. Use of probability in economics. *Economic and Political Weekly*, 1(8):317–322, 1966. ISSN 00129976.

Ruspini, E. H. A new approach to clustering. *Information and Control*, 15(1):22–32, 1969.

Shafer, G. *A mathematical theory of evidence*. Princeton university press, 1976.

Singpurwalla, N. D. and Booker, J. M. Membership functions and probability measures of fuzzy sets. *Journal of the American Statistical Association*, 99:867–877, 2004.

Smets, P. The transferable belief model for quantified belief representation. In Gabbay, D. and Smets, P., editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, chapter 1. Kluwer, Doordrecht, 1998.

Smets, P. and Kennes, R. The transferable belief model. *Artif. Intell.*, 66(2):191–234, 1994.

Sorensen, R. Vagueness. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2013 edition, 2013.

Stock, J. H. and Watson, M. W. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33, 02 2007.

Stock, J. H. and Watson, M. W. Modeling inflation after the crisis. NBER Working Papers 16488, National Bureau of Economic Research, Inc, Oct. 2010.

Takagi, T. and Sugeno, M. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man and Cybernetics*, 15(1):116–132, Jan./Feb. 1985.

Tang, M., Chen, X., Hu, W., and Yu, W. Generation of a probabilistic fuzzy rule base by learning from examples. *Information Sciences*, 217(0):21–30, 2012. ISSN 0020-0255. doi: 10.1016/j.ins.2012.06.021.

Taylor, J. W. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. *Journal of Forecasting*, 19(4):299–311, July 2000.

Thavaneswaran, A., Thiagarajah, K., and Appadoo, S. Fuzzy coefficient volatility (FCV) models with applications. *Mathematical and Computer Modelling*, 45(7–8):777–786, 2007.

Thomas, S. *Fuzziness and Probability*. ACG Press, Wichita KS, USA, 1995.

Thomson Reuters/University of Michigan. University of Michigan inflation expectation (MICH), quarterly data, not seasonally adjusted. Retrieved 2013-03-15.

van den Berg, J., van den Bergh, W.-M., and Kaymak, U. Probabilistic and statistical fuzzy set foundations of competitive exception learning. In *Proceedings of the Tenth IEEE International Conference on Fuzzy Systems*, volume 2, pages 1035–1038, Melbourne, Australia, Dec. 2001.

van den Berg, J., Kaymak, U., and van den Bergh, W.-M. Fuzzy classification using probability-based rule weighting. In *Proceedings of 2002 IEEE International Conference on Fuzzy Systems*, pages 991–996, Honolulu, Hawaii, May 2002a.

van den Berg, J., Kaymak, U., and van den Bergh, W.-M. Probabilistic reasoning in fuzzy rule-based systems. In Grzegorzewski, P., Hryniewicz, O., and Gil, M. A., editors, *Soft Methods*

in Probability, Statistics and Data Analysis, Advances in Soft Computing, pages 189–196. Physica Verlag, Heidelberg, 2002b.

van den Berg, J., Kaymak, U., and van den Bergh, W.-M. Financial markets analysis by using a probabilistic fuzzy modelling approach. *International Journal of Approximate Reasoning*, 35:291–305, 2004.

van den Berg, J., Kaymak, U., and Almeida, R. J. Function approximation using probabilistic fuzzy systems. ERIM Report Series ERS-2011-026-LIS, Erasmus Research Institute of Management ERIM, Jan. 2012.

van den Berg, J., Kaymak, U., and Almeida, R. J. Conditional density estimation using probabilistic fuzzy systems. *Fuzzy Systems, IEEE Transactions on*, 21(5):869–882, 2013. ISSN 1063-6706. doi: 10.1109/TFUZZ.2012.2235839.

van den Eijkel, G. C. *Fuzzy Probabilistic Learning and Reasoning*. Ph.D. thesis, Delft University of Technology, Delft University Press, Mekelweg 4, Delft, The Netherlands, Jan. 1999.

Villani, M., Kohn, R., and Giordani, P. Regression density estimation using smooth adaptive Gaussian mixtures. *Journal of Econometrics*, 153(2):155–173, 2009. ISSN 0304–4076.

von Neumann, J. and Morgenstern, O. *Games and Economic Behavior*. Wiley, New York, 1944.

Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, pages 307–333, 1989.

Walley, P. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London, 1991.

Waltman, L., Kaymak, U., and van den Berg, J. Fuzzy histograms: A statistical analysis. In *EUSFLAT-LFA 2005 Joint 4th EUSFLAT 11th LFA Conference*, pages 605–610, 2005a.

Waltman, L., Kaymak, U., and van den Berg, J. Maximum likelihood parameter estimation in probabilistic fuzzy classifiers. In *Fuzzy Systems, 2005. FUZZ '05. The 14th IEEE International Conference on*, pages 1098–1103, May 2005b. doi: 10.1109/FUZZY.2005.1452548.

Werbos, P. J. Intelligence in the brain: A theory of how it works and how to build it. *Neural Networks*, 22(3):200–212, 2009. ISSN 0893-6080. doi: 10.1016/j.neunet.2009.03.012.

Wiener, Z. Introduction to VaR (value-at-risk). In Galai, D., Ruthenberg, D., Sarnat, M., and Schreiber, B., editors, *Risk Management and Regulation in Banking*, pages 47–63. Kluwer Academic Publishers, Boston, 1999.

Wright, A. H. Genetic algorithms for real parameter optimization. In *Foundations of Genetic Algorithms (FOGA)*, pages 205–218. Morgan Kaufmann, 1991.

Xu, D. and Kaymak, U. 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.

Yatchew, A. Nonparametric regression techniques in economics. *Journal of Economic Literature*, 36(2):669–721, 1998.

Zadeh, L. The concept of a linguistic variable and its application to approximate reasoningi. *Information Sciences*, 8(3):199–249, 1975. ISSN 0020–0255.

Zadeh, L. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:1–28, 1978.

Zadeh, L. A. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

Zadeh, L. A. Probability measures of fuzzy events. *J. Math. Anal. Appl.*, 23:421–427, 1968.

Zadeh, L. A. Fuzzy sets and information granularity. *Advances in Fuzzy Set Theory and Applications*, pages 69–129, 1979.

Zadeh, L. A. Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 37(3):271–276, 1995. ISSN 00401706.

Zadeh, L. A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*, 90(2):111–127, 1997. ISSN 0165–0114.

Zadeh, L. A. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. In Grzegorzewski, P., Hryniewicz, O., and Gil, M. A., editors, *Soft Methods in Probability, Statistics and Data Analysis*, Advances in Soft Computing, pages 27–61. Physica Verlag, Heidelberg, 2002.

Zadeh, L. A. Is there a need for fuzzy logic? *Information Sciences*, 178(13):2751–2779, 2008. ISSN 0020–0255.

Zhang, G. and Li, H.-X. A probabilistic fuzzy learning system for pattern classification. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 2336–2341, oct. 2010. doi: 10.1109/ICSMC.2010.5641997.

Zhang, G. and Li, H.-X. An efficient configuration for probabilistic fuzzy logic system. *Fuzzy Systems, IEEE Transactions on*, 20(5):898–909, oct. 2012. ISSN 1063-6706. doi: 10.1109/TFUZZ.2012.2188897.

Zheng, X. and Chen, B. M. Identification of stock market forces in the system adaptation framework. *Information Sciences*, 2013. ISSN 0020–0255. doi: 10.1016/j.ins.2013.12.028. Available online 31 December 2013.

Zivot, E. Practical issues in the analysis of univariate GARCH models. In Andersen, T., Davis, R., Krei, J.-P., and Mikosch, T., editors, *Handbook of Financial Time Series*, pages 113–155. Springer Verlag, New York, 2009.

Zmeškal, Z. Application of the fuzzy-stochastic methodology to appraising the firm value as a european call option. *European Journal of Operational Research*, 135(2):303–310, 2001.

Zmeškal, Z. Value at risk methodology under soft conditions approach (fuzzy-stochastic approach). *European Journal of Operational Research*, 161:337–347, 2005a.

Zmeškal, Z. Value at risk methodology of international index portfolio under soft conditions (fuzzy-stochastic approach). *International Review of Financial Analysis*, 14:263–275, 2005b.

