

Asymmetry and fat-tails in financial time series

Citation for published version (APA):

Laurent, S. (2002). *Asymmetry and fat-tails in financial time series*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.20020620sl>

Document status and date:

Published: 01/01/2002

DOI:

[10.26481/dis.20020620sl](https://doi.org/10.26481/dis.20020620sl)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Asymmetry and fat-tails in financial time series

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht, op gezag van de Rector Magnificus, Prof. dr. A.C. Nieuwenhuijzen Kruseman volgens het besluit van het College van Decanen, in het openbaar te verdedigen op donderdag 20 juni 2002 om 14.00 uur

door

Sébastien Laurent

Promotor: Prof. dr. F.C. Palm

Co-Promotor: Dr. J.-P. Urbain

Beoordelingscommissie: Prof. dr. P.C. Schotman (voorzitter)

Prof. dr. L. Bauwens (Université catholique de Louvain)

Dr. S. Schim van der Loeff

Contents

1	General Introduction	1
I	Modelling financial data	13
2	Skewed Student distribution and GARCH models	15
2.1	Introduction	15
2.2	Distribution choices	18
2.2.1	The normal distribution	18
2.2.2	Mixture of normal distributions	19
2.2.3	The Student distribution	21
2.2.4	Skewed densities	21
2.3	GARCH model with skewed distribution for the innovations	29
2.4	Density forecasts	31
2.5	Application	37
2.6	Asymmetry, fat-tails and Additive Outliers	42
2.7	Conclusion	46
3	Analytical scores and Gaussian QML relative efficiency	49
3.1	Introduction	49
3.2	The model	50
3.3	The log-likelihood function	50
3.4	Analytical gradients of the skewed Student density	52
3.5	Relative efficiency of QML estimator	55
3.6	APARCH specification	59
3.7	Empirical application	63
3.8	Conclusion	64
4	A new class of multivariate skewed Densities	67
4.1	Introduction	67
4.2	Univariate case	69
4.2.1	Skewed Student densities	69
4.2.2	Empirical illustration	70

4.3	Multivariate case	74
4.3.1	Multivariate symmetrical densities	75
4.3.2	Multivariate skewed densities	77
4.3.3	Simulation	84
4.3.4	Multivariate skewed densities with independent components	86
4.4	Empirical application	88
4.5	Conclusion	95
 II Applications		97
 5 Value-at-Risk for Long and Short Positions		99
5.1	Introduction	99
5.2	VaR models	102
5.2.1	Symmetric and asymmetric volatility models	102
5.2.2	VaR for long and short positions	103
5.3	Empirical application	105
5.3.1	Data	105
5.3.2	Estimating the models	105
5.3.3	In-sample VaR computation	113
5.3.4	Out-of-sample VaR computation	117
5.3.5	Expected short-fall and related measures	118
5.4	Conclusion	119
 6 Daily Value-at-Risk using realized volatility		125
6.1	Introduction	125
6.2	Data and stylized facts	127
6.2.1	Data	127
6.2.2	Realized volatility: stylized facts	129
6.3	Two competing models	133
6.3.1	The skewed Student APARCH model	134
6.3.2	Forecasting realized volatility	135
6.3.3	Assessing the VaR performance of the models	137
6.4	Empirical application	137
6.4.1	VaR, daily returns and the skewed Student APARCH	137
6.4.2	VaR, intraday returns and daily realized volatility	140
6.4.3	Which model is best?	145
6.5	Conclusion	146
 7 Central Bank Intervention and Exchange Rate Volatility		149
7.1	Introduction	149
7.2	Regime-dependent frameworks	152

7.2.1	Regime-dependent models versus single regime (G)ARCH models	152
7.2.2	Results and comparison with GARCH model	156
7.2.3	Forecasting Performance	159
7.3	The impact of central bank interventions	162
7.3.1	The TVTP model	162
7.3.2	The intervention data	164
7.3.3	The results	166
7.4	Conclusion	171
8	Conclusion	173
A	G@RCH 2.0: An Ox Package for ARCH Models	177
A.1	Introduction	177
A.2	Features of the package	179
A.2.1	Mean equation	179
A.2.2	Variance equation	180
A.3	Estimation methods	190
A.3.1	Parameters constraints	190
A.3.2	Distributions	191
A.3.3	Tests	192
A.3.4	Forecasts	193
A.3.5	Numerical accuracy	197
A.3.6	Features comparison	199
A.4	Application	199
A.4.1	Data and methodology	199
A.4.2	Using the “Full Version”	201
A.4.3	Using the “Light Version”	208
A.5	Conclusions	216
B	Appendix Chapter 2	217
C	Appendix Chapter 7	221
	Nederlandse Vertaling	243
	Curriculum Vitae	245

List of Tables

2.1	Monte Carlo analysis: skewed Student errors.	32
2.2	Monte Carlo analysis: Chi-square errors.	33
2.3	Monte Carlo analysis: Gamma errors.	34
2.4	AR(1) - APARCH (1, 1) model. Estimation results.	41
2.5	Statistics of interest.	42
2.6	Density forecast tests for the out-of-sample forecasts.	43
2.7	AR(1) - APARCH (1, 1) model. Estimation results for AO corrected returns ($C = 5.5$).	45
3.1	Relative efficiency of QML estimator	60
3.2	Skewed Student APARCH	64
4.1	ML estimation results of AR-GARCH models for the CAC40 and the NASDAQ	71
4.2	ML estimation results of AR-GARCH models for the NIKKEI and the SMI	72
4.3	QML estimation results of simple skewed Student DGP	87
4.4	ML estimation results of AR-TVC-GARCH model: normal and Student distributions	90
4.5	ML estimation results of AR-TVC-GARCH model: skewed Student and skewed Student with IC distributions	91
5.1	Descriptive statistics	111
5.2	Skewed Student APARCH	112
5.3	VaR results for NASDAQ and NIKKEI (in-sample)	115
5.4	VaR results for all indexes (in-sample)	116
5.5	VaR results	118
5.6	Expected short-fall for NASDAQ and NIKKEI	120
5.7	Average multiple of tail event to risk measure	121
6.1	Skewed Student APARCH	138
6.2	VaR results for the CAC40 and SP500	139
6.3	Asymmetric ARFIMA	141
6.4	Ex-ante standardized returns	143

6.5	VaR results for the CAC40 and SP500	144
7.1	Markov switching models: DEM (1985-1995)	157
7.2	Markov switching models: YEN (1991-1995)	158
7.3	Variance forecasts for the models	161
7.4	Official and reported central bank interventions, number of days . .	165
7.5	Cross correlations between central bank interventions	166
7.6	Cross correlations between central bank interventions	166
7.7	Central bank interventions, DEM (1985-1995)	167
7.8	Central bank interventions, YEN (1991-1995)	168
A.1	Accuracy of the GARCH procedure	197
A.2	GARCH accuracy comparison	198
A.3	Accuracy of the FIGARCH procedure	198
A.4	GARCH features comparison	200

List of Figures

2.1	Skewed Student densities with $v = 8$ and $\xi = 1, 1.5$ and 3 .	23
2.2	Skewness implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $3.5 \leq v \leq 15$.	25
2.3	Skewness implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $3.05 \leq v \leq 3.5$.	26
2.4	Kurtosis implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $4.5 \leq v \leq 15$.	26
2.5	Kurtosis implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $4.05 \leq v \leq 4.5$.	27
2.6	Normal, Student and skewed Student densities.	28
2.7	$\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $ST(0, 1, 8)$ errors. The MLEs were computed assuming normality for the innovations.	36
2.8	$\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $SKST(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming the innovations to be Student distributed.	36
2.9	$\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $SKST(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming that the innovations are skewed Student distributed.	37
2.10	$st(0, 1, \exp(-0.179), 6.039)$ (solid line), $t(0, 1, 5.57)$ (dashed line) and $N(0, 1)$ (short dashes).	43
2.11	NASDAQ (solid line) corrected for 76 AO (circles).	46
3.1	Difference between numerical and analytical scores of the log-likelihood with respect to $\ln(\xi)$ for $v = 8$.	54
3.2	Difference between numerical and analytical scores of the log-likelihood with respect to v , evaluated at $\ln(\xi) = 0$.	55
4.1	Graph of the $SKST(0, I_2, (1, 1.3), 6)$ density.	84
4.2	Panel A refers to the contours of the bivariate $SKST(0, I_2, (1, 1.3), 6)$ density illustrated in Figure 4.1. Panel B refers to the contours of a $SKST-IC(0, I_k, (1, 1.3), (6, 6))$ (see Section 4.3.4).	85

4.3	Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a normal likelihood (with 20 cells).	94
4.4	Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a skewed Student likelihood (with 20 cells).	94
5.1	CAC 40 stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 1/1/1990 - 21/12/2000.	106
5.2	DAX stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 26/11/1990 - 21/12/2000.	107
5.3	NASDAQ stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 11/10/1984 - 21/12/2000.	108
5.4	NIKKEI stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 4/1/1984 - 21/12/2000.	109
5.5	SMI stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 9/11/1990 - 21/12/2000.	110
6.1	Logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.	131
6.2	Density estimates (dashed line) and corresponding normal density (solid line) for the logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.	131
6.3	First 200 autocorrelations for the logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns. The horizontal lines show the upper limit 95% Bartlett confidence bands.	132
6.4	Regression lines for the logarithmic realized volatility (y-axis) of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns against the previous (i.e. one day before) returns (x-axis).	134
6.5	The graphs display the density distributions, i.e. empirical (dashed lines) vs normal (solid lines), for the daily returns standardized with respect to the square root of the ex-post (left panel) and the ex-ante (right panel) daily realized volatility computed for the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.	145

7.1	Conditional variances: GARCH vs. two-regime model.	162
A.1	Selecting the variables	202
A.2	Model settings	203
A.3	Selecting the starting values method	203
A.4	Standard errors estimation methods	204
A.5	Entering the starting values	204
A.6	Graphics menu	208
A.7	Graphical analysis	209
A.8	Tests dialog box	210
A.9	Forecasts from an AR(1)-APARCH(1,1).	210
A.10	Density forecast analysis	215

Acknowledgements

First of all I would like to dedicate this dissertation to Christelle and Valentin and thank them for their support, encouragement, patience and love.

I cannot refrain to thank my supervisors, Franz Palm and Jean-Pierre Urbain for their expertise and their millions of useful comments (even in the footnotes of the tables of the appendixes). It is a pleasure for me to work with both.

Many thanks go also to the Department of Quantitative Economics of the Faculty of Economics and Business Administration of Maastricht University for giving me a warm welcome (except perhaps the terrible twins A.W.J. Hecq and B.C.B. Candelon). Thank you Karin for your invaluable help.

I would like also to thank the other members of my dissertation committee, Peter Schotman, Sybrand Schim van der Loeff and Luc Bauwens for their comments and suggestions.

I had the chance to meet and work with several coauthors. They gave me the opportunity to open my eyes to new topics and to have interesting discussions. Let me present them chronologically: Christelle Lecourt, Michel Beine, Philippe Lambert, Jean-Philippe Peters, Pierre Giot and Luc Bauwens.

The main part of my research was made possible by the Department of Economics at the University of Liège, in Belgium. I would like to thank all my colleagues and the members of the Department for giving me a lot of freedom and the financial support to participate in several congresses. In particular, I thank Sergio Perelman and Pierre Pestieau for their encouragements and Bernard Lejeune for valuable comments on my research.

Part of this thesis has also been done in Louvain-la-Neuve. Indeed, I visited the CORE every Wednesday during the last year (the day of the econometric seminar). I thank Luc Bauwens for this invitation and his receptiveness but also Helena Beltran for the nice discussions during the breaks and Jeroen Rombouts for his help for translating the summary in Dutch and for his knowledge on multivariate GARCH models.

Last but not least, I would like to thank all my family for giving me moral support whenever I needed it. Thank you dad and grandmother for all your love and care.

Chapter 1

General Introduction

It has been recognized for a long time that the dynamic behavior of economic variables is difficult to understand. And this difficulty certainly increases with the observation frequency of the data.

Traditional regression tools have shown their limitation in the modelling of high-frequency (weekly, daily or intra-daily) data. Assuming that only the mean response could be changing with covariates while the variance remains constant over time often revealed to be an unrealistic assumption in practice. This fact is particularly obvious in series of financial data where clusters of volatility can be detected visually.

Understanding and predicting the temporal dependence in the second moment is crucially important for many issues in macroeconomics and finance. We may distinguish at least three main categories of applications. First, from a statistical point of view, modelling the heteroscedastic feature of the data can provide more efficient estimates of the conditional mean and more precise confidence bands for the forecasts. Testing economic (or financial) theories is the second potential application. For instance, one can mention the growing literature aiming at testing the effectiveness of central bank interventions in reducing the volatility on the foreign exchange market (see Dominguez and Frankel, 1993). Reading these first two potential applications, one could imagine that the modelling of the conditional variance has only an academic purpose and is certainly not used by practitioners. This is far from being the truth. Although at the beginning of the 1990s, it was still possible to consider it as a decision tool in an experimental phase, several institutions have now developed the necessary skills to use the econometric theory

in portfolio management (Gourieroux, 1997).

The increased importance played by risk and uncertainty considerations in modern economic theory has called for the development of new econometric time series techniques that allow for the modelling of time varying means, variances and covariances. Given the apparent lack of any structural dynamic economic theory explaining the variation in the second moment, econometricians have thus extended traditional time series tools such as Autoregressive Moving Average (ARMA) models (Box and Jenkins, 1970) for the mean to essentially equivalent models for the variance. Indeed, the dynamics observed in the dispersion is clearly the dominating feature in the data. Autoregressive Conditional Heteroscedasticity (ARCH) models (Engle, 1982) are now commonly used to describe and forecast changes in volatility of financial time series.

Although ARCH-type models have met with substantial empirical success, using this class of models is not the only way to model the time-varying conditional variance in a parametric framework. Indeed, Stochastic Volatility (SV) models (Taylor, 1986) are also popular in finance (see Shephard, 1996 for a survey). As mentioned by Palm (1996), a major difficulty arises with the estimation of SV models which are non-linear and not conditionally Gaussian. However, with the advances in simulation technology, estimation of SV models is now less cumbersome.

Another response to the overwhelming variety of parametric univariate ARCH models, is to consider and estimate nonparametric (NP) or semiparametric (SP) models. For instance, Pagan and Hong (1990) explored a NP kernel estimate of the expected value of the squared returns. Pagan and Schwert (1990) used a collection of standard NP estimation methods, including kernels, Fourier series and least squares regressions, to fit models for the relation between squared returns and past squared returns. Effectively, the main difficulty of this approach relies in the estimation of the function that links the squared returns to its past values.

Unlike nonparametric models, ARCH models are typically estimated by maximizing the associated log-likelihood function or a quasi-log-likelihood function (see Gourieroux, 1997 for a review of alternative estimation procedures of ARCH models). Consequently, one has to make an additional assumption about the innovation process. It is usual to rely on a conditional Gaussian log-likelihood since the Gaussian Quasi Maximum Likelihood (QML) method can provide consistent

estimates in the general framework of a dynamic model under correct specification of both the conditional mean and the conditional variance (see Weiss, 1986 and Bollerslev and Wooldridge, 1992 among others).

However, another striking characteristic of high-frequency financial returns is that they are often fat-tailed. For instance, Hong (1988) rejected the conditional normality claiming abnormally high kurtosis in the daily New York Stock Exchange stock returns. In fact, the kurtosis of most asset returns is higher than three, which means that extreme values are observed more frequently than for the normal distribution. While the high kurtosis of the returns is a well-established fact, the situation is much more obscure with regard to the symmetry of the distribution. Many authors do not observe anything special on this point, but other researchers (for instance Simkowitz and Beedles, 1980; Kon, 1984 and So, 1987) have drawn the attention to the asymmetry of the distribution in the sense that the unconditional mean and the unconditional mode do not coincide. When the mean is at the right (resp. left) of the mode, the series is said to be right (resp. left) skewed. For instance, French, Schwert, and Stambaugh (1987) found conditional skewness significantly different from 0 in the standardized residuals when an ARCH-type model was fitted to the daily SP500 returns. More recently, Mittnik and Paoletta (2000) have shown that an asymmetric and fat-tailed distribution is required for modelling several daily exchange rate returns of East Asian currencies against the US dollar.

Basically, searching for a more realistic assumption for the innovation process has two sources of motivation. The first *raison d'être*, is to have more efficient estimates (which is of prime importance for statistical inference). Indeed although consistent, the Gaussian QML estimator is inefficient for non-normally distributed data, with the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera, 1991). This leaves the door open for other distribution functions and/or other estimation techniques. Second, accounting for asymmetry and fat-tails is relevant for financial applications. For instance, the Capital Asset Pricing Model (CAPM) assumes that only the means, variances and covariances of returns matter in asset pricing, and, therefore, higher-order moments are unimportant. Upside and downside risks are considered equally probable by investors, but this assumption is not reasonable given that most investors have a preference for positive skewness (Peiró, 1999). Moreover, as shown by Brennan

(1979) and He and Leland (1993), if the market's portfolio rate of return has constant mean and volatility, the average investor is risk averse. This implies skewness preference that is positively valued by investors, which means that modelling the third moment is required in several financial applications. For instance, Kraus and Litzenberger (1976) extend the CAPM to include the effect of skewness on valuation, and present empirical evidence consistent with their extension.

While it might be agreed that it is desirable to allow the conditional density to be non-normal, it is not clear how to achieve this goal.

In a SP context, Gallant and Tauchen (1989) and Gallant, Hsieh, and Tauchen (1991) propose to model the joint density of the data using a series expansion with a Gaussian Vector Autoregressive (VAR) leading term. This is an innovative approach since it has the potential to reveal a lot of information concerning the underlying distribution without imposing a great deal of a priori information or structure. However, as mentioned by Hansen (1994), this method has several drawbacks: it requires very large data sets in order to have a reasonable degree of precision, the methods are computationally demanding (this method will probably remain primarily in the hands of specialists) and the techniques may be sensitive to the choice of the number of expansion terms.

Alternatively, Engle and González-Rivera (1991) propose to estimate the conditional density nonparametrically using a three steps approach. First they estimate the parameters of the model by QML using a Gaussian pseudo-likelihood function. The density of the residuals standardized by their estimated conditional standard deviations is then estimated in a second step using a linear spline with smoothness priors. The third step consists in maximizing the likelihood function considering the estimated density of the second step as the true density. In a Monte Carlo study, this approach was found to improve the efficiency beyond the QML estimator, particularly when the density was highly non-normal and skewed.

Instead of using SP or NP techniques, a third approach would be to search for a flexible parametric error distribution (coupled with an ARCH-type model for instance). In order to accommodate the excess of (unconditional-) kurtosis, GARCH models have been first combined with Student distributed errors by Bollerslev (1987). Indeed, although GARCH models generate fat-tails in the unconditional distribution, when combined with a Gaussian conditional density, they do not fully account for the excess kurtosis present in many return series. The Student density

is now very popular in the literature due to its simplicity and because it often outperforms the Gaussian density. However, the main drawback of this density is that it is symmetrical while financial time series can be skewed. To create asymmetric unconditional densities, GARCH models have been extended to include a leverage effect. For instance, the threshold ARCH (TARCH) model of Zakoian (1994), the exponential GARCH (EGARCH) of Nelson (1991) or the asymmetric power ARCH (APARCH) of Ding, Granger, and Engle (1993) allow past negative (resp. positive) shocks to have a deeper impact on current conditional volatility than past positive (resp. negative) shocks (see among others Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990). Combined with a Student distribution for the errors, this model is in general flexible enough to mimic the observed kurtosis of many stock returns but often fails in replicating the asymmetry of these series (even if it can explain a small part of it, see Section 2.5).

To account for both excess skewness and excess kurtosis, mixtures of normal or Student densities can be used in combination with a GARCH model. In general, it has been found that these densities cannot capture all the skewness and leptokurtosis (Ball and Roma, 1993; Beine and Laurent, 1999; Jorion, 1988; Neely, 1999; Vlaar and Palm, 1993), although they seem adequate in some cases. McCulloch (1985), Liu and Brorsen (1995), Mittnik, Paoletta, and Rachev (1998) and Lambert and Laurent (2000) consider the asymmetric stable density in combination with a GARCH model. A major drawback of the stable density is that, except when the tail parameter $\alpha = 2$ (i.e. normality), the variance does not exist, a fact usually not supported empirically (see Pagan, 1996). Lee and Tse (1991), Knight, Satchell, and Tran (1995) and Harvey and Siddique (1999)¹ propose alternative skewed fat-tailed densities, with respectively the Gram-Charlier Expansion, the Double-Gamma distribution and the non-central t . However, as pointed out by Bond (2000) in a recent survey on asymmetric conditional density functions, estimation of these densities in a GARCH framework often proved troublesome and highly sensitive to initial values. McDonald (1984) and McDonald (1991) introduce the exponential generalized beta distribution of the second kind (EGB2), a flexible distribution that is able to accommodate not only thick tails but also asymmetry. The usefulness of this density has been proved recently by Wang, Fawson, Barrett, and McDonald (2001) in a GARCH framework. These authors show that a more

¹This list is by no means exhaustive.

flexible density than the normal and the Student is required in the modelling of six daily nominal exchange rate returns vis-a-vis the US dollar. However, goodness-of-fit tests clearly reject the EGB2 distribution for all the currencies that they consider, even if it seems that it outperforms the normal and the Student.

Interestingly, Hansen (1994) proposes a skewed Student distribution that nests the symmetric Student when the asymmetry coefficient (λ) equals 0, with $-1 < \lambda < 1$. This density is quite easy to implement and its estimation does not face serious problems of convergence. However, Hansen (1994) does not discuss the relation between λ and higher moments. Recently, Jones and Faddy (2000) have designed another skew- t distribution. This density has two parameters (assuming zero location and unit scale parameters), say a and b . If $a = b$, the distribution is the usual symmetrical Student one, with number of degrees of freedom $\nu = 2b$ (assuming $b > 1$). If $a - b > 0$ (< 0), the density is skewed to the right (left): hence $a - b$ reflects the skewness feature of the density. A property of this skew- t density is that its long tail is thicker than its short tail (if $a > b$, the left tail behaves like $z^{-(2a+1)}$ at minus infinity, the long tail like $z^{-(2b+1)}$ at plus infinity). Jones and Faddy (2000) also provide the moments and the cumulative density function of their skew- t density.

Recently, (in a context different from the volatility literature) Fernández and Steel (1998) developed a more general tool (based on the method of inverse scaling of the probability density function on the left and the right of the mode) which has the advantages of simplicity and that all the parameters have a clear interpretation. Moreover, contrary to Hansen (1994), Fernández and Steel (1998) discuss the relation between the asymmetry coefficient and the first three moments. However, the main drawback of this density is that it is not expressed in terms of the mean and the variance but in terms of the mode and a measure of the dispersion.

The main purpose of this dissertation is to find an elegant way for modelling jointly and the most faithfully the first four conditional moments of high-frequency financial time series and to illustrate the potentiality of this specification in financial and economic applications. Priority is given to the simplicity of the approach. For this reason, but also for those given above, this thesis is not in the tradition of SP or NP techniques but is rather in line with fully parametric GARCH models.

For modelling the higher order features of the data, we propose to follow Fernández and Steel (1998) in using a standardized version of their skewed Student

density. Once again, this choice has been conducted by our concern of simplicity but also by the fact that this density seems adequate in presence of financial time series. Unlike most of its competitors, the main advantages of this family of skewed densities are that all the parameters have a clear interpretation and that it is easy to implement. For instance, the square of the asymmetry parameter of this density equals the ratio of probability masses above and below the mode and can be directly interpreted as a skewness measure. Moreover, its probability density function (pdf), cumulative density function (cdf) and quantile function are based on the corresponding functions of its symmetric versions. From an empirical point of view, this density seems to do a good job (when combined with an appropriate specification of the first two conditional moments) in modelling daily stock returns. Moreover, the results given in Chapters 5 and 6 suggest that it provides accurate VaR forecasts for the data we have considered.

The basic specification used throughout this thesis for modelling the first two conditional moments is an ARMA-APARCH. First, our choice is motivated by the fact that an AR with a low order was found to be sufficient for controlling the autocorrelation observed in the investigated series. Second, the extra flexibility of the APARCH specification (the leverage effect and the power coefficient) is justified for most of the series we have investigated. In particular, the APARCH specification models a Box-Cox transformation of the conditional standard deviation instead of the conditional variance. It has been motivated by a stylized fact detected by Taylor (1986) who first observed that the absolute returns (a proxy of the conditional standard deviation) in financial time series are positively autocorrelated, even at long lags. Ding, Granger, and Engle (1993) found that the closer the power coefficient is to 1, the larger the memory of the process is. Note that an exception to this choice is made in Chapter 4 where a simple GARCH model will be used to make easier the comparison with the multivariate analysis.

This allows us to introduce the second contribution of this thesis. Even if the dissertation focuses chiefly on univariate models, our main concern was to check whether the technique proposed by Fernández and Steel (1998) to introduce skewness in any univariate unimodal density could be extended in a multivariate context. Indeed, financial volatilities move together over time across assets and markets. Recognizing this commonality through a multivariate modelling framework can lead to obvious gains in efficiency and to improved financial decision

making than working with separate univariate models. As far as financial applications are concerned, it is of primary importance to base modelling and inference on a more suitable distribution than the multivariate normal. The challenge to econometricians is to design multivariate distributions that are both easy to use for inference and of course compatible with other properties of financial returns (e.g. autocorrelation, skewness, kurtosis, ...). Otherwise it is very likely that the parameter estimators will not be consistent (see Newey and Steigerwald, 1997).

The dissertation is split in two parts: a methodological part and a part presenting financial applications.

The first part focuses on methodological issues of the models used to analyze daily financial data and is made up of three chapters.

In the second chapter, entitled “Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations” and based on Lambert and Laurent (2001), we examine the issue of both skewness and fat-tails in financial time series. We first retain our attention on three candidate distributions: the normal, the Bernoulli-mixture of normals and the Student. Then, we propose an extended version of the skewed Student density of Fernández and Steel (1998) and show the usefulness of this density (coupled with an APARCH) when modelling a stock index (the NASDAQ in our example) and forecasting not only the mean and the variance but the whole density of this series. The results clearly suggest that the skewed Student density better fits the NASDAQ and is more appropriate to produce density forecasts of this series.

Chapter three, entitled “Analytical scores of the APARCH skewed Student model and Gaussian QML relative efficiency” (based on Laurent, 2001), derives analytical expressions for the score of the univariate skewed Student density as well as the score of the APARCH model presented in Chapter 2. The use of asymmetric and fat-tailed densities is growing in the literature. However, all the existing applications rely on numerical techniques to calculate the gradients. This chapter shows that the use of analytical gradients highly speeds up maximum-likelihood estimation and improves the numerical accuracy. It illustrates also the loss of efficiency of the Gaussian QML estimator when the innovations are skewed and/or fat-tailed.

Chapter four, entitled “A New Class of Multivariate Skewed Densities, with Application to GARCH Models” and based on Bauwens and Laurent (2002), pro-

poses a multivariate generalization of the family of skewed densities presented in Chapter 2. We describe a practical and flexible solution to introduce skewness in multivariate symmetrical distributions. Applying this procedure to the multivariate Student density leads to a “multivariate skewed Student” density, for which each marginal has a different asymmetry coefficient. Combined with a multivariate GARCH model, this new family of distributions is potentially useful for modelling stock returns. In an application to the daily returns of the French CAC40, German DAX, US NASDAQ, Japanese NIKKEI and Swiss SMI data, it is found that this density suits well the data and clearly outperforms its symmetric competitors (the multivariate normal and Student densities). This chapter ends the first part of this thesis.

The second part is made up of three chapters and is devoted to showing the usefulness of non-normal densities in two financial applications.

In recent years, the tremendous growth of trading activity and the well-publicized trading loss of well known financial institutions (see Jorion, 2000) has led financial regulators and supervisory committees of banks to favor quantitative techniques which appraise the possible loss that these institutions can incur. Value-at-Risk (VaR) has become one of the most sought-after techniques as it provides an easy-to-understand method for quantifying risk. Indeed, the VaR measures the worst expected loss over a given horizon under normal market conditions at a given confidence level, in other words it is a quantile. For instance, a bank might say that the daily VaR of its trading portfolio is \$35 million at the 99% confidence level. In other words, there is only 1 chance in a 100, under normal market conditions, for a loss greater than \$35 million to occur.

When using a full parametric approach, as the one used all along this thesis, the VaR produced by the model depends on the whole conditional density. Consequently, using a skewed density when the data are known to be skewed can be of primary importance to obtain accurate VaR forecast. Chapter five, “Value-at-Risk for long and Short Positions”, based on Giot and Laurent (2001c) and Giot and Laurent (2001b), shows that the APARCH skewed Student model performs very well in forecasting the VaR, compared with other parametric models, based on symmetric densities. Indeed, we compare the performance of this model with that of the RiskMetrics (which is a GARCH(1,1) model with fixed coefficients), normal and Student APARCH models and show that the APARCH model combined

with a skewed Student density brings about considerable improvements in correctly forecasting one-day-ahead VaR for long and short trading positions on daily stock indexes (French CAC40, German DAX, US NASDAQ, Japanese NIKKEI and Swiss SMI data). The performance of the models is assessed by computing Kupiec (1995)'s LR tests based on the empirical failure rates.

The recent availability of intraday data has led to new developments concerning the estimation of the daily volatility. The notion of realized volatility has been introduced recently in the literature by Taylor and Xu (1997) and Andersen and Bollerslev (1998) and is computed as an aggregated measure of volatility defined on intraday returns. According to these authors it offers an “error free” measure of the daily volatility. Interestingly, when one uses this realized volatility instead of the conditional variance produced by a parametric ARCH-type model, the normality assumption on the innovation process is supported. Does the use of the realized volatility invalidate the choice of a skewed Student density ? Can we use the Gaussian assumption in a VaR application based on the realized volatility ? We will answer these questions in Chapter six, “Modelling Daily Value-at-Risk using Realized Volatility and ARCH Type Models” (based on Giot and Laurent, 2001a).

The second main application is presented in Chapter seven. In this chapter, entitled “Official Central Bank Interventions and Exchange Rate Volatility: Evidence from a Regime Switching Analysis” and based on Beine, Laurent, and Lecourt (2001), we extend the static mixture of normal distributions presented in Chapter 2. Indeed, we assume that the evolution of the DEM/USD and YEN/USD exchange rate returns (in a weekly basis) depends on a latent regime variable whose dynamics is driven by a first-order Markov switching process. In contrast with previous analyzes, we allow for regime-dependent specifications and investigate whether official interventions may explain the observed volatility regime switches. The estimation results shed an interesting light on the conclusions given in the literature. It is found that depending on the prevailing volatility level, coordinated central bank interventions can have either a stabilizing or a destabilizing effect. Our results lead us to challenge the usual view that such interventions are necessarily associated with increases in volatility.

Chapter eight proposes some concluding remarks.

Finally Appendix A, entitled “G@RCH 2.0: An Ox Package for Estimating and Forecasting Various ARCH Models” (based on Laurent and Peters, 2001),

does not constitute a regular chapter. Indeed, it documents GARCH 2.0, an Ox package with a friendly dialog-oriented interface dedicated to the estimation and forecast of various univariate ARCH-type models. These models can be estimated by approximate (Q)ML under four assumptions: normal, Student- t , Generalized Error Distribution (GED) or skewed Student errors. This software should help researchers in their future applications dealing with univariate ARCH models.

Part I

Modelling financial data

Chapter 2

Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations

2.1 Introduction

During the past decade, the statistical analysis of financial time series has focused on the conditional second moment as most financial asset returns exhibit temporal bursts of volatility. ARCH models (Engle, 1982) and its various extensions (see Appendix A) are commonly used to describe the conditional variance while an ARMA structure is often considered for the conditional mean. For a survey on ARMA-ARCH type models, see Bera and Higgins (1993), Palm (1996) or Pagan (1996) among others.

Even if the choice of appropriate statistical models for the first two moments is crucial, the specification of the conditional distribution is also of primary importance. These sophisticated linear models for the conditional mean and for the conditional variance often rely on simplistic assumptions on the stochastic structure (normality). Indeed, it is widely accepted that financial returns, on a weekly, daily or intraday basis, are fat-tailed and even skewed.

ARCH-type models are usually estimated by ML with a Gaussian log-likelihood $L_{norm}(y|\Phi)$, where y and Φ denote respectively the vector of observations and the vector of parameters. It is well known that under regularity conditions, the value of Φ which maximizes L_{norm} , i.e. $\hat{\Phi}_{ML}$, is consistent, asymptotically normally

distributed and efficient. Its asymptotic covariance matrix can be consistently estimated by minus one times the inverse of the Hessian matrix evaluated at $\hat{\Phi}_{ML}$. In this respect, Lumsdaine (1996) proved the consistency and the asymptotic normality of the ML estimator of the (Integrated-)GARCH(1,1) under the condition that $E[\log(\alpha_1 \varepsilon_t^2 + \beta_1)] < 0$.¹

As explained above, the normality assumption is unrealistic with high-frequency financial data. However, if we are only interested in the first two conditional moments, this assumption may be justified by the fact that the Gaussian Quasi Maximum Likelihood (QML) can provide consistent (and asymptotically normally distributed) estimators assuming that the conditional mean and the conditional variance are specified correctly (Weiss, 1986; Bollerslev and Wooldridge, 1992). The Gaussian QML estimator of Φ is obtained by maximizing L_{norm} although the true probability density function is non-Gaussian. Lee and Hansen (1994) extended previous works and showed that if the conditional mean and conditional variance of a GARCH (1, 1) process are specified correctly, and one uses the Gaussian likelihood as a vehicle to estimate the corresponding parameters, the parameter will be consistently estimated even if the rescaled series (the residuals divided by the conditional standard deviation, i.e. z_t) is neither Gaussian nor independent. For more complicated specifications, it is fairly difficult to prove (theoretically) the consistency of the Gaussian QML estimator. For instance, Teysnière (1997) shows using Monte Carlo simulations and kernel density estimation, that the Gaussian QML of an ARFIMA-FIGARCH (Autoregressive Fractionally Integrated Moving Average - Fractionally Integrated GARCH) process seems to have nice properties: they are root-n consistent, asymptotically normal and the bias is negligible.

However, even if the QML estimator is consistent under certain conditions, it is inefficient with the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera, 1991). The asymptotic standard errors can be estimated consistently as was done by White (1982) and Gouriéroux, Monfort, and Trognon (1984), although they will not attain the Cramer-Rao bound, reflecting the penalty resulting from not knowing the true conditional density. Searching for a more suitable distribution may thus be motivated by the search of more efficient estimates.

More importantly, from a practical point of view, the issue of skewness (asym-

¹The notation will be clarified in the next sections.

metry) and kurtosis (fat-tails) is useful in many respects for financial applications. Peiró (1999) emphasizes the relevance of the modelling of higher-order features in asset pricing models, portfolio selection and option pricing theories.² Modelling skewness and kurtosis has an impact on all conditional quantiles. Therefore, not surprisingly, they are crucial in Value-at-Risk applications (see Chapters 5 and 6).