Summary in English

Conditional density estimation is an important problem in many areas such as system identification and machine learning, where the predicted density is typically highly non-linear and multimodal, or in empirical economics, macroeconomic analysis, quantitative finance and risk management, where financial time-series typically possess non-trivial statistical properties, such as fat tails, asymmetric distributions and changing variation over time. This thesis considers the general problem of conditional density estimation, *i.e.* estimating and predicting the density of the response variable as a function of covariates.

The semi-parametric models studied and developed in this thesis, namely the fuzzy GARCH model and probabilistic fuzzy systems, combine fuzzy and probabilistic representations of uncertainty, while making very few assumptions regarding the functional form of the estimated density or regarding changes across the space of covariates. These models possess sufficient generalization power to approximate a non-standard density along with the ability to describe the underlying process and to incorporate possible non-linear relations between variables using simple linguistic descriptors. Such systems can capture different properties of data, such as fat tails, skewness and multimodality in one single model.

The proposed models are applied to time series data for macroeconomic analysis, quantitative finance and risk management. The analysed time series data exhibit complex behavior and non-trivial statistical properties. A probabilistic fuzzy system is applied to multi-horizon forecasting of quarterly U.S. inflation. Point and density forecasts of inflation are of great importance for macroeconomic policy makers and financial institutions. Based on these one quarter and one year ahead forecasts, a central bank can adjust the monetary policy instruments accurately and financial institutions can quantify inflationary risk. This application shows that slowly changing patterns in inflation are accurately captured by probabilistic fuzzy systems.

Fuzzy GARCH models and probabilistic fuzzy systems are used for analysing the returns distribution of different stocks, specially of the S&P 500 index. The main focus in analysing 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 proposed models perform

well in estimating the expected gains and losses from investing in different stocks. The reason for the proposed models to successfully capture such interesting values is two fold. First, the flexible functional forms in these models allow to approximate a nonstandard returns density. Second, possibly complex effects of current market information on future returns are explained using simple linguistic descriptors coupled with stochastic models. Particular relevance is given to the interpretation of these models in the study of stylized facts on the distribution of stock returns, such as seasonality and volatility clustering.

Nederlandse Samenvatting (Summary in Dutch)

Conditionele dichtheid schatten is een belangrijk probleem in veel gebieden, waaronder systeem identificatie en machinaal leren, waar de voorspelde dichtheid vaak extreem non-lineair en multimodaal is. Andere gebieden zijn de empirische economie, macro-economische analyse en kwantitatief financieel risicomanagement, waar financile tijdreeksen typische niet-triviale statistische eigenschappen vertonen zoals fat tails, asymmetrische verdelingen en veranderende variantie in tijd. Dit proefschrift kijkt naar het algemene probleem van conditionele dichtheid schatting, namelijk het schatten en voorspellen van de respons variabele als een functie van co varianten.

De semi-parametrische modellen die wij bestuderen, namelijk het fuzzy GARCH model en probabilistische vage systemen, zijn een samensmelting van vage en probabilistische representaties van onzekerheid, met zeer weinig aannames met betrekking tot de functionele vorm van de geschatte dichtheid en veranderingen in de ruimte van co varianten. Deze modellen beschikken over voldoende mogelijkheid tot generalisatie voor het schatten van atypische dichtheid en kunnen ook het onderliggende proces beschrijven rekening houdend met de mogelijk niet-lineaire verbanden tussen variabelen door gebruik te maken van eenvoudige lingustische beschrijvingen. Zulke systemen kunnen verschillende eigenschappen van de data vangen, zoals fat tails, scheefheid en multimodaliteit in een enkel model.

De voorgestelde modellen worden toegepast op tijdreeksen voor macro-economische analyse, kwantitatief financieel risicomanagement. De geanalyseerde tijdreeksen vertonen complex gedrag en niet-triviale statistische eigenschappen. Een probabilistisch vaag model is toegepast voor het voorspellen van kwartaal inflatie in de VS over meerdere perioden. Punts- en dichtheid inflatie voorspellingen zijn van groot belang voor macro-economische beleidsmakers en financile instellingen. Gebaseerd op kwartaal- en jaarvoorspellingen kan een centrale bank haar monetair beleid aanpassen en kunnen financile instellingen inflatie risico kwantificeren. Deze toepassing toont dat langzaam veranderende patronen in inflatie kunnen worden gevangen in probabilistische vage systemen.

Fuzzy GARCH modellen en probabilistische vage systemen worden gebruikt voor het analyseren van verdelingen van returns bij verschillende aandelen en in het bijzonder de S&P 500 index. De nadruk in het analyseren van deze verdelingen ligt niet op het beoordelen van een enkel model en bijbehorende parameters voor het opstellen van beleid, maar vooral op het schatten van verwachten winsten en verliezen die kunnen optreden bij het investeren in de verschillende aandelen. De voorgestelde modellen kunnen de verwachte winsten en verliezen goed schatten. Twee redenen liggen hieraan ten grondslag. Ten eerste, de flexibele functionele formulieren in deze modellen maken het mogelijk om niet-standaard dichtheid van returns te schatten. Ten tweede, mogelijk complexe effecten van de huidige marktinformatie met betrekking tot toekomstige returns zijn verklaard door gebruik te maken van lingustische beschrijvingen in combinatie met stochastische modellen. We achten de interpretatie van deze modellen bijzonder relevant voor het bestuderen van gestileerde feiten in de verdeling van aandelen returns, zoals seizoen gebondenheid en het clusteren van volatiliteit.

Curriculum Vitae

Rui received his 5 year degree in Mechanical Engineering in 2005 from Technical University of Lisbon, Instituto Superior Técnico, Portugal. During his studies he was a research assistant in characterization of liquid films by interferometry and atomic force microscopy. In 2006 he obtained his Master of Science degree also in Mechanical Engineering from the same university. During his masters, he was a research assistant for a one year period in a project of image based classification. Directly after his masters, he started his PhD in Econometrics at the Erasmus Research Institute of Management, Erasmus University of Rotterdam. In this thesis, he worked on developing new models that combine probabilistic and fuzzy uncertainty applied to risk forecasting. During his PhD, Rui received a grant from COST Action IC072 in 2011 to perform a short term research visit at Université de Technologie de Compiègne, France, where he worked on rule-based models using the belief functions framework, and an IEEE-CIS Walter Karplus Student Research Grant in 2012 for a short term research visit at Laboratoire d'Informatique de Paris 6, Université Pierre et Marie Curie, France, where he worked on linguistic summaries of patient data to detect medical events. In 2013 he did a research visit at LISTIC, Université de Savoie, France, to work in information fusion based on expert knowledge using the belief functions framework. Rui's research has been published in several journals such as IEEE Transactions on Fuzzy Systems and Information Sciences, edited book chapters such as Advances in Intelligent Systems and Computing, Communications in Computer and Information Science or Lecture Notes in Computer Science and also international peer reviewed conference proceedings. While finalizing his thesis, Rui started to work as an assistant professor at the Eindhoven University of Technology, Information Systems Group, School of Industrial Engineering.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES RESEARCH IN MANAGEMENT

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://hdl.handle.net/1765/1

ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

DISSERTATIONS LAST FIVE YEARS

Acciaro, M., *Bundling Strategies in Global Supply Chains*, Promoter(s): Prof.dr. H.E. Haralambides, EPS-2010-197-LIS, http://hdl.handle.net/1765/19742

Akpinar, E., *Consumer Information Sharing; Understanding Psychological Drivers of Social Transmission*, Promoter(s): Prof.dr.ir. A. Smidts, EPS-2013-297-MKT, http://hdl.handle.net/1765/50140

Alexiev, A., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital*, Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-208-STR, http://hdl.handle.net/1765/20632

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: Strategy, Structure, and Performance*, Promoter(s): Prof.dr. J.Y.F. Wynstra, EPS-2014-300-LIS, http://hdl.handle.net/1765/1

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-Frequency Data*, Promoter(s): Prof.dr.D.J.C. van Dijk, EPS-2013-273-F&A, http://hdl.handle.net/1765/38240

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promoter(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://hdl.handle.net/1765/23670

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promoter(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://hdl.handle.net/1765/ 39128

Berg, W.E. van den, *Understanding Salesforce Behavior Using Genetic Association Studies,* Promoter: Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://hdl.handle.net/1765/ 39128