As pointed out by El Babsiri and Zakoian (1999), although asymmetric GARCH models can generate skewed unconditional densities by allowing positive and negative changes to have a different impact on future volatilities, the two components of the innovation have - up to a constant - the same volatilities, while it is desirable to allow an asymmetric confidence interval around the prediction value. In this respect, to model jointly the first four conditional moments in a fully parametric framework, several densities have been proposed in the literature (see Chapter 1 for a brief survey). Interestingly, Fernández and Steel (1998) develop a general tool (based on the method of inverse scaling of the probability density function on the left and the right of the mode) to introduce skewness in any continuous unimodal and symmetric density. However, the major drawback of this technique is that the skewed density is not expressed in terms of the mean and the variance but in terms of the mode and a measure of the dispersion. In order to keep in the ARCH tradition, we first re-express Fernández and Steel's (1998) density as a function of the mean and variance and derive its cumulative density function and quantile function. We also proceed to a Monte Carlo simulation to assess its practical applicability in a ML estimation procedure in the GARCH framework. Finally, we show the usefulness of this method by the analysis of the NASDAQ on the period 1985-1996. Using both in- and out-of-sample density forecast tests, we show that this density seems to be adequate in describing this database compared to the normal and the Student distributions.

The chapter is organized as follows. Section 2.2 reviews three candidate distributions before presenting the family of skewed densities proposed by Fernández and Steel (1998) and its standardization. Section 2.3 gives the results of a small Monte Carlo simulation while Section 2.4 summarizes the concept of density forecast. Section 2.5 provides our empirical investigation. Finally, Section 2.6 investigates the link between Additive Outliers and skewness and kurtosis, while Section 2.7

²In this respect, Hardle and Hafner (2000) compare several ARCH-type models in terms of option pricing on the German stock index DAX.

offers some concluding remarks.

2.2 Distribution choices

A univariate time series y_t ($t = 1, \dots, T$), known to be typically conditionally heteroscedastic, may be modelled as follows:

$$y_t = E[y_t | \Omega_{t-1}] + \varepsilon_t, \quad (2.1)$$

where ε_t is the disturbance term (or unpredictable part) and Ω_{t-1} is the information set at time $t - 1$. Without loss of generality, we can define an Autoregressive Conditional Heteroscedastic (ARCH) process, ε_t by:

$$\varepsilon_t = z_t \sigma_t \quad (2.2)$$

$$z_t \sim i.i.d. (0, 1) \quad (2.3)$$

$$\sigma_t = h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_1; \eta), \quad (2.4)$$

where z_t is an independently and identically distributed (*i.i.d.*) process with $E(z_t) = 0$, $Var(z_t) = 1$, η is a parameter vector and $h(\cdot)$ is a function giving the conditional standard deviation. By definition, ε_t is serially uncorrelated with mean zero, and its conditional variance equals σ_t^2 . To estimate this kind of model by maximum likelihood, one has to make an additional assumption on the innovation process by choosing a density function for z_t .

It is not our intention to review all the existing densities but only three of the most widespread in the literature dealing with financial time series (the normal, the Bernoulli-normal mixture and the Student) before presenting the skewed Student distribution.

2.2.1 The normal distribution

A common choice for the distribution of z_t is the normal $N(0, 1)$. The log-likelihood function of y_1, y_2, \dots, y_T is:

$$\begin{aligned} L_{norm} &= \sum_{t=1}^T [\ln g(\varepsilon_t \sigma_t^{-1}) - \ln \sigma_t] \\ &= -\frac{1}{2} \sum_{t=1}^T [\ln(2\pi) + \ln(\sigma_t^2) + z_t^2], \end{aligned} \quad (2.5)$$

where $g(\cdot)$ is in this case the Gaussian probability density function (pdf) and ε_t , σ_t^2 and z_t are given in Eq. (2.2)-(2.4). This normality assumption is to a certain extent justified by the fact that consistent estimates are found for parameters of the first two conditional moments (provided that they are correctly specified), even when normality does not hold. Note that ε_t and σ_t^2 are functions of past observations or in other terms are computed recursively. In a full ML framework, the values used to start up the recursion are considered as unknown quantities and estimated jointly with the other parameters. However, it is convenient to replace these values by their expected value or their sample mean.³ In this case we call this estimation procedure approximate ML. Note that both are equivalent asymptotically.

2.2.2 Mixture of normal distributions

Using four major daily exchange rates in dollar terms (GBP, DEM, FRF and YEN) over the period 1980-1996, Beine and Laurent (2000) find that an important number of outliers are responsible for the rejection of the normality assumption (more than 300 for the DEM). Following among others Jorion (1988) and Vlaar and Palm (1993), and in order to account for these outliers, they use a jump-diffusion ARCH-type model that assumes that the returns are drawn from a mixture of normal distributions, i.e. a diffusion process combined with an additive jump process. They show that for the DEM, the FRF and the GBP, this distribution is validated by the data and in all cases clearly outperforms the normal distribution.

The Bernoulli-normal mixture is defined as follows:

$$y_t = \bar{\mu}_t + \sigma_t z_t, \text{ with probability } (1 - \lambda_B) \quad (2.6)$$

$$y_t = \bar{\mu}_t + \sigma_t z_t + \mu_B + \sqrt{\sigma_B^2} z_t^*, \text{ with probability } \lambda_B, \quad (2.7)$$

where z_t and z_t^* are *i.i.d.* $N(0, 1)$, $E(z_t z_t^*) = 0$, λ_B stands for the probability of a jump and is drawn from a Bernoulli distribution ($0 < \lambda_B < 1$), μ_B is the mean of the jump distribution while σ_B^2 captures the variance of the jump distribution. $\bar{\mu}_t$ and σ_t^2 are the conditional mean and conditional variance of the diffusion process.⁴

³For instance, if ε_t has a MA(1) component, ε_0 is set to $E(\varepsilon_t) = 0$.

⁴This specification is similar to the one proposed by Neely (1999). However, this author considers a Bernoulli-Student distribution.

This model can be rewritten as:⁵

$$y_t = E(y_t|\Omega_{t-1}) + \varepsilon_t \quad (2.8)$$

$$\varepsilon_t \sim (1 - \lambda_B)N(-\lambda_B\mu_B, \sigma_t^2) + \lambda_B N(\mu_B - \lambda_B\mu_B, \sigma_t^2 + \sigma_B^2), \quad (2.9)$$

where $E(y_t|\Omega_{t-1}) = \bar{\mu}_t + \lambda_B\mu_B$, and $\lambda_B\mu_B$ is the conditional mean of the jump process. Notice that in this specification, λ_B is assumed to be constant over time.

The log-likelihood associated with this new distribution takes the following form (for a sample of size T):

$$\begin{aligned} L_{Bern} &= -\frac{T}{2} \ln(2\pi) \\ &+ \sum_{t=1}^T \ln \left\{ \frac{(1 - \lambda_B)}{\sigma_t^2} \exp \left[-\frac{(y_t - \bar{\mu}_t)^2}{2\sigma_t^2} \right] + \frac{\lambda_B}{\sqrt{\sigma_t^2 + \sigma_B^2}} \exp \left[-\frac{(y_t - \bar{\mu}_t - \mu_B)^2}{2(\sigma_t^2 + \sigma_B^2)} \right] \right\} \end{aligned} \quad (2.10)$$

It can be seen that σ_B^2 is the additional volatility related to the jump. It should be stressed that while the normal mixture distribution can account for skewness, its introduction will also affect the conditional fourth moment of the residuals. Indeed, Vlaar and Palm (1993) show that for a Bernoulli-normal mixture as described in Eq. (2.9), the skewness (or third moment of standardized variable) of ε is equal to:

$$\frac{(\lambda_B - \lambda_B^2)\mu_B \{(1 - 2\lambda_B)\mu_B^2 + 3\sigma_B^2\}}{\{(\lambda_B - \lambda_B^2)\mu_B^2 + E(\sigma_t^2) + \lambda_B\sigma_B^2\}^{3/2}}, \quad (2.11)$$

while its excess kurtosis (or fourth moment of standardized variable minus 3) equals:

$$\frac{3V(\sigma_t^2) + (\lambda_B - \lambda_B^2) \{3\sigma_B^4 + (6 - 12\lambda_B)\mu_B^2\sigma_B^2 + (1 - 6\lambda_B + 6\lambda_B^2)\mu_B^4\}}{\{(\lambda_B - \lambda_B^2)\mu_B^2 + E(\sigma_t^2) + \lambda_B\sigma_B^2\}^2}, \quad (2.12)$$

where $E(\sigma_t^2)$ and $V(\sigma_t^2)$ are respectively the unconditional expectation and unconditional variance of σ_t^2 (which can be estimated by their sample analog).

Note that when λ_B is different from 0 and 1, a negative (resp. positive) value of μ_B means that the innovations are negatively (resp. positively) skewed. However, the jump probability, the mean size and the variance of the jump govern the skewness and kurtosis of ε , which makes the interpretation of the parameters quite challenging.

⁵Vlaar and Palm (1993) show that under this mixture of normal distributions, $E(\varepsilon_t) = 0$. This is done by shifting the density by $\lambda_B\mu_B$. See these authors for more details.

2.2.3 The Student distribution

As reported by Palm (1996), Pagan (1996) and Bollerslev, Chou, and Kroner (1992), the use of the Student-t distribution is widespread in the literature. In particular, Bollerslev (1987), Hsieh (1989), Baillie and Bollerslev (1989) and Palm and Vlaar (1997) among others show that this distribution better captures the observed kurtosis.

As a reminder, provided that $v > 2$, z_t is distributed as a Student with mean 0, variance 1 and degree of freedom v , and denoted $z \sim ST(0, 1, v)$, if:

$$g(z|v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma\left(\frac{v}{2}\right)} \left[1 + \frac{z^2}{v-2}\right]^{-(v+1)/2}, \quad (2.13)$$

and $\Gamma(\cdot)$ is the Gamma function.

In this case, the log-likelihood function (for a sample size of T) becomes:

$$\begin{aligned} L_{Stud} &= T \left\{ \ln \Gamma\left(\frac{v+1}{2}\right) - \ln \Gamma\left(\frac{v}{2}\right) - \frac{1}{2} \ln [\pi(v-2)] \right\} \\ &\quad - 0.5 \sum_{t=1}^T \left[\ln \sigma_t^2 + (1+v) \ln \left(1 + \frac{z_t^2}{v-2}\right) \right], \end{aligned} \quad (2.14)$$

where ε_t , σ_t^2 and z_t are given in Eq. (2.2)-(2.4). Compared to the normal distribution, the Student- t implies the estimation of the additional parameter v standing for the number of degrees of freedom. The thickness of the tails is decreasing when v is increasing. The constraints on the tail parameter can be relaxed (after reparametrization) by allowing v to take values in $(0, 2]$. In these cases, the variance is infinite and σ_t^2 , which is not the variance anymore, remains a dispersion parameter.

2.2.4 Skewed densities

More recently, Fernández and Steel (1998) proposed an extension of the Student distribution by adding a skewness parameter. Their procedure allows the introduction of skewness in any continuous unimodal and symmetric (about 0) distribution $g(\cdot)$ by changing the scale at each side of the mode. To understand how to build this new family of densities, it is fruitful to express it in terms of a mixture of two truncated densities.

Construction

Let $u \in \mathfrak{R}$ be an *i.i.d.* continuous random variable with a symmetric unimodal density function $g(\cdot)$ with mean 0 and variance 1,

$$u \sim i.i.d. g(0, 1) \quad (2.15)$$

and x , a Bernoulli process, with probability of success $\frac{\xi^2}{1+\xi^2}$. Let us consider the following mixture:

$$\epsilon = x\xi |u| - (1-x) \frac{1}{\xi} |u|. \quad (2.16)$$

Using Eq. (2.15) and (2.16), the unconditional density $f(\epsilon|\xi)$ of ϵ is:

$$f(\epsilon|\xi) = Prob(x=1) f(\epsilon|\xi, x=1) + Prob(x=0) f(\epsilon|\xi, x=0) \quad (2.17)$$

where $Prob(x=1) = 1 - Prob(x=0) = \frac{\xi^2}{1+\xi^2}$.

Recalling that if $u \sim g(u)$, $|u| \sim 2g(u)I_{[0,\infty)}(u)$ and $\xi u \sim \frac{1}{\xi}g(u)$, one obtains:

$$\begin{aligned} f(\epsilon|\xi, x=1) &= f(\xi|u| |\xi) = 2\frac{1}{\xi}g\left(\frac{\epsilon}{\xi}\right)I_{[0,\infty)}(\epsilon) \\ f(\epsilon|\xi, x=0) &= f\left(-\frac{1}{\xi}|u| |\xi\right) = 2\xi g(\xi\epsilon)I_{(-\infty,0)}(\epsilon); \end{aligned}$$

Consequently, after straightforward simplifications, we have:

$$f(\epsilon|\xi) = \frac{2}{\xi + \frac{1}{\xi}} \left[g\left(\frac{\epsilon}{\xi}\right) I_{[0,\infty)}(\epsilon) + g(\xi\epsilon) I_{(-\infty,0)}(\epsilon) \right], \quad (2.18)$$

Thus, $f(\epsilon|\xi)$ is a unimodal density with the same mode as $g(\epsilon)$ and a skewness parameter $\xi > 0$ such that the ratio of probability masses above and below the mode is:

$$\frac{\Pr(\epsilon \geq 0|\xi)}{\Pr(\epsilon < 0|\xi)} = \xi^2. \quad (2.19)$$

Note that the density $f(\epsilon|1/\xi)$ is the ‘‘mirror’’ of $f(\epsilon|\xi)$ with respect to the mode. Therefore, working with $\ln(\xi)$ might be preferable to indicate the sign of the skewness. If we set $y_t = \mu_t + \epsilon_t\sigma_t$, where ϵ_t has a skewed Student distribution (as obtained by considering a Student distribution with mean 0 and variance 1 for u in Eq. (2.15)), then we obtain a distribution for y_t where all these parameters have a clear interpretation:

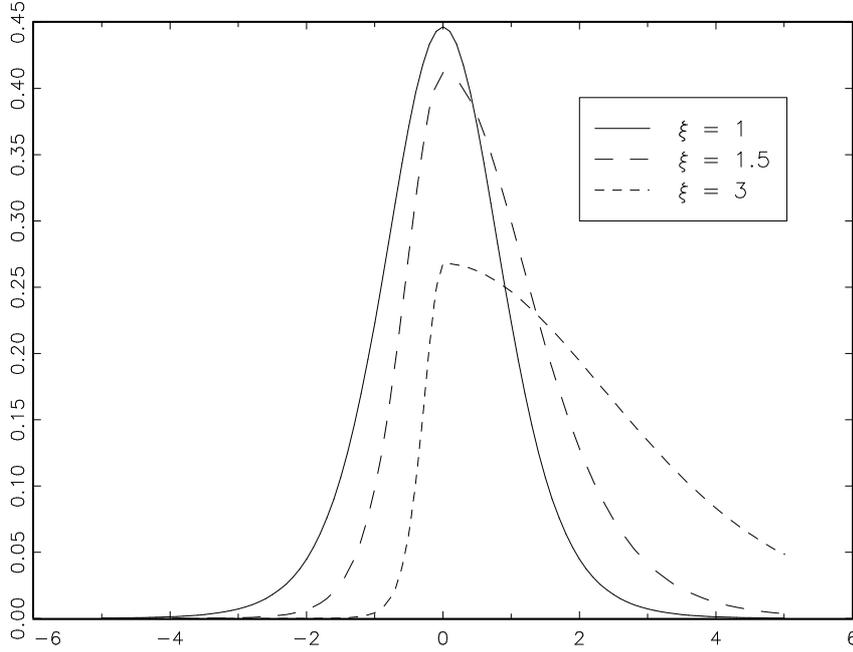


Figure 2.1: Skewed Student densities with $v = 8$ and $\xi = 1, 1.5$ and 3 .

- μ_t , as the conditional mode, models the location,
- $\sigma_t^2 > 0$ (which is not the conditional variance anymore) models the dispersion,
- $\xi > 0$ models the skewness,
- $v > 0$ models the tail thickness.

Note that the four important aspects of the distribution can thus be specified. This density has been used successfully by Lambert and Laurent (2000) on the daily exchange rate Deutsche mark US dollar and Von Rohr and Hoeschele (1999) in a (static) Bayesian framework. The skewed normal distribution is a limiting case ($v \rightarrow \infty$) of the skewed Student with the same tail properties as the traditional normal.

Note also that contrary to the skew-t of Jones and Faddy (2000), the skewed Student presented here above has the same thickness of tails at plus and minus infinity, where it behaves like $z^{-(v+1)}$.

Moments

Fernández and Steel (1998) show that if the r^{th} ($r \in \mathfrak{R}$) order moment of $g(\cdot)$ exists, the associated skewed distribution in Eq. (2.18) also has a finite r^{th} moment. In particular,

$$E(\epsilon^r|\xi) = M_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}} \quad (2.20)$$

where

$$M_r = \int_0^\infty 2s^r g(s) ds, \quad (2.21)$$

and M_r is the r^{th} order moment of $g(\cdot)$ truncated to the positive real values. Provided that these quantities are finite⁶, we can easily obtain:

$$E(\epsilon|\xi) = M_1 \left(\xi - \frac{1}{\xi} \right) \quad (2.22)$$

$$\begin{aligned} V(\epsilon|\xi) &= E(\epsilon^2|\xi) - E(\epsilon|\xi)^2 \\ &= (M_2 - M_1^2) \left(\xi^2 + \frac{1}{\xi^2} \right) + 2M_1^2 - M_2 \end{aligned} \quad (2.23)$$

$$Sk(\epsilon|\xi) = \frac{E(\epsilon^3|\xi) - 3E(\epsilon|\xi)E(\epsilon^2|\xi) + 2E(\epsilon|\xi)^3}{Var(\epsilon|\xi)^{\frac{3}{2}}} \quad (2.24)$$

$$\begin{aligned} &= \frac{\left(\xi - \frac{1}{\xi} \right) (M_3 + 2M_1^3 - 3M_1M_2) \left(\xi^2 + \frac{1}{\xi^2} \right) + 3M_1M_2 - 4M_1^3}{Var(\epsilon|\xi)^{\frac{3}{2}}} \\ Ku(\epsilon|\xi) &= \frac{E(\epsilon^4|\xi) - 4E(\epsilon|\xi)E(\epsilon^3|\xi) + 6E(\epsilon^2|\xi)E(\epsilon|\xi)^2 - 3E(\epsilon|\xi)^4}{Var(\epsilon|\xi)^2}, \end{aligned} \quad (2.25)$$

where $E(\cdot|\xi)$, $V(\cdot|\xi)$, $Sk(\cdot|\xi)$ and $Ku(\cdot|\xi)$ are respectively the mean, variance, skewness and kurtosis⁷, given ξ .

Let us now reconsider the skewed Student density of Fernández and Steel (1998), where $g(\cdot)$ in Eq. (2.18) is the Student distribution. As shown in Eq. (2.24) and (2.25), both ξ and v determine the skewness and the kurtosis. Figures 2.2 and 2.3 investigate the relation between these two parameters and the skewness (with $v > 3$ to ensure the existence of the skewness). For simplicity, we do not

⁶For instance, if $g(\cdot)$ is the Student density given in Eq. (2.13), the r^{th} order moment of ϵ exists if $v > r$.

⁷Even if a closed form of the kurtosis is theoretically available, it is not tractable.

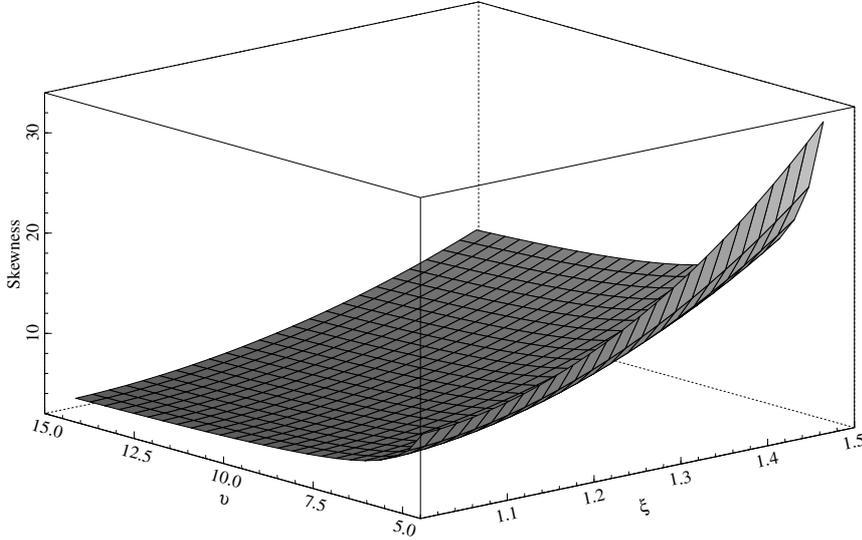


Figure 2.2: Skewness implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $3.5 \leq v \leq 15$

tackle the case $0 < \xi \leq 1$ and only report the graphs when $\xi \geq 1$. It is clear from these two figures that the dominating feature of the skewness is the ξ parameter.⁸ From Figure 2.3 we can see that skewness may be very high when v approaches 3. Figures 2.4 and 2.5 traces the kurtosis surface for several combinations of $\xi > 1$ and $v > 4$ (to insure the existence of the kurtosis). The dominating feature of the kurtosis is obviously the v parameter, even if the higher the asymmetry parameter, the higher the kurtosis. Consequently, even if both ξ and v determine skewness and kurtosis, Figures 2.2-2.5 show that skewness (resp. kurtosis) is mainly governed by ξ (resp. v).

Standardized skewed Student density

One drawback of this parameterization of the skewed Student density is that μ_t and σ_t^2 are not the conditional mean and the conditional variance but the conditional mode and some measure of conditional dispersion. In order to keep in the ARCH tradition, it is important to express the density in terms of the mean and of the variance, and, thus, to reparameterize Eq. (2.2). In such a way it will be possible

⁸This property applies also for Hansen's skewed Student density. See Jondeau and Rockinger (2000).

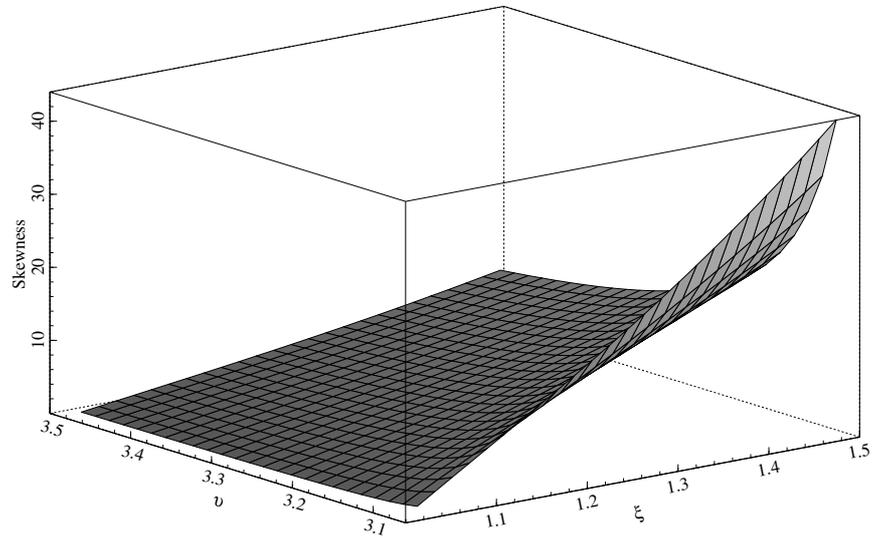


Figure 2.3: Skewness implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $3.05 \leq v \leq 3.5$

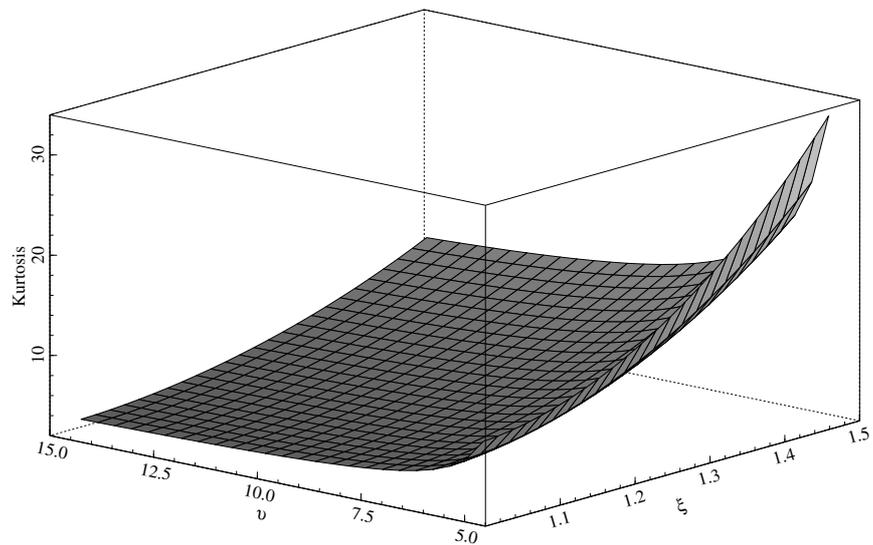


Figure 2.4: Kurtosis implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $4.5 \leq v \leq 15$

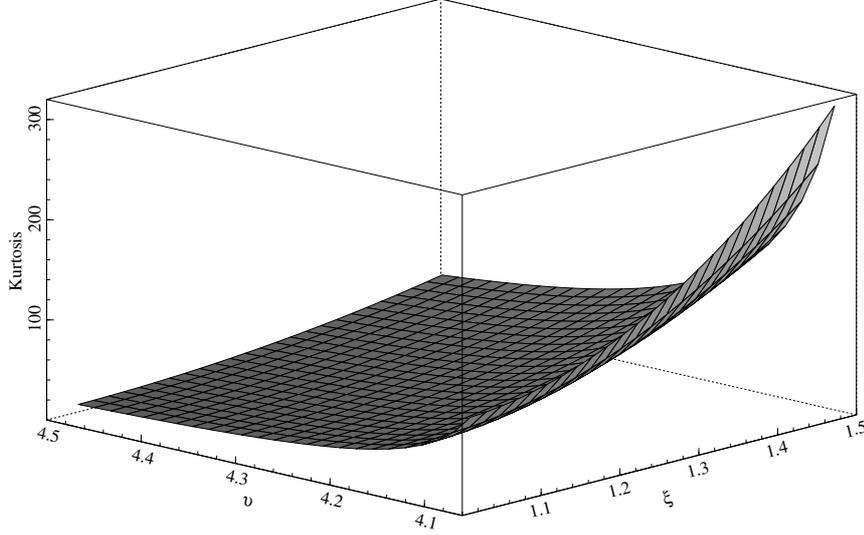


Figure 2.5: Kurtosis implied by the skewed Student density for several combinations of $1 \leq \xi \leq 1.5$ and $4.05 \leq v \leq 4.5$

to take z_t in Eq. (2.3) a skewed Student distribution with zero mean and unit variance. More specifically, assume that ϵ has a Student distribution with density $g(\cdot)$ and degree of freedom v . Then, the r^{th} moment of ϵ truncated to the positive real values is:

$$M_{r|v} = \frac{\Gamma\left(\frac{v-r}{2}\right) \Gamma\left(\frac{1+r}{2}\right) (v-2)^{\frac{1+r}{2}}}{\sqrt{\pi} (v-2) \Gamma\left(\frac{v}{2}\right)} \quad (2.26)$$

Using Eq. (2.22) and (2.23), and provided that $v > 2$, it follows that:

$$E(\epsilon|\xi, v) = \frac{\Gamma\left(\frac{v-1}{2}\right) \sqrt{v-2}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right)} \left(\xi - \frac{1}{\xi}\right) \equiv m, \quad (2.27)$$

and

$$V(\epsilon|\xi, v) = \left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2 \equiv s^2. \quad (2.28)$$

Now consider the standardized random variable

$$z_t = \frac{\epsilon_t - m}{s}. \quad (2.29)$$

Definition 1 If (i) z_t is defined by Eq. (2.29) and (ii) ϵ_t has a density given by Eq. (2.18), where $g(\cdot)$ is the Student density given by Eq. (2.13), then z_t has mean 0, variance 1 and is said to be distributed as standardized skewed Student

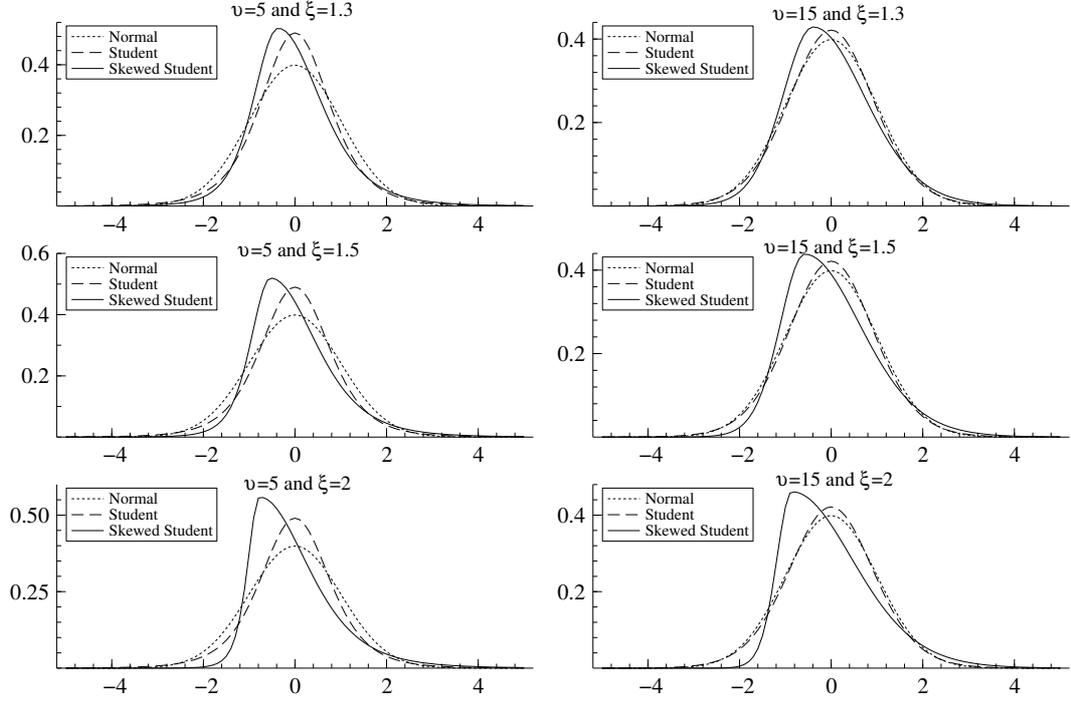


Figure 2.6: Normal, Student and skewed Student densities.

with asymmetry parameter ξ , and number of degrees of freedom $\nu (> 2)$. This is denoted $z_t \sim SKST(0, 1, \xi, \nu)$. The density of z_t is given by:

$$f(z_t|\xi, \nu) = \frac{2}{\xi + \frac{1}{\xi}} s \left\{ g \left[\xi (sz_t + m) | \nu \right] I_{(-\infty, 0)}(z_t + m/s) + g \left[(sz_t + m) / \xi | \nu \right] I_{[0, \infty)}(z_t + m/s) \right\}. \quad (2.30)$$

For a standardized skewed Student, the log-likelihood of y_1, y_2, \dots, y_T is:

$$\begin{aligned} L_{SkSt} = & T \left\{ \ln \Gamma \left(\frac{\nu + 1}{2} \right) - \ln \Gamma \left(\frac{\nu}{2} \right) - 0.5 \ln [\pi (\nu - 2)] + \ln \left(\frac{2}{\xi + \frac{1}{\xi}} \right) + \ln (s) \right\} \\ & - 0.5 \sum_{t=1}^T \left\{ \ln \sigma_t^2 + (1 + \nu) \ln \left[1 + \frac{(sz_t + m)^2}{\nu - 2} \xi^{-2I_t} \right] \right\} \end{aligned} \quad (2.31)$$

where $I_t = \begin{cases} 1 & \text{if } z_t \geq -\frac{m}{s} \\ -1 & \text{if } z_t < -\frac{m}{s} \end{cases}$. Figure 2.6 displays several standardized skewed Student densities with $\nu = 5, 15, +\infty$ and $\xi = 1, 1.3, 1.5$ and 2.

Distribution and quantile functions of a skewed distribution

Using the same notation and hypotheses as in the previous section, we can relate the cumulative distribution function (cdf) F and the quantile function F^{-1} corresponding to a standardized skewed density $f(z|\xi)$ to the cdf G and quantile function G^{-1} of the original symmetric density. We have

$$F(z|\xi) = \begin{cases} \frac{2}{1+\xi^2} \frac{1}{s} G(\xi(sz + m)) & \text{if } z < -\frac{m}{s} \\ 1 - \frac{2}{1+\xi^{-2}} \frac{1}{s} G(-\xi^{-1}(sz + m)) & \text{if } z \geq -\frac{m}{s} \end{cases} \quad (2.32)$$

for the cdf and

$$F^{-1}(p|\xi) = \begin{cases} \frac{1}{\xi} G^{-1}\left(\frac{p}{2}(1+\xi^2)\right) - m & \text{if } p < \frac{1}{1+\xi^2} \\ -\frac{\xi G^{-1}\left(\frac{1-p}{2}(1+\xi^{-2})\right) - m}{s} & \text{if } p \geq \frac{1}{1+\xi^2} \end{cases} \quad (2.33)$$

for the quantile function. These two functions are particularly interesting in Monte Carlo simulations to generate random numbers from our family of skewed densities, in Value-at-Risk applications and to check the adequacy of the conditional distribution (in the density forecast evaluation method, see Section 2.4).

2.3 GARCH model with skewed distribution for the innovations

Before analyzing real data, and in order to assess the practical applicability of the QML procedure of the skewed Student distribution, we first present the results of a simulation study. It is not our intention to provide a comprehensive Monte Carlo study of the QML. The reliability of the inference concerning the model parameters will not be examined. Our results, however, will provide some preliminary evidence with respect to the finite sample properties of the QML estimator for a skewed Student pseudo-likelihood (coupled with a GARCH model).

Note that, *since any particular probability model is unlikely to be the “correct” model, but should accurately be viewed as an approximation to the underlying probability structure, it is reasonable to report “robust” standard errors, as suggested by White (1982)... These give asymptotically valid confidence intervals for the “pseudo-true” parameter values which minimize the information distance between the true probability measure and the quasi-likelihood.* (Hansen, 1994, p. 713).

The exact asymptotic standard errors for the QML estimator $\hat{\Phi}$ are the square roots of the diagonal elements of the matrix:

$$\hat{A}_T^{-1} \hat{B}_T \hat{A}_T^{-1}, \quad (2.34)$$

where

$$\hat{A}_T = - \sum_{t=1}^T \left(\frac{\partial^2 l_t(\hat{\Phi})}{\partial \Phi \partial \Phi'} \right) \quad (2.35)$$

$$\hat{B}_T = \sum_{t=1}^T \left(\frac{\partial l_t(\hat{\Phi})}{\partial \Phi} \frac{\partial l_t(\hat{\Phi})}{\partial \Phi'} \right), \quad (2.36)$$

and $l_t(\hat{\Phi})$ is the log-likelihood of observation t , evaluated at $\hat{\Phi}$. These standard errors are robust for deviations from the distribution used in the objective function.

More specifically, consider (as DGP) the following GARCH(1,1) model:

$$y_t = \mu + \varepsilon_t \quad (2.37)$$

$$\varepsilon_t = \sigma_t z_t \quad (2.38)$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (2.39)$$

To illustrate the behavior of QML estimators, Table 2.1 reports average estimated parameters (over the Monte Carlo replications) and average robust standard errors corresponding to the model defined in Eq. (2.37) to (2.39) and estimated under three different pseudo-likelihoods: the normal, i.e. $z_t \sim N(0, 1)$, Student, i.e. $z_t \sim ST(0, 1, \nu)$ and Skewed Student, i.e. $z_t \sim SKST(0, 1, \xi, \nu)$. In this first experiment, the DGP is $\mu = 0, \omega = 0.1, \alpha_1 = 0.1, \beta_1 = 0.8$ and $z_t \sim SKST(0, 1, \exp(0.3), 8.0)$. The sample size (T) is 3000 observations⁹ and the number of replications equals 500.

From Table 2.1, it is clear that the QML method for the GARCH model, under the correct density (i.e. the skewed Student, see column 5), works reasonably well for the considered sample size. This table also illustrates the well known result of Weiss (1986) and Bollerslev and Wooldridge (1992) that (if the mean and the variance are specified correctly) the QML estimator under “pseudo-true” normal and Student distributions are respectively consistent (but inefficient) and

⁹To avoid start-up problem, the first 3000 realizations (out of 6000) were discarded for each replication.

inconsistent when the innovations are skewed.¹⁰ Moreover, the Monte Carlo results suggest that the QML estimator of the skewed Student (with a GARCH model governing the conditional variance) are only slightly biased and are more efficient than the Gaussian QML (when the true density is a skewed Student).

In Tables 2.2 and 2.3, the DGP is the same except that now, $z_t \sim \chi^2(3)$ and $z_t \sim \Gamma(1, 2)$ (respectively), where χ^2 and $\Gamma(\cdot)$ denote the Chi-square and Gamma distributions.¹¹ These two tables show that the skewed Student does a good job in modelling the first and second moments when the errors are Chi-square or Gamma distributed (that are both skewed and kurtosed) leading to very small biases in the mean and variance parameters, compared to the usual Student density.

2.4 Density forecasts

As explained above, relying on the Gaussian assumption has several advantages if we are only interested in the first two conditional moments. Furthermore, switching from a normal density to a Student density may be hazardous if this last assumption does not hold. And if it does not, it is very likely that the estimates will not be consistent (Newey and Steigerwald, 1997). As a consequence, we have to be very cautious with the choice of the density and check its appropriateness. To compare the adequacy of the different distributions, we employ in- and out-of-sample density forecasts proposed by Diebold, Gunther, and Tay (1998) (henceforward DGT).¹² The idea of density forecasts is quite simple. Let $f_i(y_i|\Omega_i)_{i=1}^m$ be a sequence of m one-step-ahead density forecasts produced by a given model, where Ω_i is the conditioning information set, and $p_i(y_i|\Omega_i)_{i=1}^m$ the sequence of densities defining the data generating process y_i (which is never observed). Testing whether this density is a good approximation of the true density $p(\cdot)$ is equivalent to testing:

$$H_0 : f_i(y_i|\Omega_i)_{i=1}^m = p_i(y_i|\Omega_i)_{i=1}^m. \quad (2.40)$$

DGT use the fact that under (2.40), the probability integral transform $\hat{\zeta}_i = \int_{-\infty}^{y_i} f_i(t)dt$ is *i.i.d.* $U(0, 1)$, i.e. independent and identically distributed uniform,

¹⁰The last two lines of Tables 2.1 to 2.3 report the skewness and kurtosis of z_t implied by the DGP and the estimated parameters of the three pseudo-likelihoods.

¹¹The Chi-square and Gamma distributions have been standardized in order to have mean 0 and variance 1.

¹²For more details about density forecasts and applications in finance, see the special issue of *Journal of Forecasting* Timmermann (2000).

Table 2.1: Monte Carlo analysis: skewed Student errors.

	DGP	normal	Student	skewed Student
μ	0.0	-0.0008 (0.0133)	-0.0531 (0.0136)	0.0009 (0.0131)
ω	0.1	0.1067 (0.0237)	0.1022 (0.0192)	0.1031 (0.0181)
α_1	0.1	0.1013 (0.0169)	0.0974 (0.0136)	0.1008 (0.0134)
β_1	0.8	0.7914 (0.0347)	0.7994 (0.0280)	0.7963 (0.0264)
$\ln(\xi)$	0.3	-	-	0.3008 (0.0209)
ν	8.0	-	7.1649 (0.7096)	8.2000 (0.8794)
Skewness	0.76	0	0	0.75
Kurtosis	5.13	3	5.44	5.03

Model: $y_t = \mu + z_t (\omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)^{1/2}$. DGP: $\mu = 0$, $\omega = 0.1$, $\alpha_1 = 0.1$, $\beta_1 = 0.8$ and $z_t \sim SKST(0, 1, \exp(0.3), 8.0)$. Robust standard errors of the estimated parameters are reported in parentheses. The last two lines report the skewness and kurtosis of z_t implied by the DGP and the estimated parameters of the three pseudo-likelihoods.

Table 2.2: Monte Carlo analysis: Chi-square errors.

	DGP	normal	Student	skewed Student
μ	0.0	0.0001 (0.0133)	-0.1779 (0.0161)	-0.0021 (0.0122)
ω	0.1	0.1049 (0.0250)	0.1035 (0.0182)	0.1023 (0.0067)
α_1	0.1	0.1027 (0.0197)	0.0968 (0.0134)	0.1018 (0.0055)
β_1	0.8	0.7919 (0.0376)	0.8095 (0.0247)	0.7990 (0.0079)
$\ln(\xi)$	-	-	-	1.8591 (0.1344)
v	-	-	3.8108 (0.2400)	7.1504 (0.6885)
Skewness	1.63	0	0	1.75
Kurtosis	7	3	11.40	9.02

Note: see Table 2.1, except that the DGP has $z_t \sim \chi^2(3)$.

Table 2.3: Monte Carlo analysis: Gamma errors.

	DGP	normal	Student	skewed Student
μ	0.0	-0.0009 (0.0133)	-0.1407 (0.0160)	-0.0056 (0.0128)
ω	0.1	0.1048 (0.0246)	0.1006 (0.0181)	0.1004 (0.0092)
α_1	0.1	0.1011 (0.0186)	0.0938 (0.0130)	0.1003 (0.0072)
β_1	0.8	0.7940 (0.0365)	0.8080 (0.0258)	0.7997 (0.0129)
$\ln(\xi)$	-	-	-	1.2207 (0.0589)
v	-	-	4.7417 (0.3487)	11.1242 (1.6085)
Skewness	1.41	0	0	1.33
Kurtosis	6	3	7.44	5.74

Note: see Table 2.1, except that the DGP has $z_t \sim \Gamma(1, 2)$.

where $\int_{-\infty}^{y_i} f_i(t)dt$ is the cumulative density function associated to $f_i(y_i|\Omega_i)$. To check H_0 , they propose to use goodness-of-fit and independence tests. The *i.i.d.*-ness property of $\hat{\zeta}_i$ can be evaluated by plotting the correlograms of $(\hat{\zeta} - \bar{\zeta})^j$, for $j = 1, 2, 3, 4, \dots$, to detect potential dependence in the conditional mean, variance, skewness, kurtosis, etc. Departure from uniformity can also be evaluated by plotting an histogram of $\hat{\zeta}_i$. According to Bauwens, Giot, Grammig, and Veredas (2000) p. 4 (in the context of duration models), *a humped shape of the $\hat{\zeta}$ -histogram would indicate that the issued forecasts are too narrow and that the tails of the true density are not accounted for. On the other hand, a U-shape of the histogram would suggest that the model issues forecasts that either under- or overestimate too frequently.*¹³ To illustrate the usefulness of this testing procedure, Figures 2.7 to 2.9 plot the $\hat{\zeta}$ -histograms (with 40 cells) of 5000 in-sample one-step-ahead forecasts based on the same DGP as in the previous section. In Figure 2.7, $ST(0, 1, 8)$ errors are generated while a Gaussian QML estimation is performed. In Figures 2.8 and 2.9 skewed Student $SKST(0, 1, \exp(0.3), 8)$ errors are generated while Student and skewed Student pseudo-likelihoods are used respectively in the QML procedure. Figures 2.7 and 2.8 clearly suggest that the assumption made on the error term is not appropriate. Moreover, Figure 2.8 shows that an *inverted S* shape of the histogram would indicate that the errors are skewed, i.e. the true density is probably not symmetric. However, from Figure 2.9, it is clear that the probability integral transform is uniformly distributed. To check the uniformity of ξ , we could also rely on the Pearson goodness-of-fit test that compares the empirical distribution with the theoretical one (see the application).

For a given number of cells denoted g , the Pearson goodness-of-fit statistics is:

$$P(g) = \sum_{i=1}^g \frac{(n_i - En_i)^2}{En_i}, \quad (2.41)$$

where n_i is the number of observations in cell i and En_i is the expected number of observations (based on the ML estimates). For *i.i.d.* observations, Palm and Vlaar (1997) show that under the null of a correct distribution the asymptotic distribution of $P(g)$ is bounded between a $\chi^2(g - 1)$ and a $\chi^2(g - k - 1)$ where k is the number of estimated parameters.

¹³Confidence intervals for the $\hat{\zeta}$ -histogram can be obtained by using the properties of the histogram under the null hypothesis of uniformity.

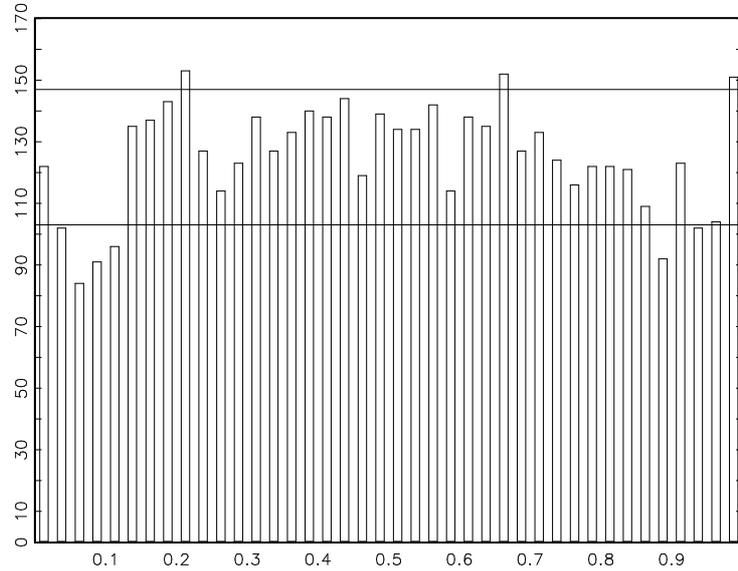


Figure 2.7: $\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $ST(0, 1, 8)$ errors. The MLEs were computed assuming normality for the innovations.

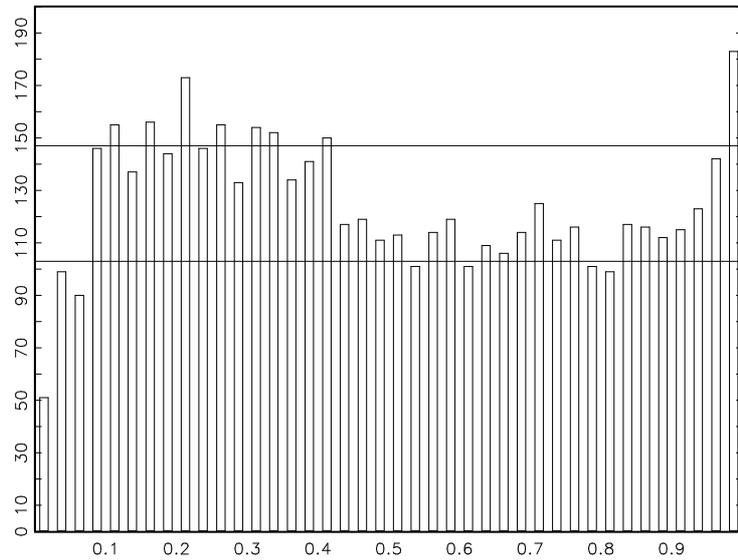


Figure 2.8: $\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $SKST(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming the innovations to be Student distributed.

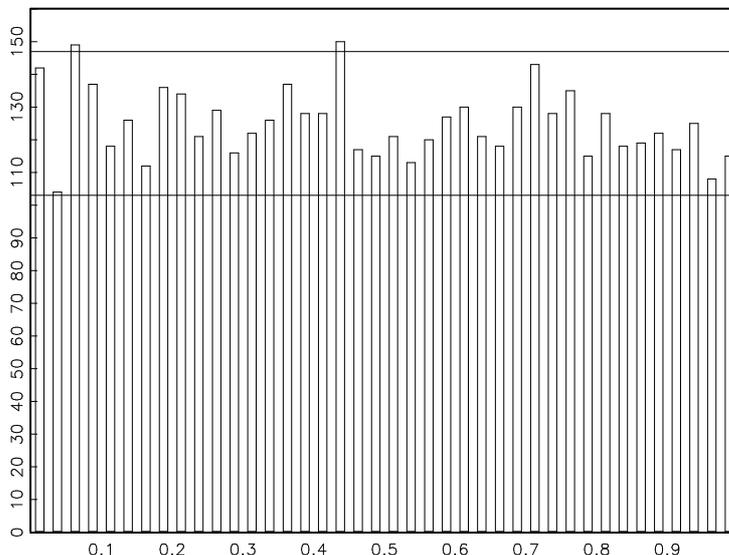


Figure 2.9: $\hat{\zeta}$ -histogram (40 cells) for 5000 one-step-ahead forecasts. DGP with $SKST(0, 1, \exp(0.3), 8)$ errors. The MLEs were computed assuming that the innovations are skewed Student distributed

2.5 Application

The analyzed database consists of 3000 observations of the NASDAQ from January 1985 until December 1996 (source: Datastream). The daily return is defined as $y_t = 100 \times (\log p_t - \log p_{t-1})$ where p_t is the stock index value of day t . Here, we propose to analyze the NASDAQ by relying on four pseudo-likelihoods: a Gaussian, a Bernoulli-normal mixture, a Student and a skewed Student. Dynamics will be introduced in the conditional mean and the conditional variance with an AR(1)-APARCH(1,1) specification:

$$y_t = \mu + \psi_1(y_{t-1} - \mu) + \varepsilon_t \quad (2.42)$$

$$\varepsilon_t = \sigma_t z_t \quad (2.43)$$

$$\sigma_t^\delta = \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta, \quad (2.44)$$

where $\mu, \psi_1, \omega, \alpha_1, \beta_1, \gamma$ and δ are parameters to be estimated.¹⁴ The APARCH is probably one of the most general ARCH-type model. Indeed, it nests at least seven GARCH models, see Ding, Granger, and Engle (1993). δ ($\delta > 0$) plays the role of a

¹⁴It is convenient to start the recursion of Eq. (2.44) by setting unobserved components to their sample average. This point will be clarified in Section 3.6.

Box-Cox transformation of σ_t , while γ ($-1 < \gamma < 1$), reflects the so-called leverage effect¹⁵ (Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990). Following Ding, Granger, and Engle (1993), if it exists, a stationary solution of (2.44) is given by:

$$E(\sigma_t^\delta) = \frac{\omega}{1 - \alpha_1 E(|z| - \gamma z)^\delta - \beta_1} \quad (2.45)$$

which depends on the density of z . Such a solution exists if $\omega > 0$ and $\alpha_1 E(|z| - \gamma z)^\delta + \beta_1 < 1$. Setting $\gamma = 0$ and $\delta = 2$ and assuming that z_t has zero mean and unit variance, one recovers the stationarity condition of the GARCH(1,1) model ($\alpha_1 + \beta_1 < 1$). Ding, Granger, and Engle (1993) derived the expression $E(|z| - \gamma z)^\delta$ in the Gaussian case. Paoletta (1997) gives expressions for various non standardized densities. Lambert and Laurent (2001) show that for the standardized skewed Student:¹⁶

$$E(|z| - \gamma z)^\delta = \left[\xi^{-(1+\delta)} (1 + \gamma)^\delta + \xi^{1+\delta} (1 - \gamma)^\delta \right] \frac{\Gamma\left(\frac{\delta+1}{2}\right) \Gamma\left(\frac{\nu-\delta}{2}\right) (\nu-2)^{\frac{1+\delta}{2}}}{\left(\xi + \frac{1}{\xi}\right) \sqrt{(\nu-2)} \pi \Gamma\left(\frac{\nu}{2}\right)} \quad (2.46)$$

Note that a closed form solution of this expression is still not available in the literature for the mixture of normal distributions case. Consequently, we replace $E(|z| - \gamma z)^\delta$ by its sample counterpart, i.e. $\frac{1}{T} \sum_{i=1}^T (|z_t| - \gamma z_t)^\delta$.

Table 2.4 hereafter presents the QML estimation results (for the first 2000 observations) of the AR(1)-APARCH(1,1) for the four pseudo-likelihoods.¹⁷ The results have been obtain with G@RCH 2.0, an Ox package with a friendly dialog-oriented interface dedicated to the estimation and forecast of various univariate ARCH-type models (this software is documented in Appendix A). Table 2.5 reports some statistics of interest.

Note that caution is necessary in interpreting conventional confidence intervals for these estimates, since although the samples are large, the asymptotic properties of the estimates are not yet well established for the APARCH model. Indeed,

¹⁵A positive (resp. negative) value of γ means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive shocks.

¹⁶Notice that setting $\xi = 1$ leads to the stationarity condition of the symmetric Student density (with unit variance).

¹⁷The formula used to compute the robust standard errors is given in Eq. (2.34).

as explained in Section 2, sufficient conditions for consistency and asymptotic normality of estimators are only available for a limited class of processes, mainly GARCH (1,1) and ARCH (p) and for a Gaussian (pseudo-)likelihood. To the best of our knowledge, nothing is known about the consistency of the APARCH model of Ding, Granger, and Engle (1993) even if this specification offers several advantages compared to the simple GARCH model. Subject to these caveats on valid inference, let us comment the results:

1. First, the extra flexibility of the APARCH specification is required. Both the asymmetry coefficient (γ) and the power (δ) estimates suggest that a usual GARCH model is not appropriate to model the NASDAQ. This is also confirmed by likelihood ratio tests (not reported here to save space).
2. Second, comparing the log-likelihood and the AIC criterion of the four distributions, one should certainly retain the skewed Student density. Indeed, despite the fact that the LR test is presumably non-standard when comparing the normal and the Student and the Bernoulli-mixture with the non-Gaussian densities, the differences are so big that there is little doubt that the skewed Student should be preferred.
3. The estimated parameters attest that the distribution of the NASDAQ is highly kurtosed and skewed. Indeed, the number of degrees of freedom of the Student is about 6, which means that the innovations are fat-tailed. On the other hand, the estimated parameter μ_B of the mixture distribution is negative and significant, which suggests that the innovations are negatively skewed as shown in Eq. (2.11). However, the jump probability (λ_B) and the size of the jump (σ_B^2) are not significant at the 5 % level, which renders the interpretation of the results difficult because under the assumption that $\lambda_B = 0$ or $\lambda_B = 1$, μ_B and σ_B^2 are not identified. The asymmetry of the innovations is reinforced by the skewed Student density, whose $\ln(\xi)$ parameter is negative and significant. It seems moreover that the asymmetry feature of the APARCH model (characterizing the conditional variance) and the “skewness” coefficient of the skewed Student density are both necessary to explain the overall asymmetry of the series. To illustrate the difference between the normal, the Student and the skewed Student, Figure 2.10 plots the

fitted densities of the innovations, namely a $SKST(0, 1, \exp(-0.179), 6.039)$ (solid line), a $ST(0, 1, 5.57)$ (dashed line) and a $N(0, 1)$ (short dashes). The asymmetry coefficient equals -0.179 (and is significant), which means the skewed Student density allocates nearly 59% of the mass to the left side of the mode.

4. The stationary condition of the APARCH model is satisfied for all the distributions, as $\alpha_1 E(|z| - \gamma z)^\delta + \beta_1 < 1$ (at the QML estimators).¹⁸
5. The AR(1)-APARCH(1,1) seems to be adequate in describing the dynamics of the first two moments of the NASDAQ, for the period of interest. Indeed, the Box-Pierce statistics¹⁹ Q_{20} and Q_{20}^2 are all non significant at any reasonable level.²⁰
6. The relevance of the skewed Student distribution is also confirmed by the Pearson goodness-of-fit statistic, $P(50)$ and $P(60)$. While the normal and the Student distributions are rejected (the p-values equal about 0), the skewed Student density seems to be supported by the data (both by the non adjusted and adjusted tests with 50 and 60 cells). The results concerning the normal-mixture are more ambiguous since the acceptance of this density is very sensitive to the significance level (5 of 10%) and the version of the test (adjusted or not).

Finally, to assess the relevance of the skewed Student density, we perform some out-of-sample forecasts. Table 2.6 gives the goodness-of-fit tests (density forecasts test) on the one-day-ahead forecasts of the AR(1)-APARCH(1,1). This test has

¹⁸This is not a formal test because when computing $\alpha_1 E(|z| - \gamma z)^\delta + \beta_1$, we substitute QML estimates for the true parameters while in fact these parameters are estimated and thus subject to an uncertainty. However, accounting for this uncertainty to compute a confidence band is not trivial due to the non-linearity of this formula (see Eq. (2.46) for the skewed Student case).

¹⁹The number of degrees of freedom of the asymptotic distribution of the Box-Pierce test has to be adjusted by the number of ARMA parameters (to test the presence of serial correlation in the standardized residuals) while the Box-Pierce statistics on the squared standardized residuals has to be adjusted by the number of GARCH parameters (Bollerslev and Mikkelsen, 1996).

²⁰This test is computed on standardized residuals (\hat{z}_t) except for the Bernoulli-normal mixture. Indeed, for this model Q_{20} and Q_{20}^2 are computed on the normalized residuals that are obtained by re-expressing Eq. (2.9) to have $N(0, 1)$ innovations (if the mixture of normal assumption holds), i.e. $z_t^n = F^{-1} \left[(1 - \lambda_B) F \left(\frac{y_t - \bar{\mu}_t}{\sigma_t} \right) + \lambda_B F \left(\frac{y_t - \bar{\mu}_t - \mu_B}{\sigma_t + \sigma_B} \right) \right]$, where $F(\cdot)$ and $F^{-1}[\cdot]$ are respectively the cumulative distribution function and the quantile function of the standard normal density.

Table 2.4: AR(1) - APARCH (1, 1) model. Estimation results.

	normal	Bernoulli-normal	Student	skewed Student
μ	0.0554 (0.0206)	0.1082 (0.0258)	0.1010 (0.0195)	0.0626 (0.0197)
ψ_1	0.2946 (0.0288)	0.2969 (0.0252)	0.2861 (0.0298)	0.2787 (0.0239)
ω	0.0533 (0.0231)	0.0332 (0.0128)	0.0374 (0.0134)	0.0371 (0.0132)
α_1	0.1641 (0.0377)	0.1364 (0.0246)	0.1451 (0.0255)	0.1503 (0.0261)
γ	0.3768 (0.1124)	0.3115 (0.1209)	0.2787 (0.1064)	0.2419 (0.0953)
β_1	0.8024 (0.0537)	0.8145 (0.0394)	0.8258 (0.0368)	0.8229 (0.0360)
δ	1.0544 (0.3794)	1.3340 (0.3286)	1.2914 (0.2929)	1.3322 (0.2967)
$\ln(\xi)$	0	-	0	-0.1789 (0.0319)
v	∞	-	5.5706 (0.7490)	6.0388 (0.8182)
λ_B	-	0.0366 (0.0264)	-	-
μ_B	-	-1.1363 (0.5070)	-	-
σ_B^2	-	1.6606 (0.8141)	-	-
<i>Log-Lik</i>	-2090.6	-2014.6	-2008.7	-1993.1

Robust standard errors are given in parentheses. *Log-Lik* refers to the log-likelihood value at maximum.

Table 2.5: Statistics of interest.

	normal	Bernoulli-normal	Student	skewed Student
$\alpha_1 E(z - \gamma)^\delta + \beta_1$	0.934	0.944	0.940	0.967
Q_{20}	25.009	24.566	23.926	26.435
Q_{20}^2	10.185	23.985	11.625	10.804
$P(50)$	122.120	63.356	78.176	50.648
	(0.000)	(0.081)	(0.005)	(0.408)
	[0.000]	[0.008]	[0.000]	[0.120]
$P(60)$	110.828	78.668	85.183	52.210
	(0.000)	(0.044)	(0.014)	(0.722)
	[0.001]	[0.004]	[0.001]	[0.388]
AIC	2.099	2.024	2.019	2.004

Q_{20} and Q_{20}^2 are respectively the Box-Pierce statistics at lag 20 of the standardized and squared standardized residuals except for the Bernoulli-normal mixture for which it is computed on normalized and squared normalized residuals. $P(50)$ and $P(60)$ are the Pearson goodness-of-fit statistics based on 50 and 60 cells respectively. P-values of the non-adjusted and adjusted test are given respectively in parentheses and in brackets.

been conducted on the last 1000 observations (about 4 years), using the estimated parameters reported in Table 2.4. From Table 2.6, it is obvious that the normal, the Bernoulli-mixture and the Student pdf's are not adequate for density forecast purposes. On the other hand, the skewed Student passes this test (with less evidence for the adjusted version with 50 cells).

2.6 Asymmetry, fat-tails and Additive Outliers

The preceding estimation results from the ARMA-APARCH model suggest that the normal, mixture of normal and the Student distributions are not appropriate for modelling the NASDAQ. Indeed, goodness-of-fit tests fail to validate these distributions while the skewed Student seems to be appropriate for the period under investigation. To a certain extent, these results are not surprising given the very long sample period. This period includes many important events that are thought to have disrupted the smooth dynamics of the NASDAQ and lead to an important number of "Level and Volatility Outliers" (see Hotta and Tsay, 1998).

In addition to the previous study, it may also be interesting to attempt to

Table 2.6: Density forecast tests for the out-of-sample forecasts.

	normal	Bernoulli-normal	Student	skewed Student
$P(50)$	122.120 (0.006) [0.000]	76.425 (0.007) [0.000]	78.100 (0.005) [0.001]	57.600 (0.187) [0.035]
$P(60)$	117.915 (0.027) [0.001]	90.960 (0.002) [0.001]	88.160 (0.008) [0.005]	56.120 (0.582) [0.256]

Density forecast test - Pearson goodness-of-fit test. P-values of the non-adjusted and adjusted test are given respectively in parentheses and in brackets. These tests have been conducted on the last 1000 observations (about 4 years).

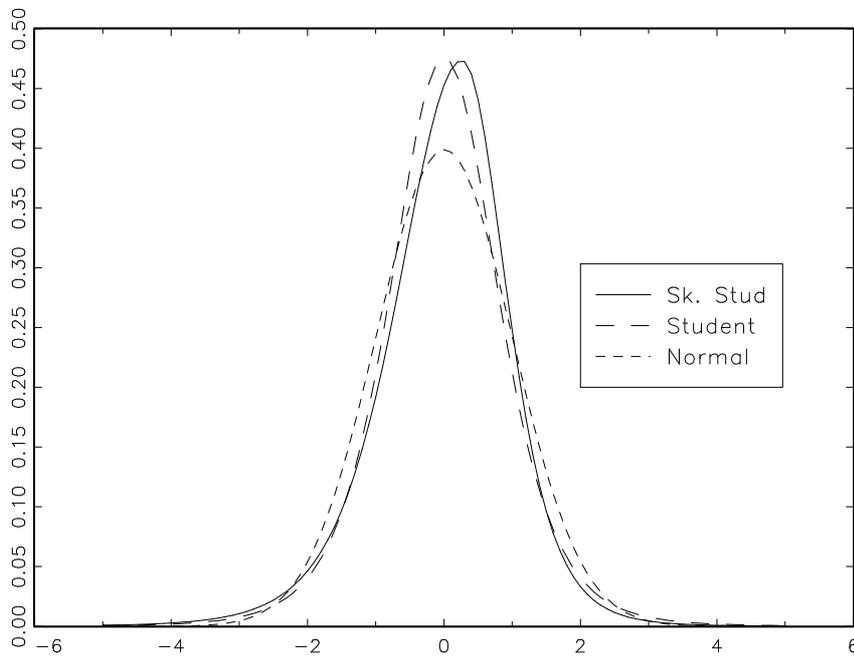


Figure 2.10: $st(0, 1, \exp(-0.179), 6.039)$ (solid line), $t(0, 1, 5.57)$ (dashed line) and $N(0, 1)$ (short dashes).

identify these outliers in a more formal way in order to estimate their magnitude and check whether these outliers are responsible for the asymmetry and the fat-tails observed on this series.

An interesting approach to identifying “additive outliers” (AO) in the volatility has been proposed by Franses and Ghijssels (1999). All the details concerning this method of detection and correction of AO as well as its extension to the APARCH model are given in Appendix B. Franses and Ghijssels (1999) extend the work of Chang, Tiao, and Chen (1988) (originally in an ARMA framework) to a GARCH(1,1) model. These authors show that the implementation of this AO correction leads to a substantial improvement in the out-of-sample forecasting properties of the GARCH model. The procedure is carried out in a sequential way and requires five steps. The first step involves the estimation of the model with the “raw data”. In the second step, a statistic $\hat{AO}(\tau^*)$ is computed for each observation and compared to a predetermined value C . When $\hat{AO}(\tau^*)$ exceeds C , the impact of the AO is said to be significant. We use a conservative value of $C = 5.5$ for the test statistic (the authors use a value of $C = 4$). In the third step, the outlier-adjusted residual is computed for the observation corresponding to the most significant outlier. Using this residual, the fourth step computes the additive outlier-corrected returns corresponding to this observation. Finally, the model is re-estimated with the new data and the procedure is carried out again until no more outliers are detected. Applying Franses and Ghijssels’ approach to our data allows us to quantify the number of “aberrant observations”, to identify these outliers and to yield AO corrected returns. The procedure leads to the identification of 76 outliers for the NASDAQ (for the first 2000 observations): 23 positive outliers and 53 negative outliers. Figure 2.11 plots the AO corrected returns (solid line) and the AO (circles). It is then possible to re-estimate the ARMA-APARCH model using the AO corrected returns. QML results are reported in Table 2.7 for three pseudo-likelihood functions: normal, skewed Student and skewed normal (with $v = \infty$). Note that the QML results of the skewed Student pseudo-likelihood obtained on the raw data (see Table 2.4) are reported in the column “skewed Student raw data” to facilitate the comparison.

The results from Table 2.7 suggest that the presence of AO is primarily responsible for the rejection of the normality assumption: adjusting for these outliers leads to a dramatic decrease of excess kurtosis. Indeed, the estimated number of degrees

Table 2.7: AR(1) - APARCH (1,1) model. Estimation results for AO corrected returns ($C = 5.5$).

	skewed Student raw data	normal AO	skewed Student AO	skewed normal AO
μ	0.0626 (0.0197)	0.0821 (0.0176)	0.0792 (0.0175)	0.0792 (0.0175)
ψ_1	0.2787 (0.0239)	0.2845 (0.0223)	0.2756 (0.0222)	0.2754 (0.0223)
ω	0.0371 (0.0132)	0.0315 (0.0107)	0.0281 (0.0100)	0.0282 (0.0100)
α_1	0.1503 (0.0261)	0.0889 (0.0164)	0.0911 (0.0167)	0.0907 (0.0166)
γ	0.2419 (0.0953)	0.3214 (0.0978)	0.2755 (0.0910)	0.2761 (0.0908)
β_1	0.8229 (0.0360)	0.8803 (0.0246)	0.8818 (0.0246)	0.8820 (0.0245)
δ	1.3322 (0.2967)	0.9039 (0.2765)	0.9778 (0.2811)	0.9786 (0.2794)
$\ln(\xi)$	-0.1789 (0.0319)	0	-0.1324 (0.0329)	-0.1324 (0.0318)
v	6.0388 (0.8182)	∞	130.7839 (35.4572)	∞
<i>Log-Lik</i>	-1993.1	-1792.2	-1784.0	-1783.9

Robust standard errors are given in parentheses. *Log-Lik* refers to the log-likelihood value at maximum.

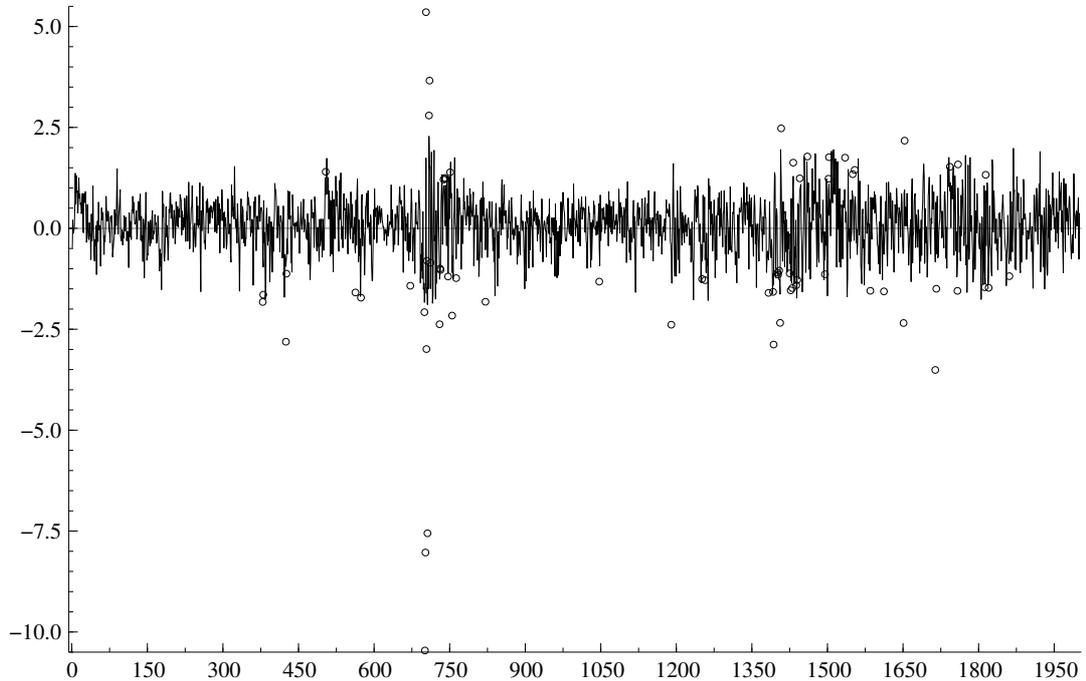


Figure 2.11: NASDAQ (solid line) corrected for 76 AO (circles).

of freedom is very high (130). Comparing the log-likelihoods, one clearly sees that the skewed Student and the skewed normal densities are indistinguishable. However, even after controlling for 76 AO, the corrected returns are still skewed: $\ln(\xi)$ is still significantly different from 0 (but slightly lower than for the raw data). Moreover, performing a LR test between the normal and skewed normal densities clearly rejects the former in favor of the latter. This suggests that the AO outliers are responsible for the high degree of kurtosis but not the asymmetry.

2.7 Conclusion

In this chapter we have first parameterized the skewed Student density proposed by Fernández and Steel (1998) in terms of the mean and of the variance parameters. This density is very promising for modelling financial series that exhibit skewness and excess kurtosis. First, we have shown its practical applicability in a QML estimation procedure in the GARCH framework using a Monte Carlo simulation. Moreover, this density, based on a mixture of two truncated symmetric

densities, is easy to implement. Indeed, its pdf, cdf and inverse cdf are based on the corresponding functions of its symmetric versions (which are available in most statistical packages). We have shown the practical advantages of the skewed Student distribution by analyzing the NASDAQ on a 12 year period (on a daily basis). Pearson goodness-of-fit tests reject the normal and the Student densities, but not the skewed Student distribution. On the other hand, the adequacy of the mixture of normal distributions is questionable since the results of the goodness-of-fit test are ambiguous. The performance of the skewed Student density has been reinforced by out-of-sample one-step-ahead density forecast tests. Finally, we have investigated the effect of AO on the skewness and kurtosis observed on the NASDAQ and found that they are responsible for the fat-tails of this series but not for the asymmetry.²¹

Several extensions of the methodology presented in this chapter could be investigated. First, the skewness and the tail properties were assumed to be time-invariant (ξ and v do not depend on t). This assumption might be unrealistic in practice and might have to be generalized just as ARMA and GARCH-type models were found to mimic the dynamics observed in the first two conditional moments of high frequency financial time series. Why should we stop at the second moment? This question has been raised by Hansen (1994), who proposes to generalize the GARCH specification to higher moments. He introduces dynamics through the 3rd and 4th order moments by conditioning the asymmetry and fat-tail parameters on past residuals and their square. This specifications has been used by Jondeau and Rockinger (2000)²² and extended recently by Harvey and Siddique (1999) who condition the skewness on past cubed residuals and past conditional skewness.²³ Recently, Lambert and Laurent (2000) have proposed a General Dynamic Model for Skewness (GDMS) to allow skewness to change over time in a

²¹Note that the asymmetry features of the NASDAQ are not only valid for daily data but also for weekly data. Indeed results, not reported in the thesis, concerning mid-week data are very similar in the sense that a leverage effect in combination with a skewed Student is found to be relevant for modelling this series.