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promoter(s): Prof.dr. B. Krug, EPS-2012-262-ORG, http://hdl.handle.net/1765/32345

Binken, J.L.G., *System Markets: Indirect Network Effects in Action, or Inaction*, Promoter(s): Prof.dr. S. Stremersch, EPS-2010-213-MKT, http://hdl.handle.net/1765/21186

Blitz, D.C., *Benchmarking Benchmarks*, Promoter(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://hdl.handle.net/1765/226244

Boons, M.,*Working Together Alone in the Online Crowd: The Effects of Social Motivations and Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promoter: Prof.dr. H.G. Barkema, EPS-2014-306-S&E, http://hdl.net/1765/1

Borst, W.A.M., *Understanding Crowdsourcing: Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities*, Promoter(s): Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-221-LIS, http://hdl.handle.net/1765/ 21914

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds*, Promoter(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-185-F&A, http://hdl.handle.net/1765/18126

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promoter(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://hdl.handle.net/1765/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit, Coworker Satisfaction, and Relational Models*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://hdl.handle.net/1765/41508

Camacho, N.M., *Health and Marketing; Essays on Physician and Patient Decision-making*, Promoter(s): Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://hdl.handle.net/1765/23604

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promoter(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://hdl.handle.net/1765/50005

Carvalho, L., *Knowledge Locations in Cities; Emergence and Development Dynamics*, Promoter(s): Prof.dr. L. van den Berg, EPS-2013-274-S&E, http://hdl.handle.net/1765/ 38449

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*, Promoter(s): Prof.dr. P.M.A.R. Heugens & Prof.dr. J. van Oosterhout, EPS-2010-205-ORG, http://hdl.handle.net/1765/19882

Cox, R.H.G.M., *To Own, To Finance, and to Insure; Residential Real Estate Revealed*, Promoter(s): Prof.dr. D. Brounen, EPS-2013-290-F&A, http://hdl.handle.net/1765/40964

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries*, Promoter(s): Prof.dr. H.E. Haralambides, EPS-2010-204-LIS, http://hdl.handle.net/1765/19881

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promoter(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://hdl.handle.net/1765/ 31174

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promoter(s): Prof.dr. D. De Cremer & Prof.dr. E. van Dijk, EPS-2011-232-ORG, http://hdl.handle.net/1765/23268

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and Their Relationship with the Performance of Sales Managers*, Promoter(s): Prof.dr. W.J.M.I. Verbeke, EPS-2010-215-MKT, http://hdl.handle.net/1765/21188

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promoter(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://hdl.handle.net/1765/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promoter(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-258-STR, http://hdl.handle.net/1765/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory*, Promoter(s): Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://hdl.handle.net/1765/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promoter(s): Prof.dr. A. de Jong, EPS-2011-240-F&A, http://hdl.handle.net/1765/26041

Duursema, H., *Strategic Leadership; Moving Beyond the Leader-follower Dyad*, Promoter(s): Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://hdl.handle.net/1765/ 39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promoter(s): Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://hdl.handle.net/1765/26509

Essen, M. van, *An Institution-Based View of Ownership*, Promoter(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://hdl.handle.net/1765/22643

Feng, L., *Motivation, Coordination and Cognition in Cooperatives*, Promoter(s): Prof.dr. G.W.J. Hendrikse, EPS-2010-220-ORG, http://hdl.handle.net/1765/21680

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promoter(s): Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://hdl.handle.net/1765/ 37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://hdl.handle.net/1765/ 38028

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promoter(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-256-MKT, http://hdl.handle.net/1765/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promoter(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://hdl.handle.net/1765/37170

Hakimi, N.A, *Leader Empowering Behaviour: The Leader's Perspective: Understanding the Motivation behind Leader Empowering Behaviour*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2010-184-ORG, http://hdl.handle.net/1765/17701

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*, Promoter(s): Prof.dr. A. Jolink & Prof.dr. S.J. Magala, EPS-2010-193-ORG, http://hdl.handle.net/1765/19494

Hernandez Mireles, C., *Marketing Modeling for New Products*, Promoter(s): Prof.dr. P.H. Franses, EPS-2010-202-MKT, http://hdl.handle.net/1765/19878