²²Jondeau and Rockinger (2000) express skewness and kurtosis of Hansen's GARCH model as a function of the underlying parameters. The cost of such a flexibility is that for a dataset of about 7,000 observations, they have to impose not less than 20,000 restrictions to ensure that the corresponding conditional skewness and kurtosis exist. This difficult estimation problem is solved using a recent sophisticated sequential quadratic optimization algorithm.

²³Contrary to Jondeau and Rockinger (2000), Harvey and Siddique (1999) do not impose the existing constraints of the skewness and kurtosis.

totally different way than in previous works. The GDMS is based on the skewed Student density presented in this chapter and uses the fact that ξ^2 is the ratio of probability masses above and below the mode. Similar to an ARMA specification, the GDMS expresses $\ln(\xi^2)$ as²⁴ a function of its past values (the AR term), its past empirical counterparts (the MA term) and a constant term.²⁵ This extension will not be pursued in this thesis but is currently being investigated in Lambert and Laurent (2000).

A common feature of nearly all the empirical applications that rely on a non-normal density for the innovations and/or a complex specification for the conditional variance is that these models are estimated by maximum likelihood methods and use numerical techniques to approximate the derivatives of the likelihood function with respect to the parameter vector. To avoid numerical inefficiencies and highly speed-up maximum-likelihood estimation, the purpose of the next chapter is to provide numerically reliable analytical expressions for the score vector when the likelihood function is a (standardized) skewed Student density and the conditional variance follows an APARCH(p, q) specification.

²⁴As in the Exponential GARCH (EGARCH) model of Nelson (1991), the \ln transformation is used to avoid to care about the positivity constraints of ξ^2 .

²⁵The empirical counterpart of ξ^2 used by Lambert and Laurent (2000) is the number of times an observation has been observed above and below the corresponding (predictive) conditional mode up to and including time $t - 1$.

Chapter 3

Analytical scores and Gaussian QML relative efficiency for the APARCH skewed Student model

3.1 Introduction

It has been shown in the previous chapter that daily financial returns are heteroscedastic, fat-tailed and can also be skewed. To account for these three stylized facts, we have shown that an APARCH specification combined with a skewed Student density does a good job in modelling daily returns of the NASDAQ. This choice will be confirmed in Chapters 5 and 6.

The estimation of this model was done using maximum likelihood methods, relying on numerical techniques to approximate the derivatives of the likelihood function with respect to the parameter vector (the score or gradient vector). This is indeed the usual approach when one deals with non-linear models and especially a non-normal likelihood function. However, as shown by Fiorentini, Calzolari, and Panattoni (1996), and McCullough and Vinod (1999), using analytical scores in the estimation procedure should improve the numerical accuracy of the resulting estimates and speed-up maximum-likelihood estimation.

This chapter derives analytical expressions for the score of univariate APARCH models when the innovation process has a skewed Student distribution and illustrates the loss of efficiency of the Gaussian QML estimator when the innovations are skewed and/or fat-tailed.

The rest of the chapter is organized in the following way. In Section 3.2 and 3.3,

we briefly review the (standardized) skewed Student density. Section 3.4 provides analytical scores of this density. In section 3.5, QML results are summarized and a Gaussian QML relative efficiency is investigated when the true density is a skewed Student. Section 3.6 presents the APARCH model and the associated gradients. Finally, Section 3.7 provides an empirical application and Section 3.8 concludes.

3.2 The model

High frequency financial returns (y_t) are known to be heteroskedastic. y_t ($t = 1, \dots, T$) is typically modelled as follows:

$$y_t = \bar{\mu}_t + \varepsilon_t \quad (3.1)$$

$$\varepsilon_t = \sigma_t z_t \quad (3.2)$$

$$\bar{\mu}_t = c(\mu | \Omega_{t-1}) \quad (3.3)$$

$$\sigma_t = h(\mu, \eta | \Omega_{t-1}), \quad (3.4)$$

where $\bar{\mu}_t$ and σ_t^2 are respectively the conditional mean and conditional variance of y_t and $c(\cdot | \Omega_{t-1})$ and $h(\cdot | \Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t - 1$) depending on unknown vectors of parameters μ and η . Note that depending on the choice of $h(\cdot)$, constraints on η are often needed to insure that $Pr(\sigma_t^2 > 0) = 1$ for all t .

It is also widely accepted that high frequency financial returns are fat-tailed and even skewed. To accommodate these stylized facts, and following the discussion of Chapter 2, let us assume that conditional on Ω_{t-1} , z_t is *i.i.d.* $SKST(0, 1, \xi, \nu)$, i.e. z_t is independent and identically distributed as a standardized (with mean 0 and unit variance) skewed Student (SKST), with asymmetry parameter ξ and number of degree of freedom $\nu > 2$. Recall that when $\xi = 1$, one recovers the symmetric Student density.

3.3 The log-likelihood function

Let $\Phi = (\mu', \eta', \xi, \nu)'$ denote the vector of parameters of interest. The approximate ML estimator of Φ , denoted $\hat{\Phi}$, can thus be obtained by maximizing (apart from initial conditions) the corresponding likelihood (for a sample size of T):

$L_T(\Phi) = \sum_{i=1}^T l_t(\Phi)$, where

$$\begin{aligned}
 l_t(\Phi) &= \ln \left(\frac{2}{\xi + \frac{1}{\xi}} \right) + \ln \Gamma \left(\frac{\nu + 1}{2} \right) - 0.5 \ln [\pi(\nu - 2)] - \ln \Gamma \left(\frac{\nu}{2} \right) + \ln [s(\xi, \nu)] \\
 &- 0.5 \ln [\sigma_t^2(\mu', \eta')] - 0.5 (1 + \nu) \ln [g_t(\Phi)]
 \end{aligned} \tag{3.5}$$

and

$$\begin{aligned}
 g_t(\Phi) &= 1 + \frac{z_t^{*2}}{\nu - 2} \\
 z_t^* &= [s(\xi, \nu)z_t + m(\xi, \nu)] \xi^{-I_t} \\
 I_t &= \begin{cases} 1 & \text{if } z_t^* \geq 0 \\ -1 & \text{if } z_t^* < 0 \end{cases},
 \end{aligned}$$

where parameters $m = m(\xi, \nu)$ and $s = s(\xi, \nu)$ are respectively the mean and the standard deviation of the non-standardized skewed Student of Fernández and Steel (1998), i.e. $SKST(m, s^2, \xi, \nu)$, and are defined in Eq. (2.27) and (2.28).

For conditional heteroskedastic models and even more in non-Gaussian cases, the derivatives of the likelihood function with respect to the parameter vector are usually obtained using numerical techniques. However, as shown by Fiorentini, Calzolari, and Panattoni (1996) and McCullough and Vinod (1999), the use of analytical scores in the estimation procedure could:

- improve the numerical accuracy of the resulting estimates.¹ In this respect, Fiorentini, Sentana, and Calzolari (2000) show that it is very difficult to numerically distinguish a Student t with 100 degrees of freedom from another with 5,000 degrees of freedom, even when the sample size is large;
- and speed up maximum-likelihood estimation. As explained by Gable, Van Norden, and Vigfusson (1997), the computation of numerical gradients typically requires $N + 1$ evaluations of the likelihood function to calculate the N elements of the score, and $N^2 + 1$ to calculate the Hessian (the matrix of second

¹A clarification is needed when we talk about the accuracy of the numerical and analytical gradients. Indeed, when programming the analytical scores we also rely on numerical tools. Due to the complexity of the calculus, the computation is usually done by a computer (and not by hand), using a particular software (or programming language). For instance, the computation of $\ln x$ is not perfectly exact but gives an approximation of the true solution, with a given precision. Consequently, the computation of the analytical gradient is not “error free”. However, for simplicity, we will consider the analytical gradient as the benchmark

derivatives). By using analytical gradients, the number of calculations required to evaluate either of these objects can be greatly reduced. This in turn considerably speeds up maximum-likelihood estimation of such models with no loss in accuracy.

3.4 Analytical gradients of the skewed Student density

This section proposes an analytical formula for the score of a skewed Student density, irrespective of the specification used in the conditional mean and conditional variance.

To do that, let us look at the elements of the score vector separately:

$$\frac{\partial l_t(\Phi)}{\partial \mu}, \frac{\partial l_t(\Phi)}{\partial \eta}, \frac{\partial l_t(\Phi)}{\partial \xi} \text{ and } \frac{\partial l_t(\Phi)}{\partial \nu},$$

where $l_t(\Phi)$ is given in Eq. (3.5).

After standard algebraic manipulations, the partial derivatives of $l_t(\Phi)$ with respect to the conditional mean and conditional variance parameters μ and η are:

$$\begin{aligned} \frac{\partial l_t(\Phi)}{\partial \mu} &= -\frac{0.5}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \mu} \left[1 - \frac{(v+1)}{(v-2)} g_t(\Phi)^{-1} s \xi^{-I_t} z_t z_t^* \right] \\ &\quad - \frac{\partial \varepsilon_t}{\partial \mu} \frac{(v+1)}{(v-2)} g_t(\Phi)^{-1} s \xi^{-I_t} z_t^* \sigma_t^{-1} \end{aligned} \quad (3.6)$$

$$\frac{\partial l_t(\Phi)}{\partial \eta} = -\frac{0.5}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \eta} \left[1 - \frac{(v+1)}{(v-2)} g_t(\Phi)^{-1} s \xi^{-I_t} z_t z_t^* \right], \quad (3.7)$$

where $\frac{\partial \varepsilon_t}{\partial \mu}$ and $\frac{\partial \sigma_t^2}{\partial \eta}$ depend on the particular specification adopted in Eq. (3.3) and (3.4). When set equal to zero, these partial derivatives have been solved by Engle (1982) for the simple ARCH model, by Fiorentini, Calzolari, and Panattoni (1996) for the GARCH model and by Chung (1999) for the ARFIMA-FIGARCH (autoregressive fractionally integrated moving average - fractionally integrated GARCH) model.² Note also that as pointed out by Fiorentini, Calzolari, and Panattoni (1996), Eq. (3.3) and (3.4) are often recursively defined and thus require to choose

²Recently, Lombardi and Gallo (2001) have derived analytic expressions for the second-order derivatives of the Gaussian log-likelihood function of FIGARCH processes.

some initial values to start up the recursion. It is thus important to account for these starting values when computing the analytical gradients (see Fiorentini, Calzolari, and Panattoni, 1996, for more details).

Similarly, one can also show that differentiating with respect to the asymmetry parameter ξ and the number of degrees of freedom ν gives:

$$\frac{\partial l_t(\Phi)}{\partial \xi} = \frac{1 - \xi^2}{\xi^3 + \xi} + \frac{1}{s} \frac{\partial s}{\partial \xi} - \frac{(\nu + 1)}{(\nu - 2)} g_t^{-1}(\Phi) z_t^* \xi^{-I_t} \left[\frac{\partial s}{\partial \xi} z_t + \frac{\partial m}{\partial \xi} - \frac{I_t}{\xi} (s z_t + m) \right], \quad (3.8)$$

where $\frac{\partial s}{\partial \xi} = (\xi - \xi^{-3} - m \frac{\partial m}{\partial \xi}) s^{-1}$, $\frac{\partial m}{\partial \xi} = \frac{\Gamma(\frac{\nu-1}{2}) \sqrt{\nu-2}}{\sqrt{\pi} \Gamma(\nu/2)}$

and

$$\begin{aligned} \frac{\partial l_t(\Phi)}{\partial \nu} &= 0.5 \left[F\left(\frac{\nu+1}{2}\right) - F\left(\frac{\nu}{2}\right) - \frac{1}{\nu-2} \right] - \frac{m}{s^2} \frac{\partial m}{\partial \nu} - 0.5 \ln [g_t(\Phi)] \\ &\quad - 0.5 \frac{(\nu+1)}{(\nu-2)} g_t^{-1}(\Phi) \left(2 z_t^* \frac{\partial z_t^*}{\partial \nu} - \frac{z_t^{*2}}{\nu-2} \right), \end{aligned} \quad (3.9)$$

where $F(x) = \frac{\partial \ln \Gamma(x)}{\partial x}$ is the di-gamma function, $\frac{\partial z_t^*}{\partial \nu} = \xi^{-I_t} \left(\frac{\partial s}{\partial \nu} z_t + \frac{\partial m}{\partial \nu} \right)$, $\frac{\partial s}{\partial \nu} = -\frac{m}{s} \frac{\partial m}{\partial \nu}$ and $\frac{\partial m}{\partial \nu} = 0.5 \left(\xi - \frac{1}{\xi} \right) \frac{\Gamma(\frac{\nu-1}{2}) \sqrt{\nu-2}}{\sqrt{\pi} \Gamma(\frac{\nu}{2})} \left[\frac{1}{\nu-2} + F\left(\frac{\nu+1}{2}\right) - F\left(\frac{\nu}{2}\right) \right]$.

Notice that, as suggested in Chapter 2, working with $\ln(\xi)$ might be preferable to indicate the sign of the skewness. Consequently, the analytical gradient $\frac{\partial l_t(\Phi)}{\partial \ln(\xi)} = \xi \frac{\partial l_t(\Phi)}{\partial \xi}$.

To judge the usefulness of using analytical scores, Fiorentini, Sentana, and Calzolari (2000) have shown that numerical gradients of the number of degrees of freedom (ν) of a Student *t* likelihood are very poor approximations for the score function, especially when $\nu \rightarrow \infty$ (or $\frac{1}{\nu} \rightarrow 0$). Figures 1 and 2 plot the difference between numerical and analytical scores of the asymmetry parameter $\ln(\xi)$.^{3,4} In Figure 3.1, $\ln(\xi)$ ranges from -2 to 2 when ν is set to 8. From this figure, one can see

³The gradients are evaluated for different values of $\ln(\xi)$ while $\bar{\mu}_t$ and σ_t are set respectively to 0 and 1.

⁴Computations have been done with the software package GAUSS 3.5. The numerical scores are obtained using the standard *GRADP* procedure. Very similar results (but not reported to save space) are obtained using the *GRADRE* procedure. This procedure, provided with the optimization library OPTMUM (Aptech Systems, Inc.), implements the Richardson Extrapolation, an iterative process which updates a derivative based on values calculated in a previous iteration. This is slower than *GRADP*, but can in general, return values that are accurate to about 8 digits of precision.

that the difference between the numerical and analytical scores of the asymmetry parameter increases when $\ln(\xi)$ tends to 0. Similarly, Figure 3.2 plots the difference between numerical and analytical scores of the asymmetry parameter evaluated at $\ln(\xi) = 0$, for various values of ν (from 2.01 to 300).⁵ From this figure it is clear that the advantage of using analytical scores increases with the value of ν .

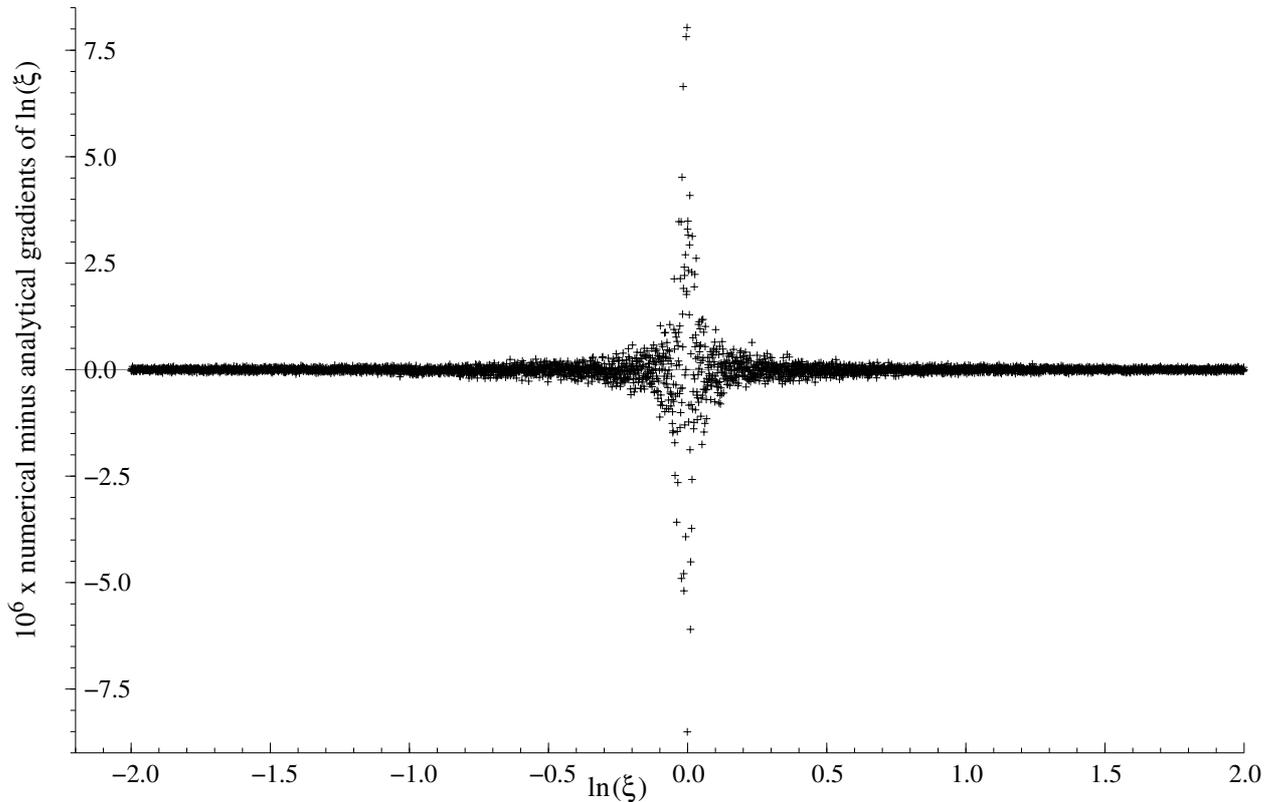


Figure 3.1: Difference between numerical and analytical scores of the log-likelihood with respect to $\ln(\xi)$ for $\nu = 8$.

Basically, these two figures reinforce the need of using analytical scores at least in two situations:

- in the estimation procedure, when the innovation process is nearly gaussian but estimated using a (skewed) Student density (estimating the unrestricted model to perform a Likelihood Ratio Test (LRT) for instance);

⁵Once again, $\bar{\mu}_t$ and σ_t are set respectively to 0 and 1.

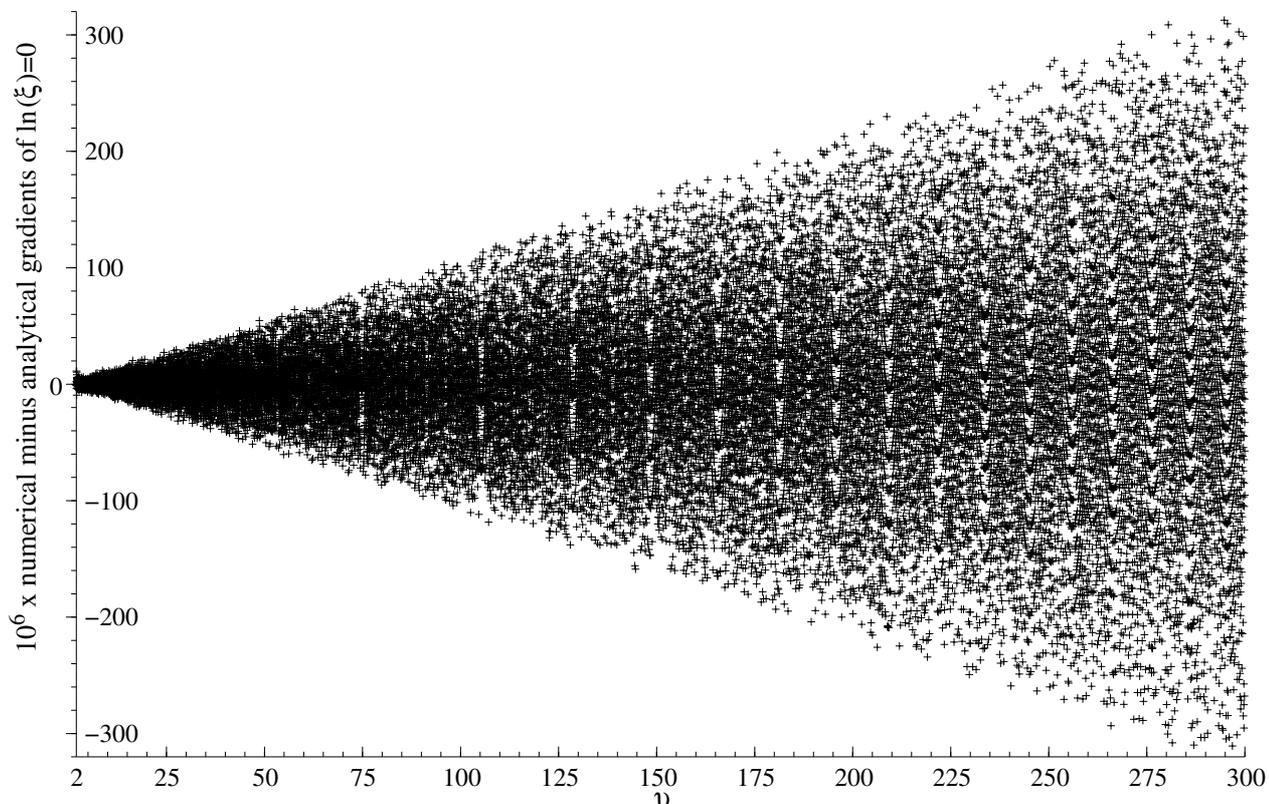


Figure 3.2: Difference between numerical and analytical scores of the log-likelihood with respect to v , evaluated at $\ln(\xi) = 0$.

- or when one has to evaluate the gradient vector under the null hypothesis $\frac{1}{v} = 0$ and/or $\ln(\xi) = 0$ to perform a Lagrange Multiplier (LM) test. See Fiorentini, Sentana, and Calzolari (2000) for an application of a LM test of multivariate normality.

3.5 Relative efficiency of QML estimator

In their seminal papers, Bollerslev and Wooldridge (1992) and Weiss (1986) studied the QML, of (multivariate) (G)ARCH models. They proved the consistency and asymptotic normality of the QML estimator under some regularity conditions.

Even if the normality assumption does not hold, maximizing the Gaussian log-likelihood function of a GARCH model provides consistent estimates of the

parameters. However, the standard errors have to be adjusted. Let $\hat{\Phi}$ be the estimate that maximizes the gaussian log-likelihood function and let Φ_0 be the true value that characterizes the GARCH model. Under certain regularity conditions:

$$\sqrt{T} \left(\hat{\Phi} - \Phi_0 \right) \xrightarrow{L} N \left(0, A_0^{-1} B_0 A_0^{-1} \right), \quad (3.10)$$

where \xrightarrow{L} means “converges in distribution to”. In other words, the asymptotic covariance matrix of $\sqrt{T} \left(\hat{\Phi} - \Phi_0 \right)$ is equal to $A_0^{-1} B_0 A_0^{-1}$, where A_0 is the information matrix evaluated at the true parameter vector Φ_0 , i.e.

$$A_0 = -T^{-1} \sum_{t=1}^T E \left(\frac{\partial^2 l_t(\Phi_0)}{\partial \Phi \partial \Phi'} \right), \quad (3.11)$$

and B_0 is the expected value of the outer product of the score matrix,

$$B_0 = T^{-1} \sum_{t=1}^T E \left(\frac{\partial l_t(\Phi_0)}{\partial \Phi} \frac{\partial l_t(\Phi_0)}{\partial \Phi'} \right). \quad (3.12)$$

Obviously, when the conditional density is truly normal, the matrices A_0 and B_0 are identical (except for the sign) and the asymptotic covariance matrix of the ML estimator is given by A_0^{-1} .

The matrices A_0 and B_0 can be consistently estimated by:

$$\hat{A}_T(\hat{\Phi}) = -T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 l_t(\hat{\Phi})}{\partial \Phi \partial \Phi'} \right) \quad (3.13)$$

$$\hat{B}_T(\hat{\Phi}) = T^{-1} \sum_{t=1}^T \left(\frac{\partial l_t(\hat{\Phi})}{\partial \Phi} \frac{\partial l_t(\hat{\Phi})}{\partial \Phi'} \right). \quad (3.14)$$

An analytical solution of $\hat{A}_T(\hat{\Phi})$ and $\hat{B}_T(\hat{\Phi})$ is provided by Engle (1982) in the case of an ARCH (q) model, Fiorentini, Calzolari, and Panattoni (1996) in the case of a GARCH (p, q) model and Lombardi and Gallo (2001) for the FIGARCH (p, d, q).

Even when the QML procedure provides a tractable solution to find a consistent estimator, this estimator is inefficient, with the degree of inefficiency increasing with the degree of departure from normality (see Engle and González-Rivera, 1991). To gain in efficiency one could thus search for a more appropriate distribution. A possible candidate distribution is the skewed Student density presented

in Chapter 2. To quantify the potential efficiency gain in using a skewed Student density (if this assumption holds), one can compute for a parameter ϕ , its ratio of Relative Efficiency (RE) of the QML estimator. Following Engle and González-Rivera (1991), the RE is defined as follows:

$$RE_\phi = \frac{\text{var} \left(\hat{\phi}_{MLE} \right)}{\text{var} \left(\hat{\phi}_{QML} \right)} \quad (3.15)$$

and is the ratio of the asymptotic variance of estimators of ϕ when the true density is known (the skewed Student) to its asymptotic variance when normality has been assumed (QML). $\text{var} \left(\hat{\phi}_{QML} \right)$ is obtained by using Eq. (3.10) (replacing A_0 and B_0 by $\hat{A}_T(\hat{\Phi})$ and $\hat{B}_T(\hat{\Phi})$) while $\text{var} \left(\hat{\phi}_{MLE} \right)$ is obtained using the standard MLE techniques, i.e. it is the asymptotic variance of the MLE estimator based on the true density. A consistent estimate of the variance covariance matrix of $\hat{\Phi}_T$ is given by $\hat{A}_T^{-1}(\hat{\Phi})$ or $\hat{B}_T^{-1}(\hat{\Phi})$ where l_t in Eq. (3.13) and (3.14) is the skewed Student log-likelihood function reported in Eq. (3.5).

The RE ratio is bounded: $0 < RE \leq 1$.⁶ Obviously, if the density is truly normal, $A_\phi^0 = -B_\phi^0$ and thus $RE_\phi = A_\phi^{0^{-1}} A_\phi^0 B_\phi^{0^{-1}} A_\phi^0 = 1$ and consequently, as the efficiency of the QML estimator decreases, RE tends towards 0.

In Table 3.1, we report the RE results for different values of the parameters of a skewed Student GARCH model. This model is specified as follows:

$$y_t = \mu + \varepsilon_t \quad (3.16)$$

$$\varepsilon_t = \sigma_t z_t \quad (3.17)$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad (3.18)$$

where z_t is *i.i.d.* $SKST(0, 1, \xi, v)$.

The covariance matrix of the set of parameters $\Phi = (\mu, \omega, \alpha, \beta)'$ is first obtained using the QML estimator (see Eq. (3.10)) and second using the MLE (with the true density) as the inverse of expectation of the outer product of the score. The computation of the score vector and the Hessian matrix in the QML proce-

⁶This result holds asymptotically. However, in small sample it is possible that this constraint does not hold.

ture can be done using numerical approximations.⁷ Instead, and in order to have more accurate estimations, we follow Fiorentini, Calzolari, and Panattoni (1996) who provide the analytical scores and the analytical Hessian of the GARCH (p, q) model when the innovations are normally distributed.⁸ To compute the variance covariance matrix of the MLE (skewed Student), we use analytical scores presented in section 3.4 and again follow Fiorentini, Calzolari, and Panattoni (1996) for the computation of the partial derivatives with respect to the GARCH model. In this experiment, we use two sets of parameter values for the GARCH model. In the first one (columns 3 to 6), $\mu = 0.1, \omega = 0.1, \alpha_1 = 0.1, \beta = 0.8$ while in the second one (columns 7 to 10) $\mu = 0.1, \omega = 0.1, \alpha_1 = 0.4, \beta = 0.5$. On the other hand, the innovations z_t are skewed Student distributed, i.e. $z_t \sim i.i.d. SKST(0, 1, \xi, \nu)$. To illustrate the link between the skewness and fat-tails parameters, we use a set of 24 combinations of $\ln(\xi)$ and ν , i.e. $\ln(\xi) = 0, 0.1, \dots, 0.5$ and $\nu = 30, 15, 8, 5$. The number of observations $T = 10^6$.

From Table 3.1, one clearly sees that the efficiency of the QML estimator increases when ν increases and $\ln(\xi)$ decreases (with less evidence for the mean parameter μ), in other words, when the skewed Student density tends to the normal. These results are in line with those of Engle (1982) who present similar results but for Student and Gamma innovations (the latter is both skewed and fat-tailed). For example, when $\ln(\xi) = 0.1$ (which means that the skewed Student density allocates 55 % of the mass to right side of the mode) and $\nu = 5$, the coefficient of skewness of z_t is 0.44 (0 for the Gaussian density), while its kurtosis equals 9.35 (3 for the normal).⁹ The resulting RE ratio is about 0.44 for the GARCH parameters which means that in that case the asymptotic variance of the QML estimator is around 2.5 times larger than the variance of the MLE estimator (minimum variance).¹⁰

These results hold also for left skewed innovations, i.e. $\ln(\xi) = 0, -0.1, \dots, -0.5$. Note that this procedure requires the evaluation of the score function of the skewed Student likelihood when this density is nearly Gaussian ($\nu = 30$ and $\ln(\xi) = 0$).

⁷The Hessian matrix can be obtained numerically in GAUSS using the standard procedure *HESSP*.

⁸The computation of the Hessian matrix has been done using a slightly modified version of the gauss procedures provided by Franses and van Dijk (2000). The programs can be downloaded from <http://www.few.eur.nl/few/people/franses>.

⁹The formulas used to compute the skewness and kurtosis of the innovation process, for a given value of ξ and ν , are given in Chapter 2, Eq. (2.24) and (2.25).

¹⁰Different values of α and β do not seem to affect much the RE ratios.

Recalling from the previous section that numerical techniques are known to give very poor results in this situation justify our choice for the use of analytical gradients.

QML can thus provide consistent estimators of the asymptotic covariance matrix. However, the results presented in the above table suggest that using the correct density in the maximization procedure may provide more accurate estimators of the covariance matrix (at least asymptotically) and thus may increase the behavior of the tests statistics based on this estimator.¹¹

3.6 APARCH specification

We have shown in Chapter 2 that an APARCH model, combined with a skewed Student density performs very well in modelling and forecasting daily financial returns. The APARCH (Ding, Granger, and Engle, 1993) is an extension of the GARCH model of Bollerslev (1986). It is probably one of the most promising ARCH-type model. Indeed, it nests at least seven GARCH specifications. The APARCH(p, q) can be defined as follows:

$$y_t = x'_{1,t}\mu + \varepsilon_t \quad (3.19)$$

$$\varepsilon_t = \sigma_t z_t \quad (3.20)$$

$$\sigma_t^\delta = x'_{2,t}\omega + \sum_{i=1}^q \alpha_i k(\varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta \quad (3.21)$$

$$k(\varepsilon_{t-i}) = |\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}, \quad (3.22)$$

where $x_{1,t}$ and $x_{2,t}$ are two vectors of respectively n_1 and n_2 weakly exogenous variables (including the intercept), μ, ω, α_i 's, γ_i 's, β_j 's and δ are parameters (or vectors of parameters) to be estimated. δ ($\delta > 0$) plays the role of a Box-Cox transformation of σ_t , while the γ_i 's allow a different effect of a positive and a negative shock on volatility. The properties of the APARCH model have been studied recently by He and Teräsvirta (1999a, 1999b). It is convenient to start the recursion of Eq. (3.21) by setting unobserved components to their sample average, i.e. setting

$$k(\varepsilon_{t-i})^\delta = \frac{1}{T} \sum_{s=1}^T (|\varepsilon_s| - \gamma_i \varepsilon_s)^\delta \text{ for } t \leq i \text{ and } \sigma_t^\delta = \left(\frac{1}{T} \sum_{s=1}^T \varepsilon_s^2 \right)^{\frac{\delta}{2}} \text{ for } t \leq 0.$$

¹¹As mentioned by Engle (1982), the precision of the forecast can also be affected.

Table 3.1: Relative efficiency of QML estimator

$\ln(\xi)$	v	$\mu = 0.1$	$\omega = 0.1$	$\alpha = 0.1$	$\beta = 0.8$	$\mu = 0.1$	$\omega = 0.1$	$\alpha = 0.4$	$\beta = 0.5$
0	30	1.00	1.00	0.95	0.98	1.00	0.99	0.95	0.98
0	15	0.99	1.00	0.96	0.98	0.97	0.96	0.94	0.94
0	8	0.98	0.83	0.82	0.82	0.92	0.82	0.81	0.81
0	5	0.89	0.44	0.44	0.43	0.74	0.44	0.44	0.43
0.1	30	1.00	0.96	0.97	0.96	0.99	0.95	0.95	0.94
0.1	15	0.98	0.97	0.92	0.94	0.96	0.92	0.92	0.90
0.1	8	0.98	0.83	0.80	0.80	0.90	0.81	0.80	0.80
0.1	5	0.88	0.43	0.42	0.42	0.70	0.42	0.42	0.41
0.2	30	0.99	0.86	0.86	0.85	0.98	0.89	0.91	0.90
0.2	15	0.98	0.82	0.84	0.81	0.95	0.88	0.87	0.87
0.2	8	0.97	0.73	0.71	0.73	0.89	0.71	0.70	0.67
0.2	5	0.87	0.34	0.35	0.32	0.66	0.41	0.38	0.38
0.3	30	0.98	0.86	0.80	0.85	0.95	0.84	0.81	0.82
0.3	15	0.98	0.72	0.74	0.71	0.93	0.78	0.74	0.74
0.3	8	0.96	0.63	0.59	0.61	0.86	0.62	0.61	0.59
0.3	5	0.88	0.32	0.33	0.31	0.69	0.38	0.32	0.32
0.4	30	0.98	0.69	0.70	0.68	0.92	0.74	0.74	0.72
0.4	15	0.95	0.63	0.65	0.61	0.88	0.67	0.67	0.63
0.4	8	0.95	0.43	0.43	0.42	0.82	0.43	0.43	0.39
0.4	5	0.86	0.24	0.23	0.21	0.63	0.32	0.26	0.26
0.5	30	0.97	0.61	0.60	0.60	0.90	0.64	0.63	0.61
0.5	15	0.96	0.49	0.56	0.49	0.87	0.54	0.55	0.49
0.5	8	0.93	0.39	0.38	0.38	0.79	0.39	0.41	0.35
0.5	5	0.81	0.16	0.17	0.16	0.44	0.13	0.21	0.16

Model: $y_t = \mu + \varepsilon_t$, $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ and $z_t = \frac{\varepsilon_t}{\sigma_t} \sim i.i.d. SKST(0, 1, \xi, v)$ where $\ln(\xi) = 0, 0.1, \dots, 0.5$ and $v = 30, 15, 8, 5$. The number of observations $T = 10^6$. The first two columns report the parameter values of $\ln(\xi)$ and v used in the Data Generating Process. The next four columns report the RE ratio for the first DGP while the last columns concern the second DGP.

To the best of our knowledge, up to now, the analytical gradients of the APARCH model have not been provided in the literature. This is probably due to the high degree of nonlinearity of this specification which makes their computation less trivial than in the ARCH case.

To achieve this goal, let us define $\gamma = (\gamma_1, \dots, \gamma_q)$ the vector of q “leverage effect” parameters, $d_t = (x_{2,t}, k(\varepsilon_{t-1})^\delta, \dots, k(\varepsilon_{t-q})^\delta, \sigma_{t-1}^\delta, \dots, \sigma_{t-p}^\delta)'$ and $\eta = (\vartheta', \gamma', \delta)'$ the vector of $(n_2 + 2q + p + 1)$ unknown parameters of the conditional dispersion equation, where $\vartheta = (\omega', \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)'$.¹²

From Eq. (3.6), one can see that differentiating the log-likelihood function with respect to μ requires an analytical expression for $\frac{\partial \varepsilon_t}{\partial \mu}$. In our case, the solution is trivial and is: $\frac{\partial \varepsilon_t}{\partial \mu} = -x'_{1,t}$. As shown in Eq. (3.6) and (3.7), differentiating the log-likelihood function with respect to μ and η also requires the computation of $\frac{\partial \sigma_t^2}{\partial \mu}$ and $\frac{\partial \sigma_t^2}{\partial \eta}$ while in the APARCH specification, a power transform of the conditional variance is modelled (σ_t^δ) . One can solve this problem by re-writing σ_t^2 as $(\sigma_t^\delta)^{\frac{2}{\delta}}$ which leads to:

$$\frac{\partial \sigma_t^2}{\partial (\mu', \vartheta', \gamma')} = \frac{2\sigma_t^2}{\delta \sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial (\mu', \vartheta', \gamma')} \quad (3.23)$$

and

$$\frac{\partial \sigma_t^2}{\partial \delta} = \frac{2\sigma_t^2}{\delta \sigma_t^\delta} \left(\frac{\partial \sigma_t^\delta}{\partial \delta} - \frac{\sigma_t^\delta \ln(\sigma_t^\delta)}{\delta} \right). \quad (3.24)$$

Our goal is thus to find a tractable solution of $\frac{\partial \sigma_t^\delta}{\partial \mu}$ and $\frac{\partial \sigma_t^\delta}{\partial \eta}$ which can be done in four steps.

- First step. Given the choice we made for the initial values of the pre-sample terms $k(\varepsilon_{t-i})^\delta$ and σ_t^δ , differentiating with respect to the conditional mean

¹²Note that the APARCH model is compatible with the parameterization given in Eq. (3.3) and (3.4). Indeed, in this case $c(\mu|\Omega_{t-1}) = x'_{1,t}\mu$ and

$$h(\mu, \eta|\Omega_{t-1}) = \left[x'_{2,t}\omega + \sum_{i=1}^q \alpha_i k(\varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta \right]^{\frac{1}{\delta}}.$$

parameters (μ) gives:

$$\begin{aligned} \frac{\partial \sigma_t^\delta}{\partial \mu} &= \delta \sum_{i=1}^q \alpha_i [k(\varepsilon_{t-i})^{\delta-1} (I_{t-i}^* + \gamma_i) x_{1,t-i}]^{\mathfrak{S}(t-i)} \\ &\times \left[\frac{1}{T} \sum_{s=1}^T (|\varepsilon_s - \gamma_i \varepsilon_s|)^{\delta-1} (I_s^* + \gamma_i) x_{1,s} \right]^{1-\mathfrak{S}(t-i)} \\ &+ \sum_{j=1}^p \beta_j \left(\frac{\partial \sigma_{t-j}^\delta}{\partial \mu} \right)^{\mathfrak{S}(t-j)} \left[-\frac{\delta}{T} \left(\frac{1}{T} \sum_{s=1}^T \varepsilon_s^2 \right)^{\frac{\delta-2}{2}} \sum_{s=1}^T \varepsilon_s x_{1,s} \right]^{1-\mathfrak{S}(t-j)} \end{aligned} \quad (3.25)$$

where

$$I_t^* = \begin{cases} -1 & \text{if } \varepsilon_t > 0 \\ 1 & \text{if } \varepsilon_t < 0 \end{cases} \quad \text{and} \quad \mathfrak{S}_t = \begin{cases} 1 & \text{if } t > 0 \\ 0 & \text{if } t \leq 0 \end{cases}.$$

Note that $I_t^* = \frac{\partial |\varepsilon_t|}{\partial \mu}$ and is not defined for $\varepsilon_t = 0$. However, even in this situation it is possible, it is almost unlikely in practice.

- Second step.

$$\frac{\partial \sigma_t^\delta}{\partial \vartheta} = d_t + \sum_{j=1}^p \beta_j \frac{\partial \sigma_{t-j}^\delta}{\partial \vartheta}, \quad (3.26)$$

where $\frac{\partial \sigma_t^\delta}{\partial \vartheta} = 0$ for $t \leq 0$.

- Third step. In a similar way, one can show that:

$$\frac{\partial \sigma_t^\delta}{\partial \gamma} = d_t^* + \sum_{j=1}^p \beta_j \frac{\partial \sigma_{t-j}^\delta}{\partial \gamma}, \quad (3.27)$$

where d_t^* is a $(1 \times q)$ vector whose i^{th} element is $\alpha_i \frac{\partial k(\varepsilon_{t-i})^\delta}{\partial \gamma_i}$ with

$$\frac{\partial k(\varepsilon_{t-i})^\delta}{\partial \gamma_i} = \begin{cases} -\delta k(\varepsilon_{t-i})^{\delta-1} \varepsilon_{t-i} & \text{if } t > 0 \\ -\frac{\delta}{T} \sum_{s=1}^T (|\varepsilon_s - \gamma_i \varepsilon_s|)^{\delta-1} \varepsilon_s & \text{if } t \leq 0 \end{cases} \quad (3.28)$$

and $\frac{\partial \sigma_t^\delta}{\partial \gamma} = 0$ for $t \leq 0$.

- Last step. Finally, differentiating with respect to the δ gives:

$$\begin{aligned} \frac{\partial \sigma_t^\delta}{\partial \delta} &= \sum_{i=1}^q \alpha_i [k(\varepsilon_{t-i})^\delta \ln k(\varepsilon_{t-i})]^{\mathfrak{S}(t-i)} \left[\frac{1}{T} \sum_{s=1}^T (|\varepsilon_s| - \gamma_i \varepsilon_s)^\delta \ln (|\varepsilon_s| - \gamma_i \varepsilon_s) \right]^{1-\mathfrak{S}(t-i)} \\ &+ \sum_{j=1}^p \beta_j \left(\frac{\partial \sigma_{t-j}^\delta}{\partial \delta} \right)^{\mathfrak{S}(t-j)} \left[0.5 \left(\frac{1}{T} \sum_{s=1}^T \varepsilon_s^2 \right)^{\frac{\delta}{2}} \ln \left(\frac{1}{T} \sum_{s=1}^T \varepsilon_s^2 \right) \right]^{1-\mathfrak{S}(t-j)}. \end{aligned} \quad (3.29)$$

3.7 Empirical application

In this empirical application we consider daily data for a stock market indexes, i.e. the NIKKEI stock index for the period 4/1/1984 - 21/12/2000 (4246 observations, source: Datastream).

We consider an APARCH(1, 1) specification:

$$\begin{aligned} y_t &= \mu + \varepsilon_t \\ \sigma_t^\delta &= \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta \\ z_t &\sim i.i.d. SKST(0, 1, \xi, \nu). \end{aligned}$$

As in Chapter 2, estimation has been first considered using numerical gradients. In a second step (using the same starting values as in the first case), estimation has been carried out using the analytical gradients presented in the previous sections.

Table 3.2 presents QML estimation results of the APARCH(1,1) with a skewed Student pseudo-likelihood.

Several comments are in order.

1. First, the extra flexibility of the APARCH specification is required. Both the asymmetry coefficient (γ) and the power (δ) estimates suggest that a usual GARCH model is not appropriate to model the NIKKEI. This is also confirmed by Likelihood Ratio (LR) tests for the null hypothesis $H_0 : \delta = 1$ and $\gamma = 0$ (not reported here to save space).
2. Likelihood ratio tests (not reported) and standard t-test clearly suggest that the skewed Student density outperforms the normal and Student densities. The distribution of the NIKKEI is highly kurtosed and left skewed. It seems moreover that the asymmetry feature of the APARCH model (characterizing

Table 3.2: Skewed Student APARCH

	Numerical Score	Analytical Score
μ	0.03425 (0.01434)	0.03401 (0.01496)
ω	0.02468 (0.00445)	0.02470 (0.00449)
α_1	0.10664 (0.01097)	0.10664 (0.01098)
γ	0.48505 (0.06881)	0.48521 (0.06937)
β_1	0.89495 (0.01028)	0.89495 (0.01030)
δ	1.21853 (0.14072)	1.21835 (0.14106)
$\ln(\xi)$	-0.05315 (0.02210)	-0.05326 (0.02228)
v	6.47140 (0.58282)	6.47150 (0.58389)
Time (in sec.)	19.28	7.69

the conditional variance) and the “skewness” coefficient of the unconditional density are both necessary to explain the overall asymmetry of the series.

3. Comparing columns 2 and 3, one can see that numerical scores give very similar results to the analytical ones. This result is not surprising due to the fact that v is quite low in this example (see section 3.4). However, one can see that using the analytical scores highly speeds up the estimation procedure (using the same starting values in both cases). Using the analytical scores is about three times faster.

3.8 Conclusion

In the empirical literature, various densities have been proposed to account for the asymmetry and the fat-tails that we generally observe for high-frequency financial returns. These densities are in general combined with a complex specification for the conditional variance because these series are known to be heteroscedastic.

A common feature of nearly all the empirical applications that rely on a non-normal density for the innovations and/or a complex specification for the conditional variance is that these models are estimated by (approximate) maximum likelihood methods and use numerical techniques to approximate the derivatives

of the likelihood function with respect to the parameter vector. To avoid numerical inefficiencies and highly speed-up maximum-likelihood estimations we provide numerically reliable analytical expressions for the score vector when the likelihood function is a (standardized) skewed Student density and the conditional variance follows an APARCH(p, q) specification (which nests at least seven GARCH models). This choice has been motivated by the fact that this density is flexible enough to be skewed *and* fat-tailed, two features shared with most high-frequency financial time-series. We have also illustrated the loss of efficiency of the Gaussian QML estimator when the innovations are skewed and/or fat-tailed.

Up to now, attention has been restricted to univariate ARCH-type models and inevitably univariate densities. In this univariate framework, the skewed Student density appears to be a promising specification to accommodate both the high kurtosis and the skewness inherent to most asset returns. Given the interpretation of shocks as news and the fact that at least certain items affect various assets simultaneously, it might be suggested that the volatility of different assets moves together over time. It could thus be interesting to consider multivariate ARCH-type models to describe the volatility of these assets jointly. The estimation of these multivariate models often relies on the normality assumption (for simplicity). The next chapter will be devoted to show that the methodology we have presented in the previous chapter can be extended to provide a tractable solution to introduce skewness in any continuous unimodal and symmetric multivariate distribution (provided that the first two conditional moments of the marginal distributions exist).

Chapter 4

A New Class of Multivariate Skewed Densities, with Application to GARCH Models

4.1 Introduction

Many time series of asset returns can be characterized as serially dependent. This is revealed by the presence of positive autocorrelation in the squared returns, and sometimes to a much smaller extent by autocorrelation in the returns. We have shown in the previous chapters that the most widespread modelling approach to capture these properties is to specify a dynamic model for the conditional mean and the conditional variance, such as an ARMA-GARCH model or one of its various extensions (see the seminal paper of Engle, 1982). However, the first two conditional moments are not the only game in town. Indeed, Peiró (1999) emphasizes the relevance of modelling of higher-order moments for asset pricing models, portfolio selection and option pricing theories. Moreover, for asset returns that are skewed and fat-tailed, it is crucial to account for these features in order to obtain accurate Value-at-Risk forecasts (see Chapters 5 and 6).

Although there is a huge literature on univariate ARCH models, much less papers are concerned with their multivariate extensions. For this reason, Geweke and Amisano (2001) argue that “while univariate models are a first step, there is an urgent need to move on to multivariate modelling of the time-varying distribution of asset returns”. Indeed, financial volatilities move together over time across assets and markets. Recognizing this commonality through a multivariate modelling

framework can lead to obvious gains in efficiency and to more relevant financial decision making than working with separate univariate models.

Among the most widespread multivariate GARCH models, we find the Constant Conditional Correlations model (CCC) of Bollerslev (1990), the Vech of Kraft and Engle (1982) and Bollerslev, Engle, and Wooldridge (1988), the BEKK of Engle and Kroner (1995), the Factor GARCH of Ng, Engle, and Rothschild (1992), the General Dynamic Covariance (GDC) model of Kroner and Ng (1998), the Dynamic Conditional Correlations (DCC) model of Engle (2001) and the Time-Varying Correlation (TVC) model of Tse and Tsui (1998).¹

The estimation of multivariate GARCH models is commonly done by maximizing a Gaussian likelihood function. Even if it is unrealistic in practice, the normality assumption may be justified by the fact that the Gaussian QML estimator is consistent provided the conditional mean and the conditional variance are specified correctly. In this respect, Jeantheau (1998) has proved the strong convergence of the QML estimator of multivariate GARCH models, extending previous results of Lee and Hansen (1994) and Lumsdaine (1996).

As far as financial applications are concerned, and in order to gain statistical efficiency, it is of primary importance to base modelling and inference on a more suitable distribution than the multivariate normal. The challenge to econometricians is to design multivariate distributions that are both easy to use for inference and compatible with the skewness and kurtosis properties of financial returns. Otherwise it is very likely that the estimators will not be consistent (see Newey and Steigerwald, 1997).

To the best of our knowledge, asymmetric and fat-tailed k -variate distributions with support on the full Euclidian space of dimension k are uncommon. The main contribution of this chapter is to propose a practical and flexible method to introduce skewness in multivariate symmetric distributions by generalizing the technique presented in the previous chapter. Applying this procedure to the multivariate Student density leads to a “multivariate skewed Student” density, in which each marginal has a specific asymmetry coefficient. Combined with a multivariate GARCH model, this new family of distributions is potentially useful for modelling

¹Alternatively, Harvey, Ruiz, and Shephard (1994) propose a multivariate stochastic variance model, which has been extended in various ways. Even if this kind of model is also attractive, we limit our attention to multivariate GARCH models.

stock returns. In an application to the daily returns of the CAC40, NASDAQ, NIKKEI and the SMI, it is found that this density suits well the data and clearly outperforms its symmetric competitors.

The chapter is organized in the following way. In Section 4.2, we briefly review the univariate skewed Student density proposed by Fernández and Steel (1998) and extended in Chapter 2. In Section 4.3, we describe the new family of multivariate skewed densities, and in Section 4.4 we apply it in a multivariate GARCH framework. Finally, we offer our conclusions and ideas for further developments in Section 4.5.

4.2 Univariate case

A series of financial returns y_t ($t = 1, \dots, T$), known to be typically conditionally heteroscedastic, may be modelled as follows:

$$y_t = \mu_t + \varepsilon_t \quad (4.1)$$

$$\varepsilon_t = \sigma_t z_t \quad (4.2)$$

$$\mu_t = c(\mu|\Omega_{t-1}) \quad (4.3)$$

$$\sigma_t = h(\mu, \eta|\Omega_{t-1}), \quad (4.4)$$

where $c(\cdot|\Omega_{t-1})$ and $h(\cdot|\Omega_{t-1})$ are functions of Ω_{t-1} (the information set at time $t - 1$), depending on unknown vectors of parameters μ and η , and z_t is an independently and identically distributed (*i.i.d.*) process with $E(z_t) = 0$, $Var(z_t) = 1$. Assuming that their corresponding conditional moments exist, μ_t is the conditional mean of y_t and σ_t^2 is its conditional variance.

4.2.1 Skewed Student densities

Another well established stylized fact of financial returns, at least when they are sampled at high frequencies, is that they exhibit fat-tails, which corresponds to a kurtosis coefficient larger than three. Furthermore, in general these series are not symmetrically distributed (see Hansen, 1994, and Peiró, 1999 among others). To accommodate the unconditional skewness and excess kurtosis, we have proposed in Chapter 2 to replace the normal distribution used originally in GARCH models by the skewed Student distributions (see Definition 1).

The main advantages of this density are its ease of implementation, that its parameters have a clear interpretation, and that it performs well on financial datasets (see Paoletta, 1997; Lambert and Laurent, 2001; Giot and Laurent, 2001a and Giot and Laurent, 2001b). Moreover, we have shown how to obtain the cumulative distribution function (cdf) and the quantile function of a standardized skewed density from the cdf and quantile function of the corresponding symmetric density.

4.2.2 Empirical illustration

In this illustration, we consider four stock market indexes: the French CAC40, US NASDAQ, Japanese NIKKEI and Swiss SMI from January 1991 to December 1998 (1816 daily observations; source: Datastream). The daily return is defined as $y_t = 100 \times (\ln p_t - \ln p_{t-1})$ where p_t is the stock index value of day t .

We use the model defined by Eq. (4.1-4.4) with the following conditional mean and variance equations:

$$\mu_t = \mu + \phi(y_{t-1} - \mu) \quad (4.5)$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2, \quad (4.6)$$

where μ, ϕ, ω, β , and α are parameters to be estimated. An autoregressive model of order one is chosen for the conditional mean to allow for possible autocorrelation in the daily returns, while a GARCH(1,1) specification -see Bollerslev (1986)- is chosen for the conditional variance to account for volatility clustering in a simple way. We have shown in the first chapter that an APARCH(1,1) model seems to be indicated with daily returns of the NASDAQ. More sophisticated ARCH models could also be used (see Appendix A for a review of the major specifications). However, we rely on a simple GARCH specification to make easier the comparison with the multivariate model.

To account for possible skewness and fat tails, we estimated the AR(1)-GARCH (1,1) model assuming a skewed Student density for the innovations. In order to assess the practical relevance of this density, we compare the estimation results with two other assumptions regarding the innovations density: the normal (obtained when v tends to infinity and $\xi = 1$), and the symmetric Student (obtained by setting $\xi = 1$). Results concerning the CAC40 and the NASDAQ are gathered in Table 4.1 and those concerning the NIKKEI and the SMI are reported in Table 4.2. Several comments are in order:

Table 4.1: ML estimation results of AR-GARCH models for the CAC40 and the NASDAQ

	Normal	Student	skewed Student
	CAC40; NASDAQ	CAC40 ; NASDAQ	CAC40 ; NASDAQ
μ	0.051 ; 0.111 (0.029) ; (0.027)	0.057 ; 0.137 (0.027) ; (0.023)	0.053 ; 0.099 (0.027) ; (0.023)
ϕ	0.052 ; 0.177 (0.026) ; (0.026)	0.044 ; 0.171 (0.023) ; (0.024)	0.044 ; 0.152 (0.023) ; (0.024)
ω	0.094 ; 0.092 (0.074) ; (0.036)	0.043 ; 0.055 (0.028) ; (0.026)	0.042 ; 0.053 (0.027) ; (0.025)
β	0.860 ; 0.766 (0.076) ; (0.063)	0.915 ; 0.827 (0.037) ; (0.052)	0.915 ; 0.826 (0.037) ; (0.052)
α	0.078 ; 0.153 (0.034) ; (0.043)	0.056 ; 0.124 (0.022) ; (0.035)	0.056 ; 0.128 (0.022) ; (0.035)
$\ln(\xi)$	0 ; 0	0 ; 0	-0.014 ; -0.158 (0.031) ; (0.034)
v	∞ ; ∞	8.657 ; 5.685 (1.918) ; (0.753)	8.714 ; 5.938 (1.933) ; (0.817)
Q_{20}	27.511 ; 17.830	30.652 ; 17.925	27.289 ; 19.526
Q_{20}^2	8.682 ; 7.815	11.302 ; 10.720	10.990 ; 10.983
P_{20}	30.608 ; 62.344 (0.044) ; (0.000)	10.531 ; 37.815 (0.938) ; (0.006)	17.782 ; 12.338 (0.537) ; (0.870)
SIC	3.229 ; 2.802	3.197 ; 2.728	3.202 ; 2.720
$Log-Lik$	-2911.6 ; -2524.6	-2879.9 ; -2453.7	-2879.8 ; -2442.9

Each column reports the ML estimates of the model defined by Eq. (4.1)-(4.2)-(4.5)-(4.6), with robust standard errors underneath in parentheses. The column headed “Normal” corresponds to $z_t \sim N(0,1)$, “Student” to $z_t \sim ST(0,1,v)$ as in Eq. (2.13), “Skewed Student” to $z_t \sim SKST(0,1,\xi,v)$ as in Eq. (2.30), and in all cases z_t is an *i.i.d.* process. Q_{20} is the Box-Pierce statistic of order 20 on the standardized residuals, Q_{20}^2 is the same for their squares, P_{20} is the Pearson goodness-of-fit statistic (using 20 cells) with the associated p-value underneath in parentheses (see footnote 2). SIC is the Schwarz information criterion (divided by the sample size), and $Log-Lik$ is the log-likelihood value at the maximum. The sample size is equal to 1816.

Table 4.2: ML estimation results of AR-GARCH models for the NIKKEI and the SMI

	Normal	Student	skewed Student
	NIKKEI; SMI	NIKKEI ; SMI	NIKKEI ; SMI
μ	-0.004 ; 0.110 (0.030) ; (0.023)	-0.021 ; 0.123 (0.026) ; (0.020)	-0.031 ; 0.102 (0.028) ; (0.021)
ϕ	-0.014 ; 0.070 (0.025) ; (0.026)	-0.026 ; 0.039 (0.023) ; (0.025)	-0.028 ; 0.028 (0.023) ; (0.026)
ω	0.061 ; 0.147 (0.036) ; (0.066)	0.040 ; 0.063 (0.016) ; (0.024)	0.039 ; 0.058 (0.015) ; (0.022)
β	0.894 ; 0.731 (0.035) ; (0.058)	0.902 ; 0.810 (0.018) ; (0.049)	0.903 ; 0.817 (0.018) ; (0.047)
α	0.080 ; 0.141 (0.024) ; (0.028)	0.082 ; 0.136 (0.016) ; (0.035)	0.082 ; 0.134 (0.016) ; (0.033)
$\ln(\xi)$	0 ; 0	0 ; 0	-0.035 ; -0.101 (0.034) ; (0.034)
v	∞ ; ∞	5.950 ; 6.273 (0.845) ; (1.085)	5.895 ; 6.364 (0.825) ; (1.096)
Q_{20}	14.198 ; 12.271	14.29 ; 11.212	14.372 ; 11.836
Q_{20}^2	5.876 ; 1.377	6.617 ; 2.433	6.662 ; 2.454
P_{20}	47.225 ; 51.567 (0.000) ; (0.000)	13.837 ; 26.818 (0.793) ; (0.108)	15.843 ; 16.570 (0.667) ; (0.618)
SIC	3.562 ; 2.894	3.494 ; 2.788	3.498 ; 2.787
$Log-Lik$	-3214.1 ; -2607.9	-3148.9 ; -2507.6	-3148.4 ; -2503.2

Note: see Table 4.1.

-
- The AR(1)-GARCH(1,1) specification seems to be adequate for capturing the dynamics of the four series. Indeed, looking at the Box-Pierce statistics with 20 lags on the standardized residuals (Q_{20}) and the squared standardized residuals (Q_{20}^2), one cannot reject the assumption of lack of autocorrelation in the innovation process and its square (except perhaps for the CAC40 where the standardized residuals are still slightly serially correlated);
 - The estimated number of degrees of freedom ν is about 6 for the NASDAQ, NIKKEI and SMI and about 9 for the CAC40, which indicates that the returns are fat-tailed. Moreover, the differences between the likelihood of the normal and the Student densities are so big that there is little doubt that the latter should be preferred to the former (despite the fact that the LR test is presumably non-standard);
 - The estimated skewness parameter $\ln(\xi)$ is negative and different from 0 at conventional levels of significance for the NASDAQ and the SMI, while it is not different from 0 for the CAC40 and the NIKKEI. The distribution of returns of the NASDAQ and the SMI is therefore characterized by negative skewness, while the other series appear to be symmetrically distributed over the period under consideration. Notice however that since the skewed Student density has the symmetric Student density as a limiting case, it is also adequate for the CAC40 and the NIKKEI (resulting perhaps in a small loss of efficiency);
 - Using the Schwarz information criterion to discriminate between the three densities, one should select the skewed Student for the NASDAQ and the SMI and the Student for the others;
 - Finally and more importantly, the relevance of the skewed Student distribution is also confirmed by the Pearson goodness-of-fit statistics.² This test is in fact equivalent to an in-sample density forecast test, as proposed recently by Diebold, Gunther, and Tay (1998). While the normal and the Student distributions are clearly rejected for the NASDAQ (the p-values being very small), the skewed Student density seems to be supported (p-value = 0.87).

²Recall that the asymptotic distribution of $P(g)$ is bounded between a $\chi^2(g-1)$ and a $\chi^2(g-k-1)$ where g is the number of cells and k is the number of estimated parameters. Since our conclusions hold for both critical values, we report the significance levels relative to the first one.

Similarly, one can see that the skewed Student density is appropriate for modelling the SMI. Unsurprisingly, the normal density is rejected for the CAC40 and the NIKKEI while the Student and the skewed Student are not rejected at conventional levels of significance.

This example illustrates the potential usefulness of the skewed Student distribution in a univariate volatility model. The skewness parameters of the four series are different, but the numbers of degrees of freedom are almost identical for the NASDAQ, NIKKEI and SMI, while the innovations of the CAC40 seem to have less kurtosis. For modelling jointly the four series, it could therefore be useful to have a multivariate density that would allow for different skewness and perhaps different tail properties on each series.

4.3 Multivariate case

Consider a time series vector y_t , with k elements, $y_t = (y_{1t}, y_{2t}, \dots, y_{kt})'$. A multivariate dynamic regression model with time-varying means, variances and covariances for the components of y_t generally takes the form:

$$y_t = \mu_t + \Sigma_t^{1/2} z_t \quad (4.7)$$

$$\mu_t = C(\mu | \Omega_{t-1}) \quad (4.8)$$

$$\Sigma_t = \Sigma(\mu, \eta | \Omega_{t-1}) \quad (4.9)$$

where $z_t \in \mathfrak{R}^k$ is an *i.i.d.* random vector with zero mean and identity variance matrix. It follows that $E(y_t | \mu, \Omega_{t-1}) = \mu_t$ and $Var(y_t | \mu, \eta, \Omega_{t-1}) = \Sigma_t^{1/2} (\Sigma_t^{1/2})' = \Sigma_t$, i.e. μ_t is the conditional mean vector (of dimension $k \times 1$) and Σ_t the conditional variance matrix (of dimension $k \times k$).

Under the assumption of a correct specification of the conditional mean and variance matrix, the efficient estimation of the above model is obtained by the ML method, assuming z_t to be a *i.i.d.* with a correctly specified distribution that may depend upon a few unknown parameters. When the distribution of z_t is assumed to be the standard normal, the ML estimator obtained from the corresponding likelihood function is consistent even if the normality assumption is incorrect (see Bollerslev and Wooldridge, 1992). This well-known Gaussian QML procedure has the advantage of robustness with respect to the distributional assumption of the

model. The QML estimator relying on a normal distribution is, however, inefficient, with the degree of inefficiency increasing with the degree of departure from normality (see Engle and González-Rivera, 1991).

4.3.1 Multivariate symmetrical densities

Like in the univariate case, a natural candidate, apart from the normal density, is the multivariate Student density with at least two degrees of freedom v (in order to ensure the existence of second moments). It may be defined as

$$g(z_t|v) = \frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{v}{2}\right) [\pi(v-2)]^{\frac{k}{2}}} \left[1 + \frac{z_t' z_t}{v-2}\right]^{-\frac{k+v}{2}}, \quad (4.10)$$

where $\Gamma(\cdot)$ is the Gamma function. This density is denoted $ST(0, I_k, v)$.

The density function of y_t , easily derived from the density of z_t by using the transformation in Eq. (4.7), is given by

$$f(y_t | \mu, \eta, v, \Omega_{t-1}) = \frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{v}{2}\right) [\pi(v-2)]^{\frac{k}{2}}} |\Sigma_t|^{-\frac{1}{2}} \left[1 + \frac{(y_t - \mu_t)' \Sigma_t^{-1} (y_t - \mu_t)}{v-2}\right]^{-\frac{k+v}{2}}. \quad (4.11)$$

While non-Gaussian QML methods provide more efficient estimators than the Gaussian QML when the assumption made on the innovation process holds, it has the main disadvantage that unlike the Gaussian QML, it does not provide a consistent estimator when this assumption does not hold (see Newey and Steigerwald, 1997).

To overcome this problem, there is a need for skewed densities in the multivariate case. Such densities can be defined by introducing skewness in symmetric densities by means of new parameters, such that the symmetric density results as a particular case. In Section 4.3.2, we propose a simple and intuitive method to introduce skewness into a multivariate “symmetric” unimodal density (with zero mean and unit variance). Before that, we define the notion of symmetry that we rely on.

In the univariate case, the symmetry property corresponds to $g(x) = g(-x)$ assuming $g(x)$ is a unimodal probability density function and $E(x) = 0$. In the multivariate case, we use the following definition of symmetry of a standardized density $g(x)$:

Definition 2 (M-symmetry): *The unimodal density $g(x)$ defined on \mathfrak{R}^k , such that $E(x) = 0$, and $Var(x) = I_k$, is symmetrical if and only if for any x , $g(x) = g(Qx)$, for all diagonal matrices Q whose diagonal elements are equal to $+1$ or to -1 . If x is a random vector with a density satisfying this definition, we write*

$$x \sim M\text{-Sym}(0, I_k, g). \quad (4.12)$$

In the bivariate case, this definition means that

$$g(x_1, x_2) = g(-x_1, x_2) = g(x_1, -x_2) = g(-x_1, -x_2), \quad (4.13)$$

and in the trivariate case

$$\begin{aligned} g(x_1, x_2, x_3) &= g(-x_1, x_2, x_3) = g(-x_1, -x_2, x_3) = g(-x_1, -x_2, -x_3) \\ &= g(x_1, -x_2, x_3) = g(x_1, -x_2, -x_3) = g(x_1, x_2, -x_3) = g(-x_1, x_2, -x_3). \end{aligned} \quad (4.14)$$

Spherically symmetric (SS) densities, defined by the property that the density depends on x through $x'x$ only, i.e.

$$g(x) \propto k(x'x), \quad (4.15)$$

for an appropriate integrable positive function $k(\cdot)$, are M-symmetric. The most well known examples of SS-densities³ are the standard normal density and the standard Student density $ST(0, I_k, \nu)$. However, there exist other distributions that have the desired property while not being spherically symmetric. A large class is defined by

$$g(x) = \prod_{i=1}^k g_i(x_i), \quad (4.16)$$

where $g_i(\cdot)$, $\forall i$, is a univariate symmetric density (unimodal, with mean 0 and unit variance). If $g_i(\cdot)$ ($\forall i$) is standard normal, there is no difference between (4.16) and (4.15) with $g(\cdot) = N(0, I_k)$. Nevertheless, if $g_i(\cdot) = ST(0, 1, \nu)$ ($\forall i$) and $g(\cdot) = ST(0, I_k, \nu)$, there is a difference between (4.16) and (4.15) since the elements of (4.15) are not mutually independent whereas those of (4.16) are. Notice that both multivariate densities have the same univariate marginal densities.

³Johnson (1987), chapter 6, provides graphical illustrations of several bivariate SS-densities.

4.3.2 Multivariate skewed densities

Literature review

Jones (2000) has generalized the univariate skew- t density of Jones and Faddy (2000), briefly described at the end of Section 2.1, to the multivariate case. His multivariate skew- t density is such that each marginal is a univariate skew- t as defined by Jones and Faddy (2000). However, his multivariate density has necessarily positive covariances, and is therefore useless for a model such as defined by Eq. (4.7), where it is essential that $\text{Var}(z_t) = I_k$.

Mauleón and Perote (1999) use the bivariate Edgeworth-Sargan density for z_t in a bivariate constant correlation GARCH model, where each conditional variance is specified like in a univariate GARCH(1,1) model. The Edgeworth-Sargan density has as leading term a bivariate standard normal density, to which are added terms that create the non-normality (these terms involve Hermite polynomials in each of the marginal densities of the leading term). However, they use only a symmetrical version of their density, because they choose not to include odd-order terms in the expansion (such terms would induce asymmetry). Actually they include four even-order terms in the expansion on each element of z_t , under the motivation that these terms induce fatter tails than for the leading normal density. This appears to us to be a costly way, in term of the number of parameters, to introduce the possibility of having fat tails. A multivariate Student density requires just one extra parameter, with the drawback of constraining the same thickness of tails on each element of z_t , but this is easily extended by taking a product of independent Student densities in the spirit of Eq. (4.16) (the last solution would require 2 parameters instead of 8 in the bivariate case). Moreover, Mauleón and Perote (1999) report some difficulties in obtaining the convergence of the numerical maximization of the log-likelihood function based on their Edgeworth-Sargan density. At least for the time being, this does not seem to be a fruitful approach.