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promoter(s): Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://hdl.handle.net/1765/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://hdl.handle.net/1765/32167

Hoever, I.J., *Diversity and Creativity: In Search of Synergy*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://hdl.handle.net/1765/37392

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promoter(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://hdl.handle.net/1765/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promoter(s): Prof.dr. D. De Cremer & Dr. M. van Dijke, EPS-2011-244-ORG, http://hdl.handle.net/1765/26228

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesign*, Promoter(s): Prof.dr. M.J.C.M. Verbeek & Prof.dr. R.J. Mahieu, EPS-2010-196-F&A, http://hdl.handle.net/1765/19674

Hytönen, K.A. *Context Effects in Valuation, Judgment and Choice*, Promoter(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://hdl.handle.net/1765/30668

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promoter(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://hdl.handle.net/1765/ 39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promoter(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://hdl.handle.net/1765/22156

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems*, Promoter(s): Prof.dr. P.J.F. Groenen, EPS-2010-195-MKT, http://hdl.handle.net/1765/19532

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promoter(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://hdl.handle.net/1765/23610

Karreman, B., *Financial Services and Emerging* Markets, Promoter(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://hdl.handle.net/1765/ 22280

Kil, J.C.M., *Acquisitions Through a Behavioral and Real Options Lens*, Promoter(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://hdl.handle.net/1765/50142

Klooster, E. van't, *Travel to Learn: The Influence of Cultural Distance on Competence Development in Educational Travel,* Promoter: Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312 – MKT, http://hdl.handle.net/1765/1

Konter, D.J., *Crossing borders with HRM: An inquiry of the influence of contextual differences in the adaption and effectiveness of HRM*, Promotor: Prof.dr. J. Paauwe, EPS-2014-305-ORG, http://hdl.handle.net/1765/1

Lam, K.Y., *Reliability and Rankings*, Promoter(s): Prof.dr. P.H.B.F. Franses, EPS-2011-230-MKT, http://hdl.handle.net/1765/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promoter(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://hdl.handle.net/1765/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promoter(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://hdl.handle.net/1765/23504

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models*, Promoter(s): Prof.dr.ir. J. Dul & Prof.dr. M.B.M. de Koster, EPS-2010-217-LIS, http://hdl.handle.net/1765/21527

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apoligize*, Promoter(s): Prof.dr. D. De Cremer, EPS-2014-301-ORG, http://hdl.handle.net/1765/1

Liang, Q., *Governance, CEO Indentity, and Quality Provision of Farmer Cooperatives*, Promoter(s): Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://hdl.handle.net/1765/1

Liket, K.C., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promoter: Prof.dr. H.R. Commandeur, EPS-2014-307-S&E, http://hdl.handle.net/1765/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones; New Frontiers in Entrepreneurship Research*, Promoter(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://hdl.handle.net/1765/ 40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promoter(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://hdl.handle.net/1765/ 22814

Lu, Y., *Data-Driven Decision Making in Auction Markets,Promotors:* Prof.dr.ir.H.W.G.M. van Heck & Prof.dr.W.Ketter, EPS-2014-314-LIS, http://hdl.handle.net/1765/1

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promoter(s): Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://hdl.handle.net/1765/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promoter(s): Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://hdl.handle.net/1765/34930

Meuer, J., *Configurations of Inter-Firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promoter(s): Prof.dr. B. Krug, EPS-2011-228-ORG, http://hdl.handle.net/1765/22745

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promoter(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://hdl.handle.net/1765/32343

Milea, V., *New Analytics for Financial Decision Support*, Promoter(s): Prof.dr.ir. U. Kaymak, EPS-2013-275-LIS, http://hdl.handle.net/1765/ 38673

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in Short-term Planning and in Disruption Management*, Promoter(s): Prof.dr. L.G. Kroon, EPS-2011-224-LIS, http://hdl.handle.net/1765/22444

Nijdam, M.H., *Leader Firms: The Value of Companies for the Competitiveness of the Rotterdam Seaport Cluster*, Promoter(s): Prof.dr. R.J.M. van Tulder, EPS-2010-216-ORG, http://hdl.handle.net/1765/21405