Another recent paper, by Branco and Dey (2000), introduces a general class of multivariate skew-elliptical distributions, and is therefore related to our work.⁴ Their work generalizes to the full class of elliptically contoured (EC) densities earlier results by Azzalini and Capitanio (1996), who have defined a multivariate

⁴Sahu, Dey, and Branco (2001) use the skew-elliptical density in Bayesian regression analysis, by assuming the error terms to have this kind of distribution, rather than a symmetrical distribution.

skew-normal distribution. Any EC-density is obtained by linear transformation of a SS-density: if z (of dimension $k \times 1$) is SS-distributed with density $g(z)$, μ is a vector of location parameters, and Ω is a $k \times k$ positive-definite symmetric scale matrix, then $x = \mu + \Omega^{1/2}z$ is elliptically contoured, which is denoted $x \sim EC(\mu, \Omega; g)$ (where g denotes the density of x). To obtain a skewed version of an EC-density, Branco and Dey (2000) start from $x^* \sim EC(\mu^*, \Omega^*; g^*)$, where $x^* = (x_0, x')'$ is a vector of $k + 1$ elements. They partition μ^* and Ω^* as x^* , i.e.

$$\mu^* = \begin{pmatrix} 0 \\ \mu \end{pmatrix}, \quad \Omega^* = \begin{pmatrix} 1 & \delta' \\ \delta & \Omega \end{pmatrix}, \quad (4.17)$$

where μ and δ are $k \times 1$ vectors, and Ω is a $k \times k$ matrix. Then they define the distribution of x conditional on $x_0 > 0$ to be the skew-elliptical distribution based on the density $g^*(\cdot)$, with parameters μ (location, or mean if it exists), Ω (scale matrix, or variance matrix if it exists), and δ (a vector of skewness parameters), i.e. $x \sim SKE(\mu, \Omega, \delta; g)$. They show that the density of this random vector (call it z) is given by

$$f(z) = 2g(z) G_*[\lambda'(z - \mu)], \quad (4.18)$$

where $g(\cdot)$ is the marginal density of x derived from the density of x^* (by properties of EC-distributions, it has the same functional form as g^*), $G_*(\cdot)$ is the (univariate) cdf of an $EC(0, 1; g_*)$, with g_* appropriately defined (essentially from the conditional density of x_0 given x), and

$$\lambda = \frac{\delta' \Omega^{-1}}{(1 - \delta' \Omega^{-1} \delta)^{1/2}}. \quad (4.19)$$

It is therefore clear that the parameters δ (a set of covariances) create the skewness. If they are all equal to 0, $G_*[\lambda'(z - \mu)] = G_*(0) = 1/2$, by symmetry of $EC(0, 1; g_*)$, and the density (4.18) becomes symmetrical. However, there is a constraint linking these skewness parameters, namely that $\delta' \Omega^{-1} \delta$ must be smaller than unity, see Eq. (4.19). This is a constraint that is likely to complicate inference. In the context of GARCH models with standardized innovations, Ω is an identity matrix (and $\mu = 0$), hence δ is a vector of correlation coefficients, and the constraint is that the sum of squared correlations is less than one. To what extent this constraint limits the degree of skewness is not known.⁵ Another drawback of this approach is that if

⁵If $k = 1$, the constraint is not limitative.

one wants to introduce some dynamics in the skewness parameters, the constraint would be different for each observation, which would complicate the estimation dramatically. We conclude on this class of skewed densities by saying that it seems an interesting, though seemingly more difficult to implement, alternative to the class of skewed densities that we propose below, and that more work is needed to compare the different classes of skewed densities

To accommodate both the skewness and kurtosis of six weekly rates of the European Monetary System (EMS) expressed in terms of the Deutsche mark, Vlaar and Palm (1993) propose to use a (Bernoulli) mixture of two multivariate normal densities (coupled with an MA(1)-GARCH(1,1) model with constant correlations, see Bollerslev, 1990).⁶ The size and the variance of the jumps are allowed to differ across currencies. However, to render the estimation feasible, they assume (and test) identical jump probability for all the series arguing that a stochastic shock leading to a jump is likely to simultaneously affect all of the currencies in the system. Even if this assumption is realistic for currencies that belong to the EMS, it is unrealistic for stock indexes, for instance. Moreover, even if this density is expressed in such a way that $E(z_t) = 0$, the covariance matrix of z_t is not an identity matrix in their specification. Another drawback of this density is that the parameters that govern the skewness and kurtosis have not a clear interpretation because for each margin the jump probability, the size and the variance of the jumps explain at the same time the variance, skewness and kurtosis in an highly non-linear way (see Vlaar and Palm, 1993 for more details). To conclude about this density, it suffers from a problem of non-identification of several parameters when the mixture is not relevant (for instance when the jump probability equals 0 or 1), which makes the testing procedures non-standard.

Finally, we cannot refrain from mentioning a class of multivariate densities that could be of interest: the so-called poly-t densities that contain the multivariate Student density as a particular case. Poly-t densities arise as posterior densities in Bayesian inference, see Drèze (1978), and can be heavily skewed, have fat tails and even be multimodal. However, more work is required to discover how the skewness of these densities depends on their parameters (see Richard and Tompa, 1980 for results on moments of poly-t densities).

⁶This density is a generalization of the Bernoulli-normal mixture presented in Section 2.2.2.

New skewed densities

We generalize to the multivariate case the method proposed by Fernández and Steel (1998) to construct a skewed density from a symmetrical one. Let us consider the k -dimensional random vector z^* defined by:

$$z^* = \lambda(\tau) |x|, \quad (4.20)$$

where

$$|x| = (|x_1|, \dots, |x_k|)', \quad (4.21)$$

and

$$x \sim M\text{-Sym}(0, I_k, g). \quad (4.22)$$

Moreover, $\lambda(\tau)$ is a $k \times k$ diagonal matrix defined by:

$$\lambda(\tau) = \tau \Xi - (I_k - \tau) \Xi^{-1}, \quad (4.23)$$

where

$$\begin{aligned} \tau &= \text{diag}(\tau_1, \dots, \tau_k), \text{ with } \tau_i \in \{0, 1\}, \\ \tau_i &\sim \text{Ber}\left(\frac{\xi_i^2}{1 + \xi_i^2}\right), \text{ with } \xi_i > 0, \\ \xi &= (\xi_1, \dots, \xi_k). \\ \Xi &= \text{diag}(\xi). \end{aligned}$$

$\text{Ber}\left(\frac{\xi_i^2}{1 + \xi_i^2}\right)$ denotes a Bernoulli distribution with probability of success $\frac{\xi_i^2}{1 + \xi_i^2}$. It is also assumed that the elements of τ are mutually independent.

For ease of exposition, we give the details of the derivation of the density of z^* in the bivariate case, before giving the general formula.

Bivariate case

We can write the density of z^* as a discrete mixture with respect to the distribution of τ :

$$\begin{aligned} f(z^*|\xi) &= \Pr(\tau_1 = 1, \tau_2 = 1)f(z^*|\xi, \tau_1 = 1, \tau_2 = 1) \\ &+ \Pr(\tau_1 = 1, \tau_2 = 0)f(z^*|\xi, \tau_1 = 1, \tau_2 = 0) \\ &+ \Pr(\tau_1 = 0, \tau_2 = 1)f(z^*|\xi, \tau_1 = 0, \tau_2 = 1) \\ &+ \Pr(\tau_1 = 0, \tau_2 = 0)f(z^*|\xi, \tau_1 = 0, \tau_2 = 0). \end{aligned} \quad (4.24)$$

By dividing the range of all possible values of $z^* \in \mathfrak{R}^2$ into the four quadrants, we can write the right hand side of Eq. (4.24) in terms of the original M-symmetric density $g(\cdot)$:

$$\begin{aligned}
 f(z^*|\xi) &= 2^2 \Pr(\tau_1 = 1, \tau_2 = 1) |\lambda(1, 1)|^{-1} g[\lambda(1, 1)^{-1} z^*] I_{(z_1^* \geq 0; z_2^* \geq 0)} \\
 &+ 2^2 \Pr(\tau_1 = 1, \tau_2 = 0) |\lambda(1, 0)|^{-1} g[\lambda(1, 0)^{-1} z^*] I_{(z_1^* \geq 0; z_2^* < 0)} \\
 &+ 2^2 \Pr(\tau_1 = 0, \tau_2 = 1) |\lambda(0, 1)|^{-1} g[\lambda(0, 1)^{-1} z^*] I_{(z_1^* < 0; z_2^* \geq 0)} \\
 &+ 2^2 \Pr(\tau_1 = 0, \tau_2 = 0) |\lambda(0, 0)|^{-1} g[\lambda(0, 0)^{-1} z^*] I_{(z_1^* < 0; z_2^* < 0)}, \quad (4.25)
 \end{aligned}$$

where e.g. $\lambda(1, 1)$ stands for $\lambda(\tau_1 = 1, \tau_2 = 1)$ and for instance $I_{(z_1^* \geq 0; z_2^* \geq 0)} = 1$ when $z_1^* \geq 0$ and $z_2^* \geq 0$, 0 otherwise. After some algebraic manipulations of (4.25) using (4.23) and the assumption of independence of τ_1 and τ_2 , we obtain:

$$\begin{aligned}
 f(z^*|\xi) &= 2^2 \frac{\xi_1}{1 + \xi_1^2} \frac{\xi_2}{1 + \xi_2^2} \left\{ g[\lambda(1, 1)^{-1} z^*] I_{(z_1^* \geq 0; z_2^* \geq 0)} \right. \\
 &+ g[\lambda(1, 0)^{-1} z^*] I_{(z_1^* \geq 0; z_2^* < 0)} + g[\lambda(0, 1)^{-1} z^*] I_{(z_1^* < 0; z_2^* \geq 0)} \\
 &\left. + g[\lambda(0, 0)^{-1} z^*] I_{(z_1^* < 0; z_2^* < 0)} \right\}, \quad (4.26)
 \end{aligned}$$

and finally,

$$f(z^*|\xi) = 2^2 \frac{\xi_1}{1 + \xi_1^2} \frac{\xi_2}{1 + \xi_2^2} g(\kappa^*), \quad (4.27)$$

where

$$\kappa^* = (\kappa_1^*, \kappa_2^*)' \quad (4.28)$$

$$\kappa_i = z_i^* \xi_i^{-I_i} \quad (i = 1, 2) \quad (4.29)$$

$$I_i = \begin{cases} 1 & \text{if } z_i^* \geq 0 \\ -1 & \text{if } z_i^* < 0. \end{cases}$$

Applying this procedure to the bivariate Student distribution given by Eq. (4.10) with $k = 2$ and x instead of z_t , i.e. $x \sim ST(0, I_2, v)$, yields a ‘‘bivariate skewed Student’’ density, in which both marginals have different asymmetry parameters, ξ_1 and ξ_2 .

Multivariate case

It is straightforward to show that for any dimension k ,

$$f(z^*|\xi) = 2^k \left(\prod_{i=1}^k \frac{\xi_i}{1 + \xi_i^2} \right) g(\kappa^*), \quad (4.30)$$

where κ^* is given in Eq. (4.28)-(4.29) for the bivariate case and is easily extended to the multivariate case. Recall that for each margin z_i^* , ξ_i has a clear interpretation since ξ_i^2 is equal to the ratio of probability masses above and below the mode. Remark also that when $k = 1$, one recovers the family of skewed densities proposed by Fernández and Steel (1998).

Moments

A convenient property of this new family of skewed densities is that the marginal moments are obtained by the same method and actually correspond to the same formulas as in the univariate case. The r -th order moment of $f(z^*|\xi)$ exists if the r -th order moment of $g(\cdot)$ exists. In particular,

$$E(z_i^{*r}|\xi) = M_{i,r} \frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}} \quad (4.31)$$

where

$$M_{i,r} = \int_0^\infty 2u^r g_i(u) du, \quad (4.32)$$

and $g_i(\cdot)$ is the marginal of x_i extracted from $g(x)$, while $M_{i,r}$ is the r -th order moment of $g_i(\cdot)$ truncated to the positive real values. Provided that these quantities are finite, we can obtain $E(z_i^*|\xi_i)$, $Var(z_i^*|\xi_i)$, $Sk(z_i^*|\xi_i)$ and $Ku(z_i^*|\xi_i)$ using the formulas given in Eq. (2.22-2.25), where $Sk(\cdot)$ and $Ku(\cdot)$ denote the skewness and kurtosis coefficients, respectively.

Finally, it is obvious that the elements of z^* are uncorrelated (since those of x are uncorrelated by assumption), so that it is easy to transform z^* so as to have any specified correlation matrix.

Standardized skewed densities

The main drawback of the skewed density defined by Eq. (4.30) is that it is not centered on 0 and the covariance matrix is a function of ξ (and of v if $g(\cdot)$

is a multivariate Student density). As in the univariate case, one can solve this problem by standardizing z^* .

Let us consider the following random vector:

$$z = (z^* - m) ./ s \quad (4.33)$$

where $m = (m_1, \dots, m_k)$ and $s = (s_1, \dots, s_k)$ are the vectors of unconditional means and standard deviations of z^* , and $./$ means element by element division. The above transformation amounts to standardize each component of z^* .

Following Lambert and Laurent (2001), if $g_i(.|v)$ is a standardized Student density (with $v > 2$),

$$m_i = \frac{\Gamma\left(\frac{v-1}{2}\right) \sqrt{v-2}}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right)} \left(\xi_i - \frac{1}{\xi_i} \right) \quad (4.34)$$

and

$$s_i^2 = \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right) - m_i^2. \quad (4.35)$$

Definition 3 *If (i) z is defined by Eq. (4.33-4.35), and (ii) z^* has a density given by Eq. (4.30), where $g(x)$ is the Student density given by Eq. (4.10), then z is said to be distributed as (multivariate) standardized skewed Student with asymmetry parameters $\xi = (\xi_1, \dots, \xi_k)$, and degrees of freedom $v(> 2)$. This is denoted $z \sim SKST(0, I_k, \xi, v)$. The density of z is given by*

$$f(z|\xi, v) = \left(\frac{2}{\sqrt{\pi}} \right)^k \left(\prod_{i=1}^k \frac{\xi_i s_i}{1 + \xi_i^2} \right) \frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma\left(\frac{v}{2}\right) (v-2)^{\frac{k}{2}}} \left(1 + \frac{\kappa' \kappa}{v-2} \right)^{-\frac{k+v}{2}}. \quad (4.36)$$

where

$$\kappa = (\kappa_1, \dots, \kappa_k)' \quad (4.37)$$

$$\kappa_i = (s_i z_i + m_i) \xi_i^{-I_i} \quad (4.38)$$

$$I_i = \begin{cases} 1 & \text{if } z_i \geq -\frac{m_i}{s_i} \\ -1 & \text{if } z_i < -\frac{m_i}{s_i}. \end{cases}$$

By construction, $E(z) = 0$ and $\text{Var}(z) = I_k$. If $\xi = I_k$, the $SKST(0, I_k, \xi, v)$ density becomes the $ST(0, I_k, v)$ one, i.e. the symmetric Student density.

Assuming that y_t is specified as in Eq. (4.7) and $z_t \sim SKST(0, I_k, \xi, v)$, the density of y_t is straightforwardly obtained (see how Eq. (4.11) is obtained from Eq. (4.10)).

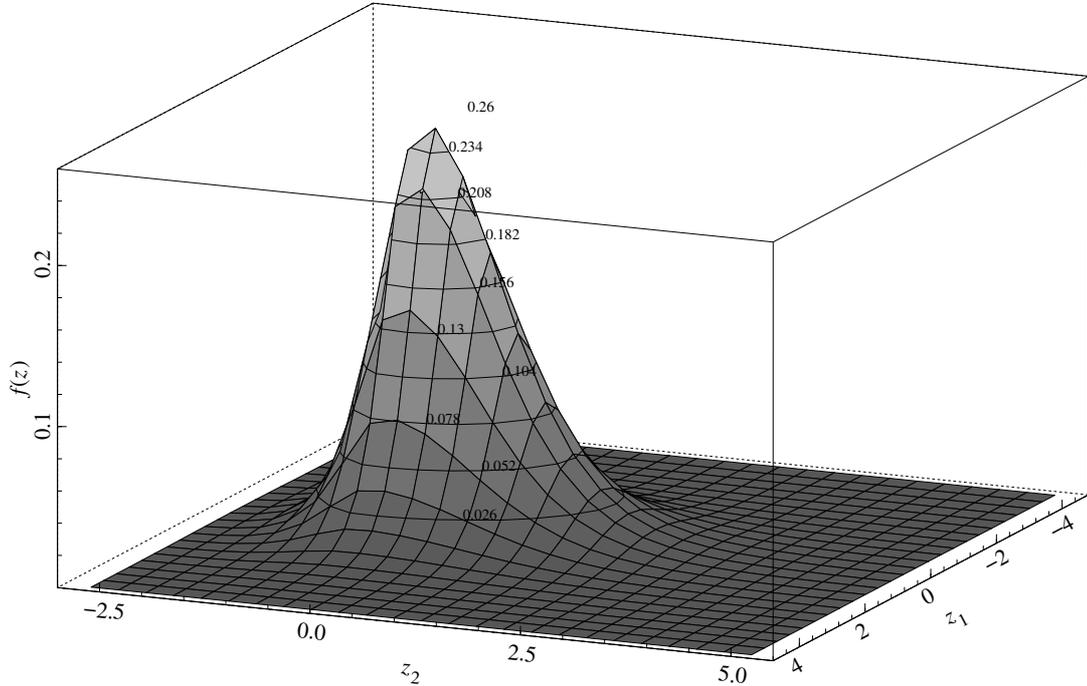


Figure 4.1: Graph of the $SKST(0, I_2, (1, 1.3), 6)$ density

To illustrate, Figure 4.1 shows a graph of the $SKST(0, I_2, \xi, 6)$ density with $\xi_1 = 1$, $\xi_2 = 1.3$, and the Panel A of Figure 4.2 shows its contours.

The first graph is oriented to show the asymmetry to the right along the z_2 -axis, while the density is symmetric in the direction of the first coordinate (z_1). The contours show more clearly the skewness properties of the density in the direction of z_2 , and its symmetry in the direction of z_1 . One also clearly sees that the mode is not centered in zero (unlike in the non-standardized version).

4.3.3 Simulation

In order to assess the practical applicability of the ML method to the estimation of the skewed Student distribution, we present the results of a small simulation study. It is not our intention to provide a comprehensive Monte Carlo study. Our results, however, provide some evidence on the properties of the MLE when a multivariate standardized skewed Student distribution is assumed for the innovations. Consider the bivariate case with $y_t = (y_{1,t}, y_{2,t})$. The data generating process is given by

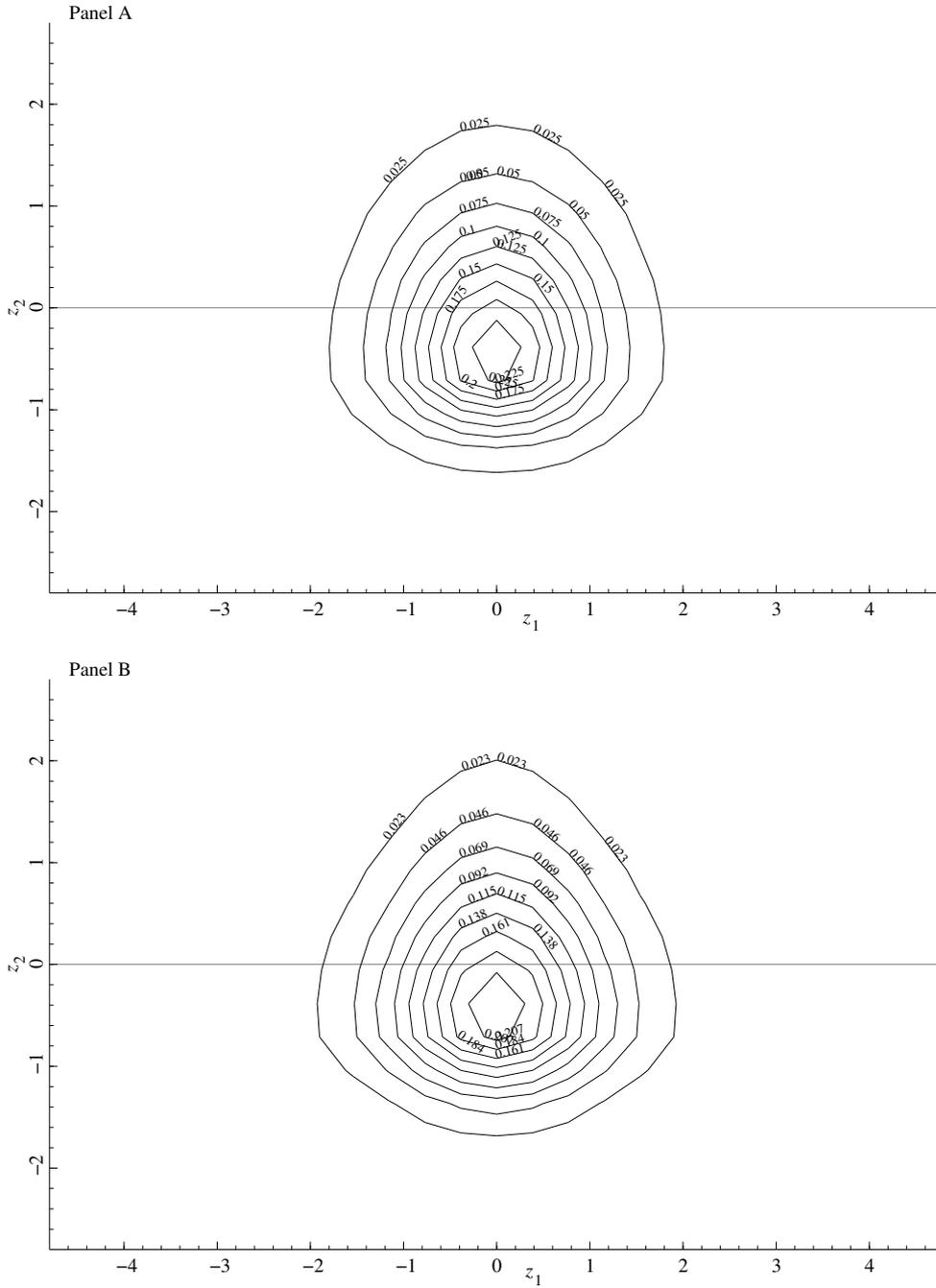


Figure 4.2: Panel A refers to the contours of the bivariate $SKST(0, I_2, (1, 1.3), 6)$ density illustrated in Figure 4.1. Panel B refers to the contours of a $SKST-IC(0, I_k, (1, 1.3), (6, 6))$ (see Section 4.3.4)

Eq. (4.7), with $\mu_t = \mu = (0, 0)'$, $\Sigma_t = \Sigma$ a correlation matrix with off-diagonal element equal to -0.2, $z_t \sim SKST(0, I_2, \xi, \nu)$, where $(\ln(\xi_1), \ln(\xi_2)) = (0.2, -0.2)$ and $\nu = 8$. This configuration implies that the innovations are skewed (with skewness amounting to 0.53 and -0.53 respectively for z_1 and z_2) and have fat-tails (the kurtosis equals 4.80 for both). The sample size is set to 20,000. Table 4.3 reports the DGP as well as the estimation results under three assumptions for the innovations: normal, Student and (standardized) skewed Student densities.

From Table 4.3, it is clear that the ML method, under the correct density (i.e. the skewed Student, see column 5), works reasonably well in the sense that the estimates are very close to the “true” values. Table 4.3 also illustrates the well known result of Weiss (1986) and Bollerslev and Wooldridge (1992) that (if the mean and the variance are specified correctly) the Gaussian QML estimator is consistent (but inefficient). Moreover, this table also confirms the result of Newey and Steigerwald (1997) that the QML estimator with a Student pseudo-likelihood is inconsistent when innovations are skewed. One can see that μ is rather strongly biased under the Student density, whereas the other parameters seem less affected in this experiment. To check the model adequacy, we use the same diagnostic tools (on each innovation separately)⁷ as in the empirical illustration of Section 2.2. These statistics suggest that the normal and Student densities are not appropriate, while the skewed Student is. Notice that rejecting that the margins are not correctly specified is sufficient to reject the assumption that the whole density is not appropriate. However, the converse is obviously not true. Indeed, accepting that the margins are well specified is necessary to accept that the whole density is appropriate, but it is not sufficient.

4.3.4 Multivariate skewed densities with independent components

An obvious variation with respect to the previous class of multivariate skewed densities is obtained by starting from the product of k independent $ST(0, 1, \nu_i)$ and applying to it the transformation defined by Eq. (4.20)-(4.21)-(4.23).

Definition 4 *If (i) z is defined by Eq. (4.33-4.35), and (ii) where ν is simply*

⁷Multivariate tests of adequacy of a distribution are more appropriate tools but are usually difficult to implement. This is the reason why we use simple diagnostic tools, which should at least help to detect a major misspecification.

Table 4.3: QML estimation results of simple skewed Student DGP

	DGP	Normal	Student	Skewed Student
μ_1	0.0	-0.001 (0.007)	-0.037 (0.007)	-0.000 (0.007)
μ_2	0.0	0.004 (0.007)	0.046 (0.007)	0.003 (0.007)
σ_1^2	1.0	0.985 (0.014)	0.982 (0.012)	0.992 (0.012)
σ_2^2	1.0	0.994 (0.013)	0.989 (0.012)	0.998 (0.012)
ρ	0.2	-0.226 (0.008)	-0.219 (0.007)	-0.213 (0.007)
$\ln(\xi_1)$	0.2	-	-	0.184 (0.010)
$\ln(\xi_2)$	-0.2	-	-	-0.194 (0.010)
v	8.0	-	7.903 (0.284)	8.316 (0.306)
Q_{20} and $Q_{20}^2(\hat{z}_1)$	-	14.791 ; 18.893	14.763 ; 17.884	14.743; 17.275
Q_{20} and $Q_{20}^2(\hat{z}_2)$	-	21.942 ; 14.492	21.852 ; 13.826	21.773; 9.936
$P_{40}(\hat{z}_1)$	-	475.768 (0.000)	316.240 (0.000)	34.504 (0.675)
$P_{40}(\hat{z}_2)$	-	585.384 (0.000)	355.068 (0.000)	30.496 (0.833)

DGP: $y_t = \mu + \Sigma^{1/2}z_t$, $t = 1, \dots, 20,000$, with $\mu = (\mu_1, \mu_2)'$, z_t *i.i.d.* $\sim SKST(0, I_2, \xi, v)$ as in Eq. (4.36), with $\xi = (\xi_1, \xi_2)$; σ_i^2 is the variance of y_i ($i = 1, 2$), and ρ is the correlation coefficient between y_1 and y_2 . The last four columns report the ML estimates (with the robust standard errors underneath in parentheses) of the parameters of the model corresponding to the DGP with different assumptions on the distribution of z_t . The column headed “Normal” corresponds to $z_t \sim N(0, I_2)$, “Student” to $z_t \sim ST(0, I_2, v)$ as in Eq. (4.10), “Skewed Student” to $z_t \sim SKST(0, I_2, [\xi_1, \xi_2], v)$. $Q_{20}(\hat{z}_i)$ and $Q_{20}^2(\hat{z}_i^2)$ are the Box-Pierce statistics of order 20 on the innovations \hat{z}_i and their squares. $P_{40}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 40 cells) with the associated p-value beside (see footnote 2). \hat{z} is given by $\hat{\Sigma}^{-1/2}(y_t - \hat{\mu})$, where $\hat{\Sigma}$ and $\hat{\mu}$ are obtained by replacing the parameters by their estimates in the corresponding formulas and $\hat{\Sigma}^{-1/2}$ is obtained from the spectral decomposition of $\hat{\Sigma}$.

replaced by v_i , z^* has a density given by Eq. (4.16), where $g_i(x)$ is the Student density given by Eq. (2.13), then z is said to be distributed as a (multivariate) skewed density with independent Student components, with asymmetry parameters $\xi = (\xi_1, \dots, \xi_k)$, and degrees of freedom $v = (v_1, \dots, v_k)$ (with $v_i > 2$). This is denoted $z \sim SKST-IC(0, I_k, \xi, v)$. The density of z is given by:

$$f(z|\xi, v) = \left(\frac{2}{\sqrt{\pi}}\right)^k \left[\prod_{i=1}^k \frac{\xi_i s_i}{1 + \xi_i^2} \frac{\Gamma(\frac{v_i+1}{2})}{\Gamma(\frac{v_i}{2}) \sqrt{v_i - 2}} \left(1 + \frac{\kappa_i^2}{v_i - 2}\right)^{-\frac{1+v_i}{2}} \right], \quad (4.39)$$

where κ_i is defined in Eq. (4.38).

Note that Eq. (4.39) is obtained equivalently by taking the product of k independent $SKST(0, 1, \xi_i, v_i)$. The main advantage of (4.39) with respect to (4.36) is that it enables a different tail behavior for each marginal, at the cost of introducing $k-1$ additional parameters. However, nothing prevents to constrain several degrees of freedom parameters to be equal. If all the degrees of freedom parameters v_i are equal to the degrees of freedom v of (4.36), the densities (4.39) and (4.36) have exactly the same marginal moments. The fact that the components of (4.36) are not independent implies that its cross-moments of order 4 or higher are functions of a common single parameter v and are thus less flexible than those of (4.39).

To illustrate, Panel B of Figure 4.2 shows the contours of the bivariate skewed density with independent Student components whose parameters are $\xi_1 = 1$, $\xi_2 = 1.3$, $v_1 = v_2 = 6$. One can notice the difference with respect to the contours of the Panel A of the same figure, which corresponds to the skewed Student with non-independent margins. In Panel B, the contours look like less “elliptic” than in Figure Panel A (see also the graphs in Johnson, 1987, Chapter 6, for the symmetric versions of these densities).

4.4 Empirical application

In this section, we jointly model the four series already used in the univariate application. The specification used to model the first two conditional moments is the time-varying correlation GARCH model (TVC-GARCH) proposed by Tse and Tsui (1998), with first-order ARMA dynamics in the conditional variances and the

conditional correlation, and an AR(1) equation for each conditional mean.⁸ This AR(1)-TVC(1,1)-GARCH(1,1) model is defined as follows:

$$y_t = \mu_t + \Sigma_t^{1/2} z_t \quad (4.40)$$

$$\mu_t = (\mu_{1,t}, \dots, \mu_{4,t})', \quad z_t = (z_{1,t}, \dots, z_{4,t})' \quad (4.41)$$

$$\mu_{i,t} = \mu_i + \phi_i(y_{i,t-1} - \mu_i) \quad (i = 1, \dots, 4) \quad (4.42)$$

$$\Sigma_t = D_t \Gamma_t D_t \quad (4.43)$$

$$D_t = \text{diag}(\sigma_{1,t}, \dots, \sigma_{4,t}) \quad (4.44)$$

$$\sigma_{i,t}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \varepsilon_{i,t-1}^2 \quad (i = 1, \dots, 4) \quad (4.45)$$

$$\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{4,t})' = y_t - \mu_t \quad (4.46)$$

$$\Gamma_t = (1 - \theta_1 - \theta_2)\Gamma + \theta_1 \Gamma_{t-1} + \theta_2 \Psi_{t-1} \quad (4.47)$$

$$\Gamma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{23} & \rho_{34} & 1 \end{pmatrix} \quad (4.48)$$

$$\Psi_{t-1} = B_{t-1}^{-1} E_{t-1} E_{t-1}' B_{t-1}^{-1} \quad (4.49)$$

$$B_{t-1} = \text{diag} \left(\sum_{h=1}^m \varepsilon_{1,t-h}^2, \dots, \sum_{h=1}^m \varepsilon_{4,t-h}^2 \right)^{1/2} \quad (4.50)$$

$$E_{t-1} = (\varepsilon_{t-1}, \dots, \varepsilon_{t-m}) \quad (4.51)$$

$$\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{4,t})' = D_t^{-1} \varepsilon_t, \quad (4.52)$$

where $\mu_i, \phi_i, \omega_i, \beta_i, \alpha_i$ ($i = 1, \dots, 4$), ρ_{ij} ($1 \leq i < j \leq 4$), and θ_1, θ_2 are parameters to be estimated.⁹ Ψ_{t-1} is thus the sample correlation matrix of $\{\varepsilon_{t-1}, \dots, \varepsilon_{t-m}\}$. Since $\Psi_{t-1} = 1$ if $m = 1$, we must take $m \geq 4$ to have a non-trivial correlation. In this application, we set $m = 4$. Note that the TVC-MGARCH model nests the constant correlation GARCH model of Bollerslev (1990). Therefore, we can test $\theta_1 = \theta_2 = 0$ to check whether the constant correlation assumption is appropriate.

The estimation results of this model are gathered in Tables 4.4 and 4.5. A QML estimation procedure has been done with four different likelihoods: normal and Student in Table 4.4, skewed Student and skewed density with independent Student components in Table 4.5.

⁸We implicitly assume that there is no Granger causality between the four series. A natural extension would be to estimate a VAR model for the mean equation and test these restrictions.

⁹The parameters θ_1 and θ_2 are assumed to be nonnegative with the additional constraint that $\theta_1 + \theta_2 < 1$.

Table 4.4: ML estimation results of AR-TVC-GARCH model: normal and Student distributions

	Normal				Student			
	CAC40	NASDAQ	NIKKEI	SMI	CAC40	NASDAQ	NIKKEI	SMI
μ_i	0.089 (0.028)	0.130 (0.025)	0.014 (0.031)	0.128 (0.025)	0.087 (0.026)	0.139 (0.022)	0.003 (0.028)	0.136 (0.021)
ϕ_i	0.014 (0.022)	0.092 (0.026)	0.024 (0.025)	0.085 (0.023)	0.017 (0.020)	0.103 (0.024)	0.012 (0.023)	0.064 (0.021)
ω_i	0.053 (0.030)	0.087 (0.033)	0.052 (0.027)	0.103 (0.069)	0.049 (0.024)	0.045 (0.029)	0.037 (0.014)	0.043 (0.022)
β_i	0.922 (0.032)	0.782 (0.058)	0.906 (0.025)	0.822 (0.089)	0.928 (0.025)	0.866 (0.062)	0.909 (0.015)	0.885 (0.044)
α_i	0.042 (0.013)	0.142 (0.040)	0.073 (0.017)	0.083 (0.035)	0.039 (0.011)	0.090 (0.036)	0.077 (0.013)	0.070 (0.023)
ρ_{ij}								
CAC40	1	0.383	0.374	0.749	1	0.286	0.234	0.663
NASDAQ	(0.103)	1	0.219	0.397	(0.038)	1	0.122	0.287
NIKKEI	(0.111)	(0.088)	1	0.383	(0.038)	(0.037)	1	0.247
SMI	(0.069)	(0.087)	(0.117)	1	(0.027)	(0.038)	(0.039)	1
θ_1		0.992 (0.005)				0.964 (0.033)		
θ_2		0.004 (0.002)				0.013 (0.007)		
$\ln(\xi_i)$		0				0		
v		∞				7.664 (0.680)		
$Q_{20}(\hat{z}_i)$	24.999	18.392	15.014	9.190	24.742	16.953	13.601	7.585
$Q_{20}(\hat{z}_i^2)$	23.753	6.698	7.998	4.302	24.244	11.814	8.133	4.197
$P_{20}(\hat{z}_i)$	26.708 (0.111)	79.909 (0.000)	45.661 (0.000)	52.074 (0.000)	10.663 (0.934)	43.809 (0.001)	17.319 (0.568)	16.548 (0.620)
SIC		11.726				11.478		
$Log-Lik$		-10544.3				-10315.2		

Each column reports the ML estimates of the model defined by Eq. (4.40)-(4.52), with robust standard errors underneath in parentheses. The column headed “Normal” corresponds to $z_t \sim N(0, I_4)$ and “Student” to $z_t \sim ST(0, I_4, \nu)$ as in Eq. (4.10). In both cases z_t is an *i.i.d.* process. $Q_{20}(\hat{z}_i)$ is the Box-Pierce statistic of order 20 on the standardized residuals \hat{z}_i , $Q_{20}(\hat{z}_i^2)$ is the same for their squares, $P_{20}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 20 cells) with the associated unadjusted p-value beside. SIC is the Schwarz information criterion (divided by the sample size $T = 1816$), and $Log-Lik$ is the log-likelihood value at the maximum.