Noordegraaf-Eelens, L.H.J., *Contested Communication: A Critical Analysis of Central Bank Speech*, Promoter(s): Prof.dr. Ph.H.B.F. Franses, EPS-2010-209-MKT, http://hdl.handle.net/1765/21061

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promoter(s): Prof.dr. G. van der Pijl & Prof.dr. H. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://hdl.handle.net/1765/34928

Oosterhout, M., van, *Business Agility and Information Technology in Service Organizations*, Promoter(s): Prof,dr.ir. H.W.G.M. van Heck, EPS-2010-198-LIS, http://hdl.handle.net/1765/19805

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on Bureaucracy and Formal Rules*, Promoter(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://hdl.handle.net/1765/23250

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private*, Promoter(s): Prof.dr. L. van den Berg, EPS-2010-219-ORG, http://hdl.handle.net/1765/21585

Ozdemir, M.N., *Project-level Governance, Monetary Incentives and Performance in Strategic R&D Alliances*, Promoter(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://hdl.handle.net/1765/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promoter(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://hdl.handle.net/1765/ 30586

Pince, C., *Advances in Inventory Management: Dynamic Models*, Promoter(s): Prof.dr.ir. R. Dekker, EPS-2010-199-LIS, http://hdl.handle.net/1765/19867

Porck, J.P., *No Team is an Island*, Promoter(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-299-ORG, http://hdl.handle.net/1765/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promoter(s): Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://hdl.handle.net/1765/30848

Potthoff, D., *Railway Crew Rescheduling: Novel Approaches and Extensions*, Promoter(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon, EPS-2010-210-LIS, http://hdl.handle.net/1765/21084

Poruthiyil, P.V., *Steering Through: How Organizations Negotiate Permanent Uncertainty and Unresolvable Choices*, Promoter(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S. Magala, EPS-2011-245-ORG, http://hdl.handle.net/1765/26392

Pourakbar, M. *End-of-Life Inventory Decisions of Service Parts*, Promoter(s): Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://hdl.handle.net/1765/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promoter(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2013-282-S&E, http://hdl.handle.net/1765/39654

Retel Helmrich, M.J., *Green Lot-Sizing*, Promoter(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-291-LIS, http://hdl.handle.net/1765/41330

Rijsenbilt, J.A., *CEO Narcissism; Measurement and Impact*, Promoter(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://hdl.handle.net/1765/ 23554

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*, Promoter(s): Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas RA, EPS-2010-190-F&A, http://hdl.handle.net/1765/18013

Roza, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of Innovation, Absorptive Capacity and Firm Size*, Promoter(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://hdl.handle.net/1765/22155

Rubbaniy, G., *Investment Behavior of Institutional Investors*, Promoter(s): Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://hdl.handle.net/1765/ 40068

Schellekens, G.A.C., *Language Abstraction in Word of Mouth*, Promoter(s): Prof.dr.ir. A. Smidts, EPS-2010-218-MKT, http://hdl.handle.net/1765/21580

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promoter(s): Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://hdl.handle.net/1765/39655

Sotgiu, F., *Not All Promotions are Made Equal: From the Effects of a Price War to Cross-chain Cannibalization*, Promoter(s): Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga, EPS-2010-203-MKT, http://hdl.handle.net/1765/19714

Sousa, M., *Servant Leadership to the Test: New Perspectives and Insight,* Promotors: Prof.dr. D. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://hdl.net/1765/1

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promoter(s): Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://hdl.handle.net/1765/41513

Srour, F.J., *Dissecting Drayage: An Examination of Structure, Information, and Control in Drayage Operations*, Promoter(s): Prof.dr. S.L. van de Velde, EPS-2010-186-LIS, http://hdl.handle.net/1765/18231

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promoter: Prof.dr. S.J. Magala, EPS-2014-308-ORG, http://hdl.handle.net/1765/50712

Stallen, M., *Social Context Effects on Decision-Making; A Neurobiological Approach*, Promoter(s): Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://hdl.handle.net/1765/ 39931

Tarakci, M., *Behavioral Strategy; Strategic Consensus, Power and Networks*, Promoter(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-280-ORG, http://hdl.handle.net/1765/ 39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promoter(s): Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://hdl.handle.net/1765/37265