Table 4.5: ML estimation results of AR-TVC-GARCH model: skewed Student and skewed Student with IC distributions

	Skewed Student				IC Skewed Student			
	CAC40	NASDAQ	NIKKEI	SMI	CAC40	NASDAQ	NIKKEI	SMI
μ_i	0.085 (0.027)	0.103 (0.023)	-0.002 (0.029)	0.119 (0.022)	0.079 (0.028)	0.111 (0.023)	-0.014 (0.029)	0.116 (0.023)
ϕ_i	0.015 (0.020)	0.081 (0.024)	0.011 (0.023)	0.058 (0.022)	0.011 (0.021)	0.075 (0.024)	0.005 (0.023)	0.060 (0.022)
ω_i	0.049 (0.024)	0.043 (0.027)	0.036 (0.014)	0.043 (0.022)	0.050 (0.029)	0.050 (0.024)	0.036 (0.014)	0.053 (0.028)
β_i	0.928 (0.025)	0.863 (0.057)	0.908 (0.014)	0.884 (0.043)	0.923 (0.032)	0.841 (0.050)	0.908 (0.016)	0.860 (0.054)
α_i	0.039 (0.011)	0.095 (0.034)	0.077 (0.013)	0.071 (0.023)	0.043 (0.014)	0.114 (0.032)	0.080 (0.014)	0.087 (0.030)
ρ_{ij}								
CAC40	1	0.288	0.234	0.661	1	0.311	0.272	0.679
NASDAQ	(0.037)	1	0.118	0.286	(0.049)	1	0.145	0.314
NIKKEI	(0.038)	(0.037)	1	0.245	(0.050)	(0.044)	1	0.280
SMI	(0.027)	(0.037)	(0.039)	1	(0.038)	(0.047)	(0.051)	1
θ_1		0.961 (0.037)				0.973 (0.032)		
θ_2		0.013 (0.007)				0.010 (0.007)		
$\ln(\xi_i)$	0.035 (0.034)	-0.186 (0.037)	-0.013 (0.036)	-0.085 (0.036)	0.025 (0.034)	-0.172 (0.037)	-0.016 (0.036)	-0.076 (0.037)
v/v_i		7.757 (0.696)			10.339 (2.172)	6.159 (0.834)	6.266 (0.906)	6.479 (1.095)
$Q_{20}(\hat{z}_i)$	24.825	20.409	13.552	7.657	25.182	21.561	12.874	7.437
$Q_{20}(\hat{z}_i^2)$	24.415	11.005	8.138	4.170	23.810	9.820	8.432	4.211
$P_{20}(\hat{z}_i)$	11.435 (0.908)	18.730 (0.474)	16.989 (0.590)	18.906 (0.462)	10.708 (0.934)	17.121 (0.581)	22.741 (0.248)	14.829 (0.733)
SIC		11.473				11.515		
$Log-Lik$		-10296.1				-10322.4		

Each column reports the ML estimates of the model defined by Eq. (4.40)-(4.52). The column headed ‘‘Skewed Student’’ corresponds to $z_t \sim SKST(0, I_4, \xi, v)$ as in Eq. (4.36), and ‘‘IC Skewed Student’’ to $z_t \sim$ Eq. (4.39) (with $k = 4$). In both cases z_t is an *i.i.d.* process. $Q_{20}(\hat{z}_i)$ is the Box-Pierce statistic of order 20 on the standardized residuals \hat{z}_i , $Q_{20}(\hat{z}_i^2)$ is the same for their squares, $P_{20}(\hat{z}_i)$ is the Pearson goodness-of-fit statistic (using 20 cells) with the associated unadjusted p-value beside. SIC is the Schwarz information criterion (divided by the sample size $T = 1816$), and $Log-Lik$ is the log-likelihood value at the maximum.

The results are in line with those obtained in the univariate case. The AR(1)-TVC(1,1)-MGARCH(1,1) specification seems adequate in describing the dynamics of the series, witness the small values of the Box-Pierce statistics of order 20 on the residuals and their squares, $Q_{20}(\hat{z}_i)$ and $Q_{20}(\hat{z}_i^2)$ respectively. The residual vector $\hat{z}_t = (\hat{z}_{i,t}, \dots, \hat{z}_{4,t})$ is defined as:

$$\hat{z}_t = \hat{\Sigma}_t^{-1/2}(y_t - \hat{\mu}_t), \quad (4.53)$$

where $\hat{\Sigma}_t$ and $\hat{\mu}_t$ are obtained by replacing the parameters by their estimates in the model formulas. $\hat{\Sigma}_t^{-1/2}$ has been obtained from the spectral decomposition of $\hat{\Sigma}_t$ (alternatively, a Cholesky factorization can be used).

A time-varying and very persistent correlation between the series is strongly supported if one looks at the estimates of θ_1 and θ_2 and the corresponding standard errors. On the first hand this justifies the use of a time-varying correlation specification and on the other hand the use of a multivariate model (comparing the sum of the univariate log-likelihoods with the corresponding multivariate likelihood, one can see that the multivariate approach increases the likelihood by more than 600 in all cases). Note that to facilitate the reading of the results concerning the unconditional correlation parameters (the matrix Γ), they are reported as in a 4 by 4 matrix. The upper triangle part of the matrix gives the estimated parameters while the lower triangle matrix (below the diagonal of ones) gives the associated standard errors. For instance, the estimated unconditional correlation between the CAC40 and the NIKKEI ($\hat{\rho}_{13}$) obtained with a Gaussian QML equals 0.374, with standard error 0.111.

It is clear from the estimation results reported in Table 4.4 that, apart from the dynamics in the first two conditional moments, the dominating feature of the four series is their fat-tail property. Indeed, the Student density increases the log-likelihood value by about 230 for only one additional parameter. Note that when comparing the standard errors related to the unconditional correlation parameters one can see that they are slightly reduced when switching from a Gaussian to a Student density. The normality assumption is also clearly rejected by the Pearson goodness-of-fit statistics (with very small p-values).¹⁰ As in the univariate case, the Student density is clearly rejected for the NASDAQ (the p-value of the Pearson

¹⁰The normality assumption is less questioned for the CAC40. This is in line with the result obtained in the univariate analysis.

goodness-of-fit statistics being equal to 0.001).

This is confirmed by the results concerning the skewed Student density (see Table 4.5). First, comparing the log-likelihood values and the information criterion values suggests that this density outperforms the symmetric Student (the log-likelihood is increased by about 19 for 4 additional parameters). Second, the Pearson goodness-of-fit statistics suggest that the skewed Student is adequate in capturing the skewness of the NASDAQ and in general that all the marginals are well described by our model specification.

The last part of Table 4.5 gives the results for the skewed density with independent Student components (see Section 4.3.4). Recall that unlike the skewed Student, this density has different degrees of freedom. The results suggest that the v_i are about 6 for the last three series (the NASDAQ, NIKKEI and SMI) and are not statistically different. Even if the number of degrees of freedom of the CAC40 is higher (about 10) the precision of this estimator is even worse and one can hardly distinguish it from the other. Note that one cannot use a LR test to discriminate between the skewed Student and the skewed Student with independent components since the models are not nested. Finally, looking at the Pearson goodness-of-fit statistics one cannot reject the assumption that this last density is also adequate for modelling the excess skewness and kurtosis observed on the four marginals.

To assess the irrelevance of the normal density and the adequacy of the skewed Student density, Figures 4.3 and 4.4 plot the histogram of the probability integral transform $\hat{\zeta}_i = \int_{-\infty}^{\hat{z}_i} f_i(t)dt$ with the 95% confidence bands.

Under weak conditions (see Diebold, Gunther, and Tay, 1998), the adequacy of a density implies that the sequence of ζ_i is independent and identically uniformly distributed on the unit interval. Departure from uniformity is directly observable in the Gaussian case for the NASDAQ, NIKKEI and SMI. On the other hand, one cannot reject the assumption that the probability integral transforms of the skewed Student density are uniformly distributed.¹¹

¹¹Confidence intervals for the ζ_i -histogram can be obtained by using the properties of the histogram under the null hypothesis of uniformity.

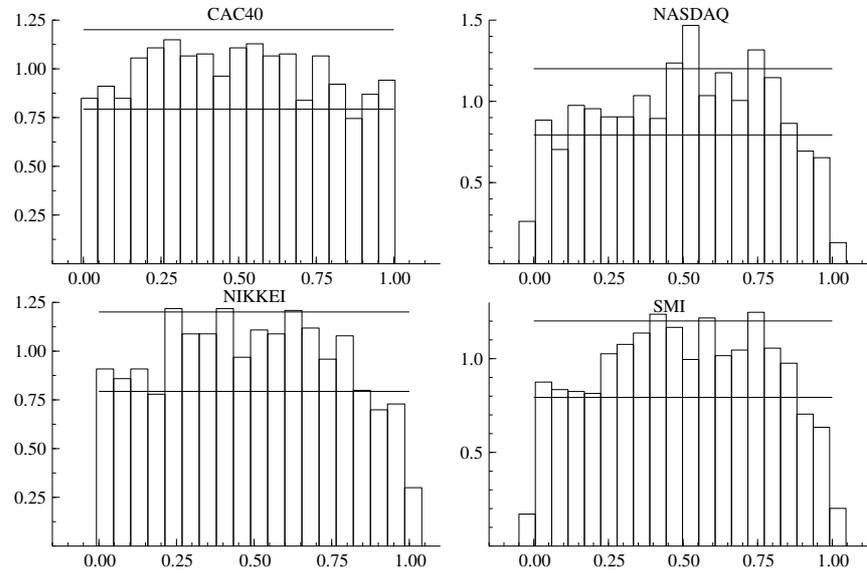


Figure 4.3: Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a normal likelihood (with 20 cells).

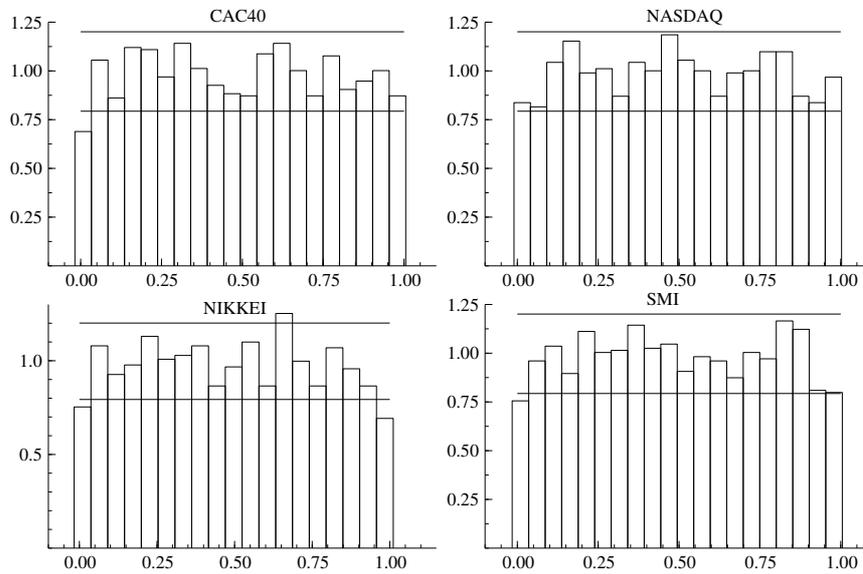


Figure 4.4: Histogram of the Probability Integral Transform of the CAC40, NASDAQ, NIKKEI and SMI innovations with a skewed Student likelihood (with 20 cells).

4.5 Conclusion

It is broadly accepted that high-frequency financial time series are heteroscedastic, fat-tailed and volatilities are related over time across assets and markets. To accommodate these stylized facts in a parametric framework a natural approach would be to rely on a multivariate GARCH or SV specification coupled with a Student density.

However, most asset returns are also skewed, which invalidates the choice of this density (it would lead to inconsistent estimates). To overcome this problem, we propose a practical and flexible method to introduce skewness in a wide class of multivariate symmetric distributions. By introducing a vector of skewness parameters, the new distributions bring additional flexibility for modelling time series of asset returns with multivariate volatility models. Applying the procedure to the multivariate Student density leads to a “multivariate skewed Student” density, in which each marginal has a different asymmetry coefficient. An easy variant provides a multivariate skewed density that can have different tail properties on each coordinate. These densities are found to outperform their symmetric competitors (the multivariate normal and Student) for modelling four daily stock market indexes, and therefore are of great potential interest for the empirical modelling of several asset returns together.

In the application, we have used a very simple specification for the first two conditional moments. First, the conditional means are assumed to follow an AR(1) and thus we implicitly assume that there is no Granger causality between the four series. To test the relevance of this restriction, one should estimate a Vector AR (VAR) model. Second, the conditional variances are estimated independently, in the sense that variances depend only on own past squared errors and on own past variances while correlations depend uniquely on the own cross-products of errors and on own past correlations. This model is thus not suited for testing causality or co-persistence in variance. Alternative specifications of the conditional covariance matrix may be more appropriate (see Bauwens, Laurent, and Rombouts, 2002 for a recent survey of multivariate GARCH models and their application in finance). Note also that we have shown in Chapter 2 that part of the unconditional asymmetry observed on daily stock returns is probably due to the so called leverage effect. Then, a natural extension of this chapter could be to use an APARCH

specification for the conditional variances.

Additional empirical studies based on these flexible distributions should be carried out to explore deeply the skewness and kurtosis properties of asset returns, including the co-skewness and co-kurtosis aspects in a multivariate framework (see Hafner, 2001).

Another potential area of application of the new densities is in Bayesian inference, for the design of simulators for Monte-Carlo integration of posterior densities that are characterized by different skewness and tail properties in different directions of the parameter space. In this respect, some of the densities we have proposed are related to the split-Student importance function proposed by Geweke (1989). This is obviously a different research topic, that we leave for further work.

Finally, a natural extension of this paper would be to generalize the GARCH specification to higher moments. Indeed, in a univariate framework Hansen (1994), introduces dynamics through the 3rd and 4th order moments by conditioning the asymmetry and fat-tail parameters on past errors and their square. In the same spirit, Harvey and Siddique (1999) and Lambert and Laurent (2000) provide alternative specifications to introduce dynamics in higher order moments. Such an extension seems feasible for the new family of skewed densities proposed in this chapter, which is less obvious for instance for the EC-density of Branco and Dey (2000).

To conclude this chapter and in the same time the first part of the thesis, this new family of multivariate skewed densities and in particular the multivariate skewed Student density seems to be a promising specification to accommodate both the high kurtosis and the skewness inherent in most asset returns.

In the second part of the thesis, we would like to investigate some economic implications of the use of non-normal distributions. On the first hand, our attention will be devoted to show that using a skewed Student density can highly improve the precision of the Value-at-Risk forecasts (Chapters 5 and 6). On the other hand, we will show that using a non-normal density can shed some light on the effectiveness of central bank interventions.

Part II

Applications

Chapter 5

Value-at-Risk for Long and Short Positions

5.1 Introduction

In recent years, the tremendous growth of trading activity and the well-publicized trading loss of well known financial institutions (see Jorion, 2000, for a brief history of these events) has led financial regulators and supervisory committees of banks to favor quantitative techniques which appraise the possible loss that these institutions can incur. Value-at-Risk (VaR) has become one of the most sought-after techniques as it provides a simple answer to the following question: with a given probability (say α), what is my predicted financial loss over a given time horizon?

It turns out that the VaR has a simple statistical definition: the VaR at level α for a sample of returns is defined as the corresponding empirical quantile at $\alpha\%$. Because of the definition of the quantile, we have that, with probability $1 - \alpha$, the returns will be larger than the VaR. In other words, with probability $1 - \alpha$, the losses will be smaller than the dollar amount given by the VaR.¹ From an empirical point of view, the computation of the VaR for a collection of returns thus requires the computation of the empirical quantile at level α of the distribution of the returns of the portfolio.

Most models in the literature focus on the computation of the VaR for negative returns (see van den Goorbergh and Vlaar, 1999 or Jorion, 2000). Indeed, it is assumed that traders or portfolio managers have long trading positions, i.e. they

¹Contrary to some wide-spread beliefs, the VaR does not specify the maximum amount that can be lost.

bought the traded asset and are concerned when the price of the asset falls. In this chapter we focus on modelling VaR for portfolios defined on long *and short* trading positions. Thus we model VaR for traders having either bought the asset (long position) or short-sold it (short position).² In the first case, the risk comes from a drop in the price of the asset, while the trader loses money when the price increases in the second case (because he would have to buy back the asset at a higher price than the one he got when he sold it). Correspondingly, one focuses in the first case on the left side of the distribution of returns, and on the right side of the distribution in the second case.

Because the distribution of returns is often not symmetric (see Section 5.3), we show that “usual” parametric VaR models of the RiskMetrics and ARCH class have a tough job in modelling correctly the left and right tails of the distribution of returns. This is also true for the so-called asymmetric GARCH models where the asymmetry refers to the relationship between the conditional variance and the lagged squared error term. Indeed, as pointed out by El Babsiri and Zakoian (1999), although such asymmetric GARCH models allow positive and negative changes to have different impacts on future volatilities, the two components of the innovation have - up to a constant - the same volatilities, while it is desirable to allow an asymmetric confidence interval around the predicted volatility in the VaR application.

To alleviate these problems, we use the skewed Student Asymmetric Power ARCH (APARCH) model presented in Chapter 2 to model the VaR for portfolios defined on long (long VaR) and short (short VaR) trading positions. We compare the performance of this new model with that of the RiskMetrics, normal and Student APARCH models and show that the new model brings about considerable improvements in correctly forecasting one-day-ahead VaR for long and short trading positions on daily stock indexes (French CAC40, German DAX, US NASDAQ, Japanese NIKKEI and Swiss SMI data). For the skewed Student APARCH model, we also compute the expected short-fall and the average multiple of tail event to risk measure as these two measures supplement the information given by the empirical failure rates.

²An asset is short-sold by a trader when it is first borrowed and subsequently sold on the market. By doing this, the trader hopes that the price will fall, so that he can then buy the asset at a lower price and give it back to the lender. See Sharpe, Alexander, and Bailey (1999) for general information on trading procedures.

While we focus exclusively on parametric models, other approaches are possible, such as Danielsson and de Vries (2000) who combine a historical simulation method (i.e. non parametric technique) for the interior of the distribution of returns with a fitted distribution based on extreme value theory for the most extreme returns. In this setting, normal and extreme events are thus modelled using two different methods. With the skewed Student APARCH model we aim to model left and right tail VaRs with a single parametric method for a wide range of values for α .

Recently, Mittnik and Paolella (2000) have introduced an APARCH model coupled with an asymmetric generalized Student distribution to model VaR for negative returns. While the analysis in their paper is sometimes similar to ours, there are some significant differences. First, we focus on the joint behavior of VaR models for long *and short* trading positions, i.e. we look at both how large negative and positive returns are taken into account by the model (Mittnik and Paolella, 2000, focus on long VaR only). Secondly, our empirical analysis deals with daily data for stock indexes, in contrast to exchange rate data for the other paper. That usual datasets such as the daily returns for European and US indexes indicate the need for these types of models is an important issue, as most studies usually focus on “exotic” series for justifying the use of these models. Thirdly, we assess the performances of the models by computing Kupiec (1995)’s LR tests on the empirical failure rates. For the new model, we also compute the expected short-fall and the average multiple of tail event to risk measure. Last, from a methodological point of view, following the methodology presented in Chapter 2, we re-express the estimated parameters in terms of the mean and variance of the skewed Student distribution (instead of the mode and the dispersion).

As indicated in Christoffersen and Diebold (2000), volatility forecastability (such as featured by ARCH class models) decays quickly with the time horizon of the forecasts. An immediate consequence is that volatility forecastability is relevant for short time horizons (such as daily trading), but not for long time horizons on which portfolio managers usually focus. In this chapter, we are consistent with these characteristics of volatility forecastability as we focus on daily returns and analyze VaR performance for daily trading portfolios made up of long and short positions.

The rest of the chapter is organized in the following way. In Section 5.2, we describe the symmetric and asymmetric VaR models. These models are applied

to daily stock indexes data in Section 5.3 where we assess their performances and characterize the long and short VaR.

5.2 VaR models

In this section we present parametric VaR models of the ARCH class. ARCH class models were first introduced by Engle (1982) with the ARCH model. Since then, numerous extensions have been put forward, see Engle (1995), Bera and Higgins (1993) or Palm (1996), but they all share the same goal, i.e. modelling the conditional variance as a function of past (squared) returns and associated characteristics. Because quantiles are direct functions of the variance in parametric models, ARCH class models immediately translate into conditional VaR models. As mentioned in the introduction, these conditional VaR models are important for characterizing short term risk for intradaily or daily trading positions.

In the first sub-section we characterize the symmetric (RiskMetrics, normal and Student APARCH) and asymmetric (skewed Student APARCH) volatility models, while we detail corresponding VaR results for negative and positive returns in the second sub-section. We stress that, by symmetric and asymmetric models, we mean a possible asymmetry in the distribution of the error term (i.e. whether it is skewed or not), and not the asymmetry in the relationship between the conditional variance and the lagged squared innovations (the APARCH model features this kind of “conditional” asymmetry whatever the chosen error term).

5.2.1 Symmetric and asymmetric volatility models

To characterize the models, we consider a collection of daily returns, y_t , with $t = 1 \dots T$. Because daily returns are known to exhibit some serial autocorrelation, we fit an $AR(n)$ structure on the y_t series for all specifications:

$$\Psi(L)(y_t - \mu) = \varepsilon_t, \tag{5.1}$$

where $\Psi(L) = 1 - \psi_1 L - \dots - \psi_n L^n$.

We now consider several specifications for the the conditional variance of ε_t .

RiskMetrics

In its most simple form, it can be shown that the basic RiskMetrics model is equivalent to a normal IGARCH (1, 1) model where the autoregressive parameter is set at a prespecified value 0.94 and the coefficient of ε_{t-1}^2 is equal to 0.06. In the RiskMetrics specification, we have:

$$\varepsilon_t = z_t \sigma_t, \quad (5.2)$$

where z_t is *i.i.d.* $N(0, 1)$ and σ_t^2 is defined as:

$$\sigma_t^2 = 0.06\varepsilon_{t-1}^2 + 0.94\sigma_{t-1}^2. \quad (5.3)$$

Normal, Student and skewed Student APARCH

The APARCH (Ding, Granger, and Engle, 1993) is an extension of the GARCH model of Bollerslev (1986). This model, already presented in Eq. (2.44), is probably one of the most promising ARCH-type model. Indeed, it nests at least seven GARCH specifications. Recall that the APARCH(1,1) is:

$$\sigma_t^\delta = \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta_1 \sigma_{t-1}^\delta, \quad (5.4)$$

where $\omega, \alpha_1, \gamma, \beta_1$ and δ are parameters to be estimated. δ ($\delta > 0$) plays the role of a Box-Cox transformation of σ_t , while γ ($-1 < \gamma < 1$), reflects the so-called leverage effect. A positive (resp. negative) value of γ means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive shocks (see Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990). The properties of the APARCH model have been studied recently by He and Teräsvirta (1999a, 1999b).

The Normal APARCH (N APARCH), Student APARCH (ST APARCH) and skewed Student APARCH (SKST APARCH) assume respectively that z_t is *i.i.d.* $N(0, 1)$, $ST(0, 1, \nu)$ and $SKST(0, 1, \xi, \nu)$.

5.2.2 VaR for long and short positions

Because the goal of the current chapter is to check the performance of the models on both the long and short sides of daily trading, we are particularly interested

in comparing the Student APARCH model with the skewed Student APARCH model regarding their performance in forecasting one step ahead long and short VaR. As indicated in the introduction, the long side of the daily VaR is defined as the VaR level for traders having long positions in the relevant equity index: this is the “usual” VaR where traders incur losses when negative returns are observed. Correspondingly, the short side of the daily VaR is the VaR level for traders having short positions, i.e. traders who incur losses when stock prices increase. How good a model is at predicting long VaR is thus related to its ability to model large negative returns, while its performance regarding the short side of the VaR is based on its ability to take into account large positive returns.

For the RiskMetrics and normal APARCH models, the one-step-ahead VaR as computed in $t - 1$ for long trading positions is given by $n_\alpha \sigma_t$, for short trading positions it is equal to $n_{1-\alpha} \sigma_t$, with n_α being the left quantile at $\alpha\%$ for the normal distribution and $n_{1-\alpha}$ is the right quantile at $\alpha\%$.³ For the Student APARCH model, the VaR for long and short positions is given by $st_{\alpha,v} \sigma_t$ and $st_{1-\alpha,v} \sigma_t$, with $st_{\alpha,v}$ being the left quantile at $\alpha\%$ for the Student distribution with v degrees of freedom and $st_{1-\alpha,v}$ is the right quantile at $\alpha\%$ for this same distribution. Because $n_\alpha = -n_{1-\alpha}$ for the normal distribution and $st_{\alpha,v} = -st_{1-\alpha,v}$ for the Student distribution, the forecasted long and short VaR will be equal in both cases.

For the skewed Student APARCH model, the VaR for long and short positions is given by $skst_{\alpha,v,\xi} \sigma_t$ and $skst_{1-\alpha,v,\xi} \sigma_t$, with $skst_{\alpha,v,\xi}$ being the left quantile at $\alpha\%$ for the skewed Student distribution with v degrees of freedom and asymmetry coefficient ξ ; $skst_{1-\alpha,v,\xi}$ is the corresponding right quantile.

Using Eq. (2.33) we can easily relate the quantile function of the (standardized) skewed Student density ($skst_{\alpha,v,\xi}$) with the one of the symmetric Student density ($st_{\alpha,v}$), i.e.

$$skst_{\alpha,v,\xi} = \begin{cases} \frac{1}{\xi} st_{\alpha,v} \left[\frac{\alpha}{2} (1+\xi^2) \right]^{-m} & \text{if } \alpha < \frac{1}{1+\xi^2} \\ \frac{-\xi st_{\alpha,v} \left[\frac{1-\alpha}{2} (1+\xi^{-2}) \right]^{-m}}{s} & \text{if } \alpha \geq \frac{1}{1+\xi^2} \end{cases}, \quad (5.5)$$

where m and s depend on ξ and v and are given in Eq. (2.27) and (2.28). If $\ln(\xi)$ is smaller than zero (or $\xi < 1$), $|skst_{\alpha,v,\xi}| > |skst_{1-\alpha,v,\xi}|$ and the VaR for long trading positions will be larger (for the same conditional variance) than the VaR

³All VaR expressions are reported for the residuals ε_t , which is equivalent to reporting the VaR centered around the expected return based on Eq. (5.1).

for short trading positions. When $\ln(\xi)$ is positive, we have the opposite result.

5.3 Empirical application

5.3.1 Data

In this empirical application we consider daily data for a collection of 5 stock market indexes (source: Datastream): the French CAC 40 stock index (CAC, 1/1/1990 - 21/12/2000), the German DAX stock index (DAX, 26/11/1990 - 21/12/2000), the U.S. NASDAQ stock index (NASDAQ, 11/10/1984 - 21/12/2000), the Japanese NIKKEI stock index (NIKKEI, 4/1/1984 - 21/12/2000) and the Swiss SMI stock index (SMI, 9/11/1990 - 21/12/2000), where the numbers in parentheses are the start and end dates for the sample at hand and the first symbol inside the parentheses designates the short notation for the index that will be used in the tables and comments below. The VaR models introduced in Section 5.2 are tested on these five datasets.

For all price series p_t , daily returns are defined as $y_t = 100 \times [\ln(p_t) - \ln(p_{t-1})]$. Descriptive statistics for the return series are given in Table 5.1. While the time spans for the five stock indexes are different, the five return series share similar statistical properties as far as third and fourth moments are concerned. More specifically, the returns series are negatively skewed and the large returns (either positive or negative) lead to a large degree of kurtosis. The Ljung-Box Q-statistic of order 10 on the squared series indicates that the conditional variances vary over time.

Descriptive graphs (level of index, daily returns, density of the daily returns and QQ-plot against the normal distribution) for each index are given in Figures 5.1-5.5. Volatility clustering is immediately apparent from the graphs of daily returns. The density graphs and the QQ-plot against the normal distribution show that all returns distributions exhibit fat tails. Moreover, the QQ-plots indicate that fat tails are not symmetric.

5.3.2 Estimating the models

In order to perform the VaR analysis in Section 5.3.3, the normal APARCH, Risk-Metrics, Student APARCH and skewed Student APARCH models are estimated

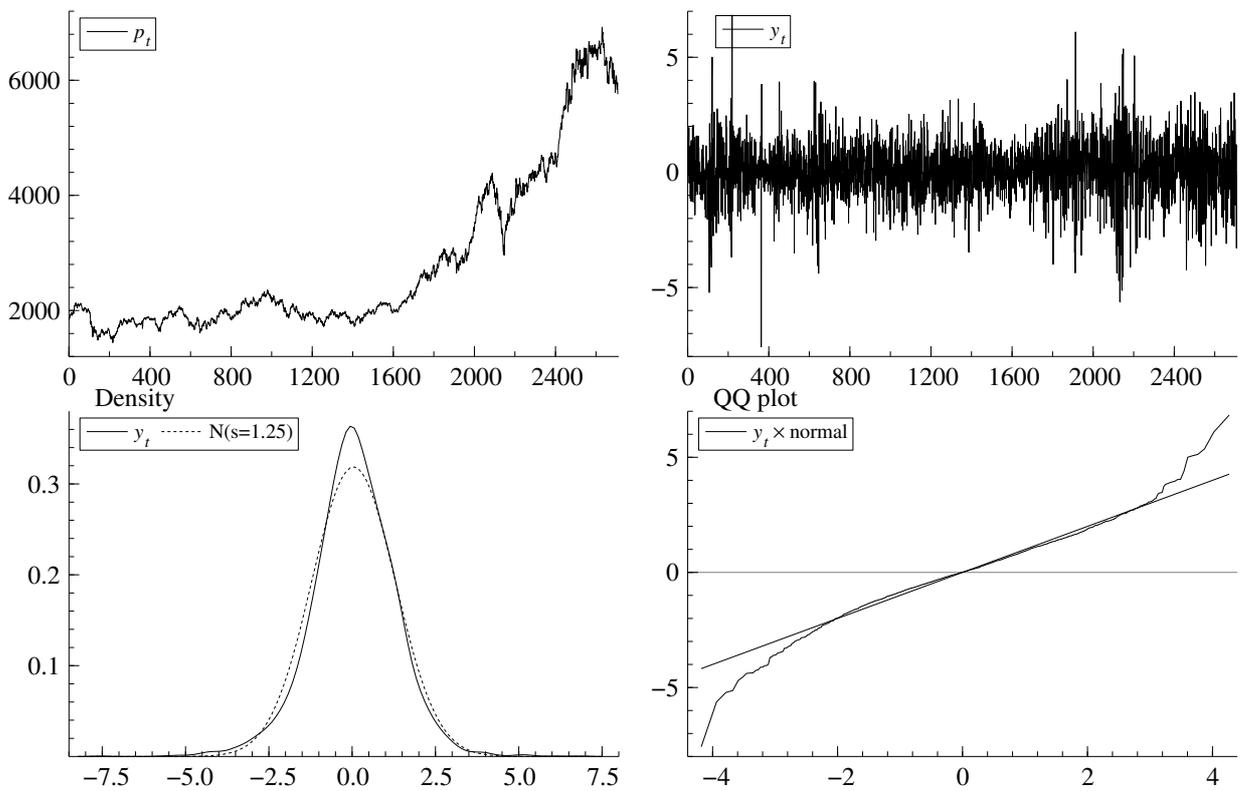


Figure 5.1: CAC 40 stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 1/1/1990 - 21/12/2000.

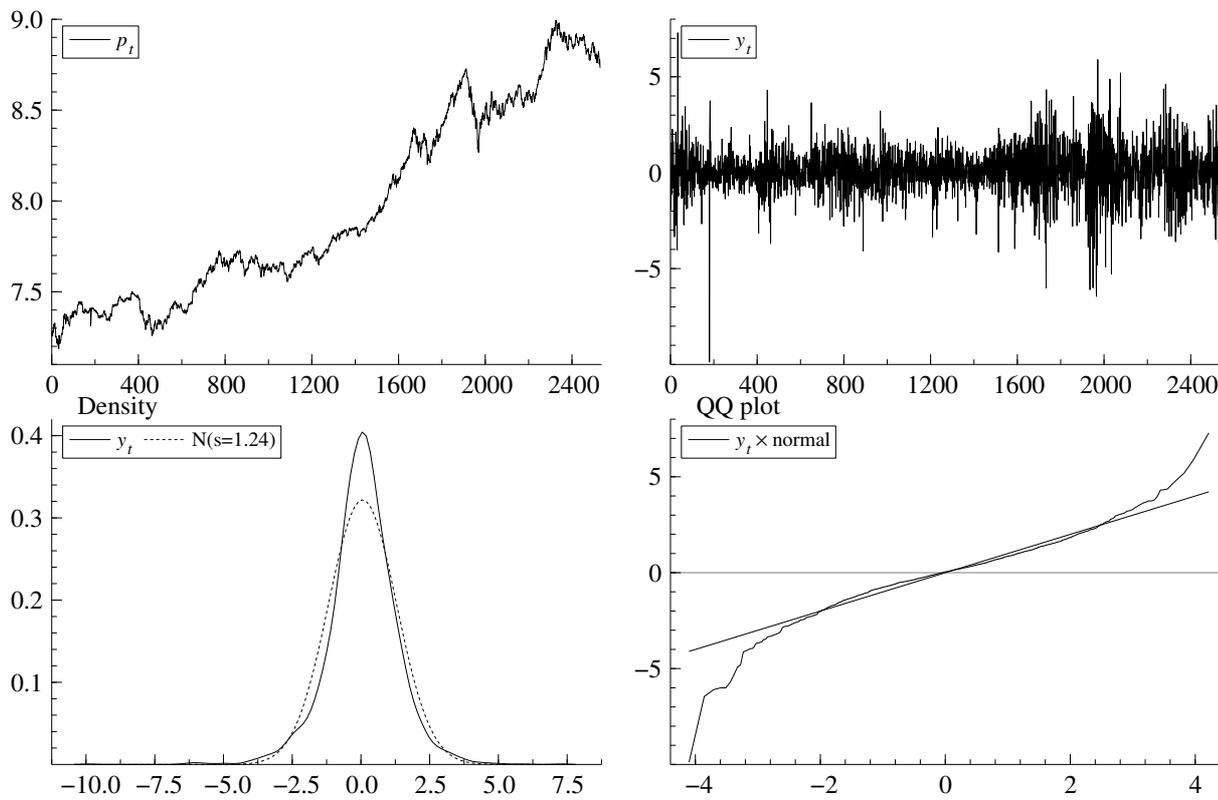


Figure 5.2: DAX stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 26/11/1990 - 21/12/2000.

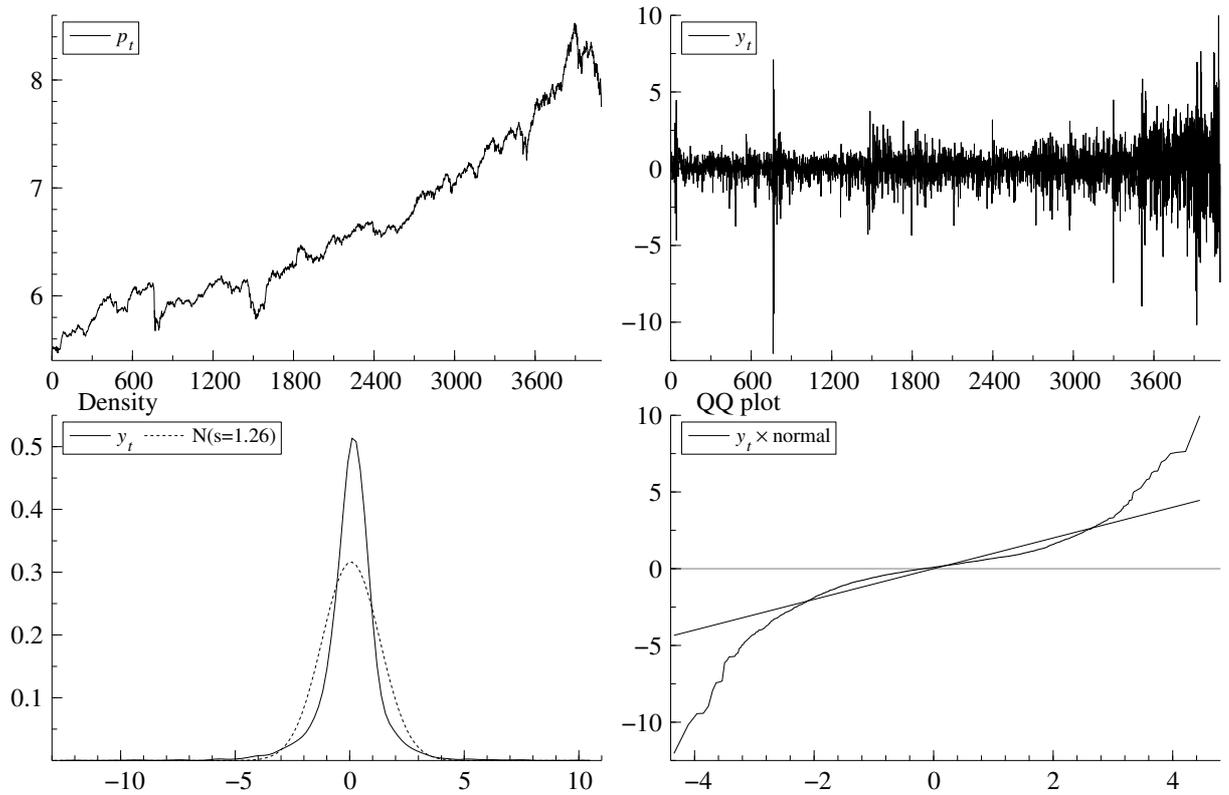


Figure 5.3: NASDAQ stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 11/10/1984 - 21/12/2000.

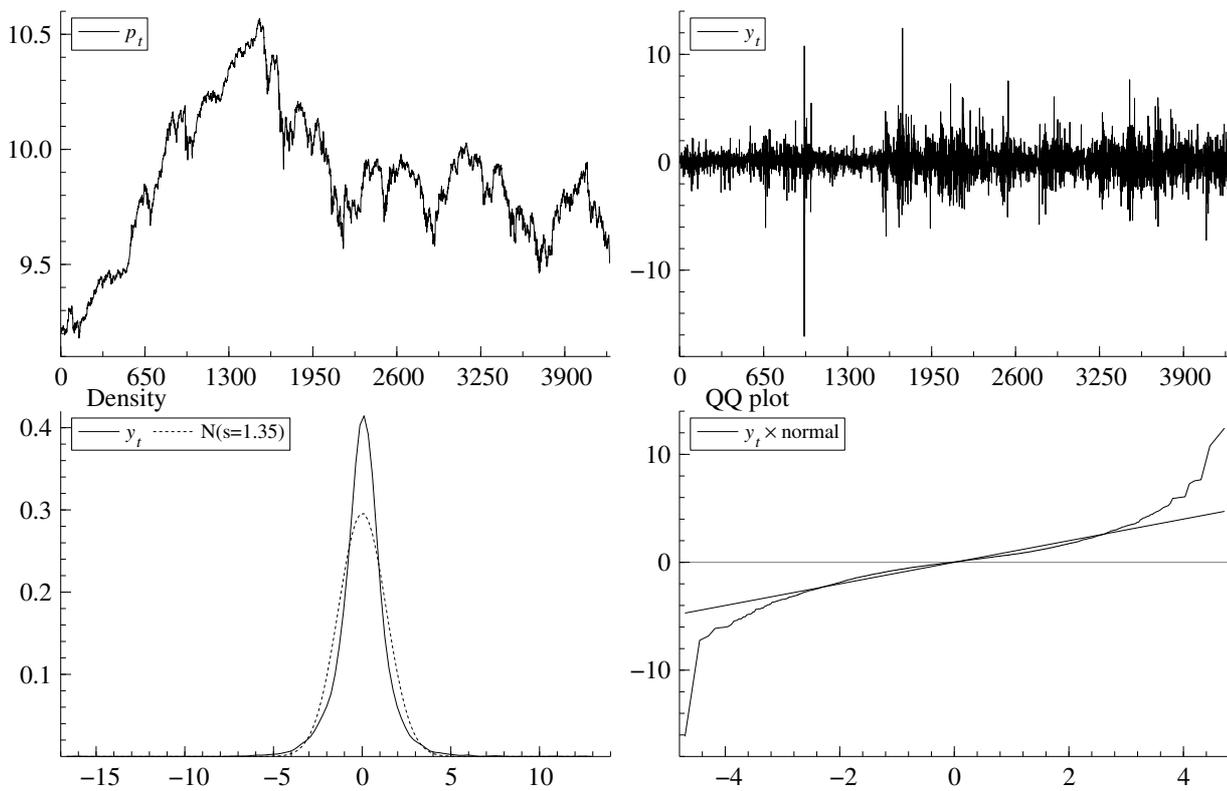


Figure 5.4: NIKKEI stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 4/1/1984 - 21/12/2000.

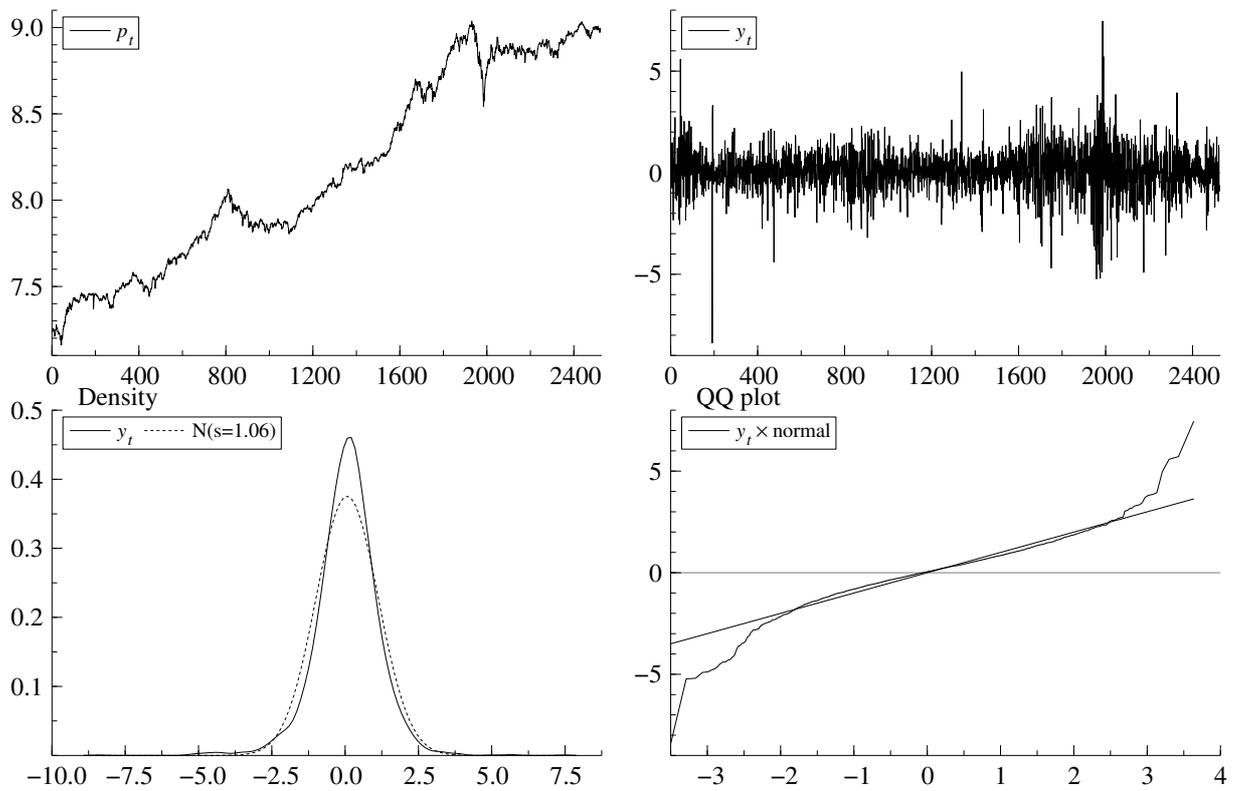


Figure 5.5: SMI stock index in level, daily returns, daily returns density and QQ-plot against the normal distribution. The time period is 9/11/1990 - 21/12/2000.

Table 5.1: Descriptive statistics

	CAC	DAX	NASDAQ	NIKKEI	SMI
Annual mean	10.66	14.53	13.90	1.79	17.44
Annual s.d.	19.87	19.67	20.03	21.38	16.87
Skewness	-0.16	-0.39	-0.74	-0.14	-0.41
Excess Kurtosis	2.09	4.15	11.25	10.15	5.13
Minimum	-7.57	-9.87	-12.04	-16.14	-8.38
Maximum	6.83	7.29	9.96	12.43	7.46
$Q^2(10)$	444.8	428.7	3269.8	635.1	849.7

Descriptive statistics for the daily returns of the corresponding stock index expressed in %. All values are computed using PcGive. $Q^2(10)$ is the Ljung-Box Q-statistic of order 10 on the squared series.

in this section. We do not report full estimation results of the normal and Student APARCH models as they are quite similar to what has been documented in the literature (see for instance Ding, Granger, and Engle, 1993 and Paoella, 1997). Furthermore, these specifications are encompassed by the skewed Student APARCH model for which we give full details below. The RiskMetrics model does not require any estimation for the conditional volatility specification as it is tantamount to an IGARCH model with some predefined values.

Table 5.2 presents the results for the (approximate QML) estimation of the APARCH model with a skewed Student pseudo-likelihood on the CAC, DAX, NASDAQ, NIKKEI and SMI data. An AR(3) was found to be sufficient to correct the serial correlation in the conditional mean. Note that to save some space, the estimated mean parameters are not reported.

The model is particularly successful in taking into account the heteroskedasticity exhibited by the data as the Ljung-Box Q-statistic computed on the squared standardized residuals is never significant.⁴ The five stock market indexes feature relatively similar volatility specifications:

- the autoregressive effect in the volatility specification is strong as β_1 is around

⁴For NASDAQ data, the decrease in the Q^2_{10} is impressive as it goes down from more than 3,000 to about 12.

Table 5.2: Skewed Student APARCH

	CAC	DAX	NASDAQ	NIKKEI	SMI
ω	0.032 (0.013)	0.013 (0.006)	0.016 (0.006)	0.024 (0.005)	0.029 (0.010)
α_1	0.056 (0.011)	0.081 (0.012)	0.128 (0.022)	0.105 (0.012)	0.089 (0.016)
γ	0.501 (0.126)	0.248 (0.083)	0.291 (0.063)	0.493 (0.079)	0.408 (0.080)
β_1	0.915 (0.017)	0.926 (0.012)	0.887 (0.023)	0.897 (0.012)	0.902 (0.022)
ν	11.325 (2.793)	7.808 (1.441)	6.737 (0.708)	6.519 (0.703)	8.184 (1.562)
$\ln(\xi)$	-0.067 (0.029)	-0.073 (0.029)	-0.185 (0.023)	-0.054 (0.023)	-0.124 (0.031)
δ	1.378 (0.207)	1.133 (0.132)	1.122 (0.178)	1.185 (0.133)	1.002 (0.153)
Q_{10}^2	9.5	0.7	12.7	13.3	2.1
$\alpha_1 E(z - \gamma z)^\delta + \beta_1$	0.981	0.992	1.003	0.986	0.979

Estimation results for the volatility specification of the skewed Student APARCH model. Robust standard errors are reported in parentheses. Q_{10}^2 is the Ljung-Box Q-statistic of order 10 computed on the squared standardized residuals.

0.9, suggesting a strong memory effects. Indeed, $\alpha_1 E(|z| - \gamma z)^\delta + \beta_1$ is just below 1 for four indexes and equals 1 for the NASDAQ (indicating that σ_t^δ may be integrated on this period).

- γ is positive and significant for all datasets, indicating a leverage effect for negative returns in the conditional variance specification;
- $\ln(\xi)$ is negative and significant for all datasets, which implies that the asymmetry in the Student distribution is needed to fully model the distribution of returns. Likelihood ratio tests (not reported) also clearly favor the skewed Student density;
- δ is between 1.002 and 1.378 and always significantly different from 2. The results suggest that instead of modelling the conditional variance (GARCH) it is more relevant to model the conditional standard deviation (indeed, δ is not significantly different from 1). This result is in line with those of Taylor (1986), Schwert (1990) and Ding, Granger, and Engle (1993) who indicate that there is substantially more correlation between absolute returns than squared returns, a stylized fact of high frequency financial returns (often called “long memory”).

These results indicate the need for a model featuring a negative leverage effect (conditional asymmetry) for the conditional variance combined with an asymmetric distribution for the underlying error term (unconditional asymmetry). The skewed Student APARCH model delivers such specifications and we study in Section 5.3.3 whether this model improves on symmetric GARCH models when the VaR for long and short returns is needed.

5.3.3 In-sample VaR computation

In this section, we use the estimation results of Section 5.3.2 and the expressions of Section 5.2.2 to compute the one-step-ahead VaR for all models. As financial returns are known to exhibit fat tails (this was confirmed in the descriptive properties of the data given in Table 5.1), we expect poor performance by the models based on the normal distribution.

All models are tested with a VaR level α which ranges from 5% to 0.25% and their performance is then assessed by computing the failure rate for the returns y_t .

By definition, the failure rate is the number of times returns exceed (in absolute value) the forecasted VaR. If the VaR model is correctly specified, the failure rate should be equal to the prespecified VaR level. In our empirical application, we define a failure rate f_l for the long trading positions, which is equal to the percentage of negative returns smaller than one-step-ahead VaR for long positions. Correspondingly, we define f_s as the failure rate for short trading positions as the percentage of positive returns larger than the one-step-ahead VaR for short positions.

Because the computation of the empirical failure rate defines a sequence of yes/no observations, it is possible to test $H_0 : f = \alpha$ against $H_1 : f \neq \alpha$, where f is the failure rate (estimated by \hat{f} , the empirical failure rate).⁵ At the 5% level and if T yes/no observations are available, a confidence interval for \hat{f} is given by $\left[\hat{f} - 1.96\sqrt{\hat{f}(1-\hat{f})/T}, \hat{f} + 1.96\sqrt{\hat{f}(1-\hat{f})/T} \right]$. In this chapter these tests are successively applied to the failure rate f_l for long trading positions and then to f_s , the failure rate for short trading positions.

In Table 5.3 we present complete VaR results (i.e. p-values for the Kupiec LR test) for the NASDAQ and NIKKEI stock indexes. In Table 5.4 we give summary results for the five stock indexes. These results indicate that:

- VaR models based on the normal distribution (RiskMetrics and normal APARCH model) have a difficult job in modelling large returns, with large positive returns being somewhat better handled than large negative returns.
- the symmetric Student APARCH model improves considerably on the performance of normal based models but its performance is still not satisfactory for large positive returns. For the NASDAQ index, its performance in general is even worse than normal based models. The reason is that the critical values of the Student distribution $st_{\alpha,v}$ and $st_{1-\alpha,v}$ are very large in this case, which leads to a high level of long and short VaR: the model is often rejected because it is too conservative.⁶

⁵In the literature on VaR models, this test is also called the Kupiec LR test, if the hypothesis is tested using a likelihood ratio test. See Kupiec (1995).

⁶For example, the empirical failure rates for the short VaR are equal to 3.59%, 1.39%, 0.37%, 0.10% and 0.05% when α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%: in all cases the model is rejected because it is too conservative.

Table 5.3: VaR results for NASDAQ and NIKKEI (in-sample)

α	5%	2.5%	1%	0.5%	0.25%
VaR for long positions (NASDAQ)					
N APARCH	0.084	0	0	0	0
RiskMetrics	0.002	0	0	0	0
ST APARCH	0	0	0.069	0.075	0.028
SKST APARCH	0.498	0.743	0.533	0.920	0.409
VaR for long positions (NIKKEI)					
N APARCH	0.563	0.336	0	0.046	0
RiskMetrics	0	0.006	0	0	0
ST APARCH	0.063	0.563	0.592	0.789	0.477
SKST APARCH	0.563	0.838	0.043	0.472	0.675
VaR for short positions (NASDAQ)					
N APARCH	0	0	0.032	0.305	0.811
RiskMetrics	0.001	0.011	0.863	0.166	0.162
ST APARCH	0	0	0	0	0.002
SKST APARCH	0.593	0.743	0.533	0.205	0.151
VaR for short positions (NIKKEI)					
N APARCH	0	0.018	0.399	0.006	0
RiskMetrics	0.004	0.094	0.023	0	0
ST APARCH	0.004	0.024	0.486	0.423	0.904
SKST APARCH	0.071	0.189	0.809	0.214	0.204

P-values for the null hypothesis $f_l = \alpha$ (i.e. failure rate for the long trading positions is equal to α , top of the table) and $f_s = \alpha$ (i.e. failure rate for the short trading positions is equal to α , bottom of the table). α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%. The models are successively the normal APARCH, RiskMetrics, Student APARCH and skewed Student APARCH models.

Table 5.4: VaR results for all indexes (in-sample)

VaR for long positions					
	CAC	DAX	NASDAQ	NIKKEI	SMI
N APARCH	40	0	20	40	20
RiskMetrics	0	0	0	0	0
ST APARCH	100	100	40	100	80
SKST APARCH	100	100	100	80	100
VaR for short positions					
	CAC	DAX	NASDAQ	NIKKEI	SMI
N APARCH	20	0	40	20	80
RiskMetrics	20	40	60	20	80
ST APARCH	60	60	0	60	20
SKST APARCH	100	80	100	100	100

Number of times (out of 100) that the null hypothesis $f_l = \alpha$ (i.e. failure rate for the long trading positions is equal to α , top of the table) is not rejected and the null hypothesis $f_s = \alpha$ (i.e. failure rate for the short trading positions is equal to α , bottom of the table) is not rejected for the five possible values of α (the level of significance is 5%). The models are successively the normal APARCH, RiskMetrics, Student APARCH and skewed Student APARCH models.

- the skewed Student APARCH model improves on all other specifications for both negative and positive returns. For the NASDAQ the improvement is substantial as the switch to a skewed Student distribution alleviates almost all problems due to the “conservativeness” of the symmetric Student distribution. The model performs correctly in 100% of all cases for the negative returns (long VaR) and for the positive returns (short VaR). As indicated in Table 5.4, the skewed Student APARCH model correctly models nearly all VaR levels for long and short positions (the success rate is 100% for four stock indexes and 80% for one index). In all cases, this is a significant improvement on the VaR performances of symmetric models.

5.3.4 Out-of-sample VaR computation

The testing methodology in the previous subsection is equivalent to back-testing the model on the estimation sample. In a “real life situation”, VaR models are used to deliver out-of-sample forecasts, where the model is estimated on the known returns (up to time t for example) and the VaR forecast is made for period $[t+1, t+h]$, where h is the time horizon of the forecasts. In this subsection we implement this testing procedure for the long and short VaR with $h = 1$ day.

We use an iterative procedure where the skewed Student APARCH model is estimated to predict the one-day-ahead VaR. The first estimation sample is the complete sample for which the data is available less the last five years. The predicted one-day-ahead VaR (both for long and short positions) is then compared with the observed return and both results are recorded for later assessment using the statistical tests. At the second iteration, the estimation sample is augmented to include one more day, the model is *re-estimated* and the VaRs are forecasted and recorded. We iterate the procedure until all days (less the last one) have been included in the estimation sample. Corresponding failure rates are then computed by comparing the long and short forecasted VaR_{t+1} with the observed return ε_{t+1} for all days in the five years period. We use the same statistical tests as in the subsection dealing with the in-sample VaR.

Empirical results for the five stock indexes are given in Table 5.5. Broadly speaking, these results are quite similar (although not as good) to those obtained for the in-sample testing procedure as the skewed Student APARCH model performs rather well for out-of-sample VaR prediction. Its combined (i.e. for long

Table 5.5: VaR results
(skewed Student APARCH, out-of-sample)

α	5%	2.5%	1%	0.5%	0.25%
VaR for long positions					
CAC	0.103	0.014	0.016	0.590	0.486
DAX	0.207	0.427	0.864	0.783	0.645
NASDAQ	0.005	0.101	0.355	0.311	0.646
NIKKEI	0.035	0.427	0.643	0.590	0.156
SMI	0.132	0.022	0.910	0.325	0.645
VaR for short positions					
CAC	0.007	0.334	0.697	0.311	0.645
DAX	0.026	0.520	0.163	0.784	0.932
NASDAQ	0	0.001	0.151	0.515	0.337
NIKKEI	0.797	0.786	0.355	0.515	0.932
SMI	0.045	0.101	0.697	0.784	0.337

P-values for the null hypotheses $f_l = \alpha$ (i.e. failure rate for the long trading positions is equal to α , top of the table) and $f_s = \alpha$ (i.e. failure rate for the short trading positions is equal to α , bottom of the table). α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%. The failure rates are computed for the skewed Student APARCH model (out-of-sample estimation).

and short VaR) success rate (at the 5% level) is equal to 70% (CAC), 90% (DAX), 70% (NASDAQ), 90% (NIKKEI) and 80% (SMI, almost 90% as one p-value is very close to the 0.05 level).

5.3.5 Expected short-fall and related measures

Our analysis in sub-sections 5.3.3 and 5.3.4 focused on the computation of empirical failure rates. In the last part of the empirical application, we now characterize the skewed Student APARCH model with respect to two other VaR related measures: the expected short-fall and the average multiple of tail event to risk measure.

The expected short-fall (see Scaillet, 2000) is defined as the expected value of the losses conditioned on the loss being larger than the VaR. The average multiple of tail event to risk measure “measures the degree to which events in the tail of the distribution typically exceed the VaR measure by calculating the average multiple of these outcomes to their corresponding VaR measures” (Hendricks, 1996). Both measures are computed for the in-sample estimation of the long and short VaR⁷ performed in sub-section 5.3.3.

For the expected short-fall, we report full estimation results for the NASDAQ and NIKKEI stock indexes⁸ in Table 5.6. These results indicate that the expected short-fall is in most cases larger (in absolute value) for the models based on the Student distribution than for the models based on the normal distribution. This is easily understood if one remembers that these models “fail” less than the ones based on the normal distribution, but, when they fail, it happens for large (in absolute value) returns: the average of these returns is correspondingly large. It should be stressed that the expected short-fall is not a tool to rank VaR models or assess models’ performances. Nevertheless it is useful for risk managers as it answers the following question: “when my model fails, how much do I lose on average?”.

A related measure is the average multiple of tail event to risk measure, which is reported in Table 5.7 for the NASDAQ and NIKKEI stock indexes. The figures in this table indicate what is the average loss/predicted loss when the VaR model fails. For example, the 1.38 for the long VaR with NASDAQ data and the skewed Student APARCH models indicates that, at the 1% level, one expects to lose 1.38 the amount given by the VaR when returns are smaller than the VaR. As for the expected short-fall, this measure does not allow a ranking of the VaR models.

5.4 Conclusion

Over short-term time horizons, conditional VaR models are usually found to be good candidates for quantifying possible trading losses. In this chapter, we ex-

⁷The expected short-fall for the long VaR is computed as the average of the observed returns smaller than the long VaR. The expected short-fall for the short VaR is computed as the average of the observed returns larger than the short VaR. Computations are similar for the average multiple of tail event to risk measure.

⁸Estimation results for the other 3 indexes are very similar to those given in Table 5.6 and are not reported.

Table 5.6: Expected short-fall for NASDAQ and NIKKEI (in-sample)

α	5%	2.5%	1%	0.5%	0.25%
Expected short-fall for long positions (NASDAQ)					
N APARCH	-2.42	-2.60	-3.00	-3.40	-3.58
RiskMetrics	-2.28	-2.40	-2.75	-2.84	-3.03
ST APARCH	-2.32	-2.54	-3.25	-3.76	-4.31
SKST APARCH	-2.42	-2.75	-3.62	-4.07	-4.42
Expected short-fall for long positions (NIKKEI)					
N APARCH	-2.66	-3.09	-3.58	-4.03	-4.28
RiskMetrics	-2.56	-2.86	-3.23	-3.52	-3.59
ST APARCH	-2.61	-3.13	-3.88	-4.48	-5.27
SKST APARCH	-2.68	-3.23	-4.06	-4.62	-5.49
Expected short-fall for short positions (NASDAQ)					
N APARCH	2.41	2.58	3.02	3.26	3.24
RiskMetrics	2.22	2.76	2.99	2.98	3.55
ST APARCH	2.33	2.66	3.38	2.95	3.66
SKST APARCH	2.08	2.52	2.82	3.42	2.48
Expected short-fall for short positions (NIKKEI)					
N APARCH	2.80	3.37	3.85	4.08	4.26
RiskMetrics	2.67	3.04	3.92	4.17	4.24
ST APARCH	2.73	3.43	4.34	4.37	5.03
SKST APARCH	2.65	3.28	4.20	4.44	4.63

Expected short-fall (in-sample evaluation) for the long and short VaR (at level α) given by the normal APARCH, Student APARCH, RiskMetrics and skewed Student APARCH models. α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%.

Table 5.7: Average multiple of tail event to risk measure for NASDAQ and NIKKEI (in-sample)

α	5%	2.5%	1%	0.5%	0.25%
AMTERM for long positions (NASDAQ)					
N APARCH	1.42	1.36	1.37	1.35	1.35
RiskMetrics	1.51	1.46	1.41	1.39	1.35
ST APARCH	1.45	1.36	1.39	1.40	1.38
SKST APARCH	1.42	1.36	1.38	1.39	1.35
AMTERM for long positions (NIKKEI)					
N APARCH	1.38	1.36	1.32	1.46	1.49
RiskMetrics	1.44	1.40	1.36	1.41	1.40
ST APARCH	1.33	1.36	1.41	1.49	1.49
SKST APARCH	1.39	1.35	1.47	1.49	1.48
AMTERM for short positions (NASDAQ)					
N APARCH	1.27	1.23	1.20	1.20	1.16
RiskMetrics	1.29	1.28	1.22	1.16	1.20
ST APARCH	1.27	1.22	1.15	1.27	1.26
SKST APARCH	1.29	1.24	1.20	1.20	1.22
AMTERM for short positions (NIKKEI)					
N APARCH	1.37	1.37	1.32	1.28	1.24
RiskMetrics	1.40	1.35	1.32	1.30	1.31
ST APARCH	1.39	1.35	1.28	1.19	1.17
SKST APARCH	1.39	1.36	1.29	1.23	1.17

Average multiple of tail event to risk measure (AMTERM, in-sample evaluation) for the long and short VaR (at level α) given by the normal APARCH, Student APARCH, RiskMetrics and skewed Student APARCH models. α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%.

tended this analysis by introducing a VaR model that could take into account losses arising from long and short trading positions. Because of the nature of long and short trading, this translates into bringing forward a statistical model that correctly models the left and right tails of the distribution of returns. The proposed model is the skewed Student APARCH model. Because density distribution of returns are usually not symmetric, it is shown that models⁹ that rely on symmetric normal or Student distributions underperform compared to the new model when the one-day-ahead VaR is to be forecasted. All models were applied to daily data for five stock indexes (CAC40, DAX, NASDAQ, NIKKEI and SMI), with an out-of-sample testing procedure confirming the results of the in-sample backtesting method: in all cases the skewed Student APARCH model performed rather well. In the last part of the chapter, we also computed the expected short-fall and the average multiple of tail event to risk measure for the models.

At this stage, several extensions can be considered. First, the performance of the new VaR model could also be assessed on multi-days period forecasts. While VaR models based on ARCH class specifications perform rather well for one-day time horizons, it is known that their performance is not as good for long time periods. Some recent work in this field has been done by Christoffersen and Diebold (2000). Secondly, the VaR for long and short trading positions could be computed using non-parametric VaR models. Computation times and quality of VaR forecasts could be compared with the results given by the skewed Student APARCH. Finally, as argued recently by Engle and Patton (1999), time varying higher conditional moments are clearly of interest. In this respect, Hansen (1994), Harvey and Siddique (1999) and Lambert and Laurent (2000) have had some success in introducing dynamics in the third and fourth moments.

To conclude, the recent availability of intraday data has led to new developments concerning the estimation of the daily volatility. The notion of realized volatility has been introduced recently in the literature by Taylor and Xu (1997) and Andersen and Bollerslev (1998) and is computed as an aggregated measure of volatility defined on intraday returns. According to these authors it offers an “error free” measure of the daily volatility. Interestingly, when one uses this realized volatility instead of the conditional variance produced by a parametric ARCH-type

⁹We considered three symmetric volatility models: the RiskMetrics, normal and Student APARCH models.

model, the normality assumption on the innovation process is supported, which questions the relevance of the skewed Student density. Consequently, in the next chapter we will try to answer the following question: Does the use of the realized volatility invalidate the choice of a skewed Student density ?

Chapter 6

Modelling Daily Value-at-Risk using Realized Volatility and ARCH Type Models

6.1 Introduction

The recent widespread availability of databases recording the intraday price movements of financial assets (stocks, indexes, currencies, derivatives) has led to new developments in applied econometrics and quantitative finance as far as the modelling of daily and intradaily volatility is concerned. Focusing solely on the modelling of daily volatility using intraday data, the recent literature suggests at least three possible methods for characterizing volatility and risk at an aggregated level, which we take to be equal to one day in this chapter.

In the spirit of what has been done in the previous chapters, the first possibility is to sample the intraday data on a daily basis so that closing prices are recorded from which daily returns are computed. In this setting, the notion of intraday price movements is not an issue, as the method is tantamount to estimating a volatility model on daily data. One of the most famous example is the ARCH model of Engle (1982) and subsequent ARCH type models such as the GARCH model of Bollerslev (1986) (see Palm, 1996, for a recent survey).

The second method is based on the notion of realized volatility which was recently introduced in the literature by Taylor and Xu (1997) and Andersen and Bollerslev (1998) and which is grounded in the framework of continuous time finance with the notion of quadratic variation of a martingale. In this case, a daily

measure of volatility is computed as an aggregated measure of volatility defined on intraday returns. More specifically, the daily realized volatility is computed as the sum of the squared intraday returns for the given trading day. We thus make explicit use of the intraday returns to compute the realized volatility, from which the daily volatility is modelled. A third possibility is to estimate a high frequency duration model on price durations for the given asset, and then use this irregularly time-spaced volatility at the aggregated level. Examples are Engle and Russell (1997) or Giot (2000). In this chapter we focus on the first two methods as our aggregation level is equal to one day, and it is not clear how duration models could be of any help in this situation.

The recent literature on realized volatility and the huge literature on daily volatility models seem to indicate that a researcher or market practitioner faces two distinct possibilities when daily volatility is to be modelled. Going one way or the other is however not a trivial question. If one decides to model daily volatility using daily realized volatility, then intraday data are needed so that corresponding intraday returns can be computed. Even today, intraday data remain relatively costly and are not readily available for all assets. Furthermore, a large amount of data handling and computer programming is usually needed to retrieve the needed intraday returns from the raw data files supplied by the exchanges or data vendors. On the contrary, working with daily data is relatively simple and the data are broadly available. However, one has the feeling that all the relevant data are not taken into account, i.e. that by going at the intraday level one could get a much better model.

In this chapter we aim to address this issue by comparing the performance of a daily ARCH type model with the performance of a model based on the daily realized volatility when the one-step ahead Value-at-Risk (VaR) measure is to be computed for a stock or market index. This exercise is done for two stock indexes (French CAC40 and US SP500 indexes) for which intraday data are available over a long time period (i.e. at least 5 years). VaR modelling is a natural application of volatility models as in a parametric framework the VaR measure (which by definition is a quantile of the conditional distribution) is a deterministic function of the volatility. See Jorion (2000) for a recent review of VaR models. Because we have intraday data over a long time period, we can retrieve the daily closing prices for the indexes and then compute daily VaR measure using ARCH type

models. When we make use of all the available data and compute intraday returns and realized volatility, we then have the competing model which uses the intraday information.

Our main results can be summarized in one sentence: yes, an (adequate) ARCH type model can deliver accurate VaR forecasts and this model performs as well as a competing VaR model based on the realized volatility. The key issue is to use a daily ARCH type model that clearly recognizes the full features of the empirical data such as a high kurtosis and skewness in the observed returns. In this chapter we use the asymmetric skewed Student APARCH model presented in Chapter 2, which has been extensively used all along this thesis and was found to be satisfactory when applied to daily data (especially in VaR applications, see Chapter 5). It is also true that the model based on the realized volatility delivers equally adequate VaR forecasts but this comes at the expense of using intraday information. Thus, for the two indexes under review, the results clearly indicate that modelling the realized volatility may be useful, but it is far from being the only game in town.

The rest of the chapter is organized in the following way. In Section 6.2, we describe the available intraday data for the two stock indexes and characterize the stylized facts of the corresponding realized volatility. In Section 6.3, we introduce the two competing models (i.e. the skewed Student APARCH model for the daily returns and the model based on the realized volatility) for computing the one-step-ahead VaR. These two models are applied to the daily stock index data in Section 6.4 where we assess their performances. Section 6.5 concludes.

6.2 Data and stylized facts

6.2.1 Data

The data are available for two stock indexes on an intraday basis and for a relatively long period of time which allows VaR modelling and testing. For both assets we consider daily returns (which are used by the skewed Student APARCH model) and intraday returns defined on a 5-minute and 15-minute time grid (these intraday returns are used to compute the daily realized volatility).

Our first asset is the French CAC40 stock index for the 1995-1999 years (1249 daily observations). It is computed by the exchange as a weighted measure of the prices of its components and is available in the database on an intraday basis with

the price index being computed every 30 seconds (approximately). For the time period under review, the opening hours of the French stock market were 10h am to 5h pm, thus 7 hours of trading per day. With the 5- (15-) minute time grid, this translates into 84 (28) intraday returns used to compute the daily realized volatility. Intraday prices at the 5- and 15-minute level are the outcomes of a linear interpolation between the closest recorded prices below and above the time set in the grid.¹ Correspondingly, all returns are computed as the first difference in the regularly time-spaced log prices of the index. Because the exchange is closed from 5h pm to 10h am the next day, the first intraday return (computed at 10h05 when working with a 5-minute time grid for example) is the first difference between the log price at 10h05 and the log price at 5h pm the day before. Daily returns in percentage are defined as 100 times the first difference of the log of the closing prices.²

Our second dataset contains 12 years (from January 1989 to December 2000, 3241 daily observations) of tick-by-tick prices for SP500 futures contracts traded on the Chicago Mercantile Exchange. Such SP500 futures contracts can be traded from 8h30 am to 15h10 pm Chicago time, i.e. from 9h30 am to 16h10 pm New York time. To conveniently define 5- and 15-minute returns, we remove all prices recorded after 16h New York time.³ As for the CAC40 dataset, intraday prices at the 5- and 15-minute level are the outcomes of a linear interpolation between the closest recorded prices (for the nearest contract to maturity) below and above the time set in the regularly time-spaced sampling grid.⁴ Returns are computed as the first difference in the regularly time-spaced log prices of the index, with the overnight return included in the first intraday return. Daily returns in percentage are defined as 100 times the first difference of the log of the closing prices.

¹In practice, the discreteness of actual securities prices can render continuous-time models poor approximations at very high sampling frequencies. Furthermore, tick-by-tick prices are generally only available at unevenly-spaced time points, so the calculation of evenly-spaced high-frequency returns necessarily relies on some form of interpolation among prices recorded around the endpoints of the given sampling intervals. It is well known that this non-synchronous trading or quotation effect may induce negative autocorrelation in the interpolated return series.

²By definition and using the properties of the log distribution, the sum of the intraday returns is equal to the observed daily return based on the closing prices.

³Thus the last recorded price for the futures at 16h corresponds more or less to the closing price of the “cash” SP500 index computed from its constituents traded on the NYSE or NASDAQ.

⁴The choice of the nearest contract to maturity means that we always select very liquid futures contracts.

6.2.2 Realized volatility: stylized facts

Estimating and forecasting volatility is a key issue in empirical finance. After the introduction of the ARCH model by Engle (1982) or the Stochastic Volatility (SV