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the Pursuit of Exploration and Exploitation through Differentiation, Integration, Contextual and Individual Attributes*, Promoter(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-191-STR, http://hdl.handle.net/1765/18457

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments*, Promoter(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. P.H.M. Vervest, EPS-2010-201-LIS, http://hdl.handle.net/1765/19868

Tröster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://hdl.handle.net/1765/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision Making*, Promoter(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://hdl.handle.net/1765/ 37542

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers about Advice Taking in Marketing*, Promoter(s): Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga, EPS-2010-211-MKT, hdl.handle.net/1765/21149

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents*, Promoter(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-212-STR, hdl.handle.net/1765/21150

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promoter(s): Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://hdl.handle.net/1765/41511

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains*, Promoter(s): Prof.dr.ir. J.A.E.E. van Nunen, EPS-2010-194-LIS, http://hdl.handle.net/1765/19594

Venus, M., *Demystifying Visionary Leadership; In Search of the Essence of Effective Vision Communication*, Promoter(s): Prof.dr. D.L. van Knippenberg, EPS-2013-289-ORG, http://hdl.handle.net/1765/ 40079

Visser, V., *Leader Affect and Leader Effectiveness; How Leader Affective Displays Influence Follower Outcomes*, Promoter(s): Prof.dr. D. van Knippenberg, EPS-2013-286-ORG, http://hdl.handle.net/1765/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promoter(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://hdl.handle.net/1765/30585

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*, Promoter(s): Prof.dr. H.W. Volberda & Prof.dr. J. Soeters, EPS-2010-189-STR, http://hdl.handle.net/1765/18012

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promoter(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248-LIS, http://hdl.handle.net/1765/26564

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promoter(s): Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://hdl.handle.net/1765/26066

Wang, Y., *Corporate Reputation Management; Reaching Out to Find Stakeholders*, Promoter(s): Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://hdl.handle.net/1765/ 38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promoters: Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://hdl.handle.net/1765/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value of Dutch Firms*, Promoter(s): Prof.dr. A. de Jong, EPS-2013-277-F&A, http://hdl.handle.net/1765/ 39127

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*, Promoter(s): Prof.dr. M.J.C.M. Verbeek, EPS-2010-188-F&A, http://hdl.handle.net/1765/18125

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promoter(s): Prof.dr. M.B.M. de Koster, EPS-2013-276-LIS, http://hdl.handle.net/1765/38766

Zhang, D., *Essays in Executive Compensation*, Promoter(s): Prof.dr. I. Dittmann, EPS-2012-261-F&A, http://hdl.handle.net/1765/32344

Zhang, X., *Scheduling with Time Lags*, Promoter(s): Prof.dr. S.L. van de Velde, EPS-2010-206-LIS, http://hdl.handle.net/1765/19928

Zhou, H., *Knowledge, Entrepreneurship and Performance: Evidence from Country-level and Firm-level Studies*, Promoter(s): Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner, EPS-2010-207-ORG, http://hdl.handle.net/1765/20634

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promoter(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://hdl.handle.net/1765/23422

Erasmus Research Institute of Management - ERiM

CONDITIONAL DENSITY MODELS INTEGRATING FUZZY AND PROBABILISTIC REPRESENTATIONS OF UNCERTAINTY

Conditional density estimation is an important problem in a variety of areas such as system identification, machine learning, artificial intelligence, empirical economics, macroeconomic analysis, quantitative finance and risk management.

This work considers the general problem of conditional density estimation, i.e., estimating and predicting the density of a response variable as a function of covariates. The semi-parametric models proposed and developed in this work combine fuzzy and probabilistic representations of uncertainty, while making very few assumptions regarding the functional form of the response variable's density or changes of the functional form across the space of covariates. These models possess sufficient generalization power to approximate a non-standard density and the ability to describe the underlying process using simple linguistic descriptors despite the complexity and possible non-linearity of this process.

These novel models are applied to real world quantitative finance and risk management problems by analyzing financial time-series data containing non-trivial statistical properties, such as fat tails, asymmetric distributions and changing variation over time.

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM PhD Series

Research in Management

Erasmus Research Institute of Management - ERiM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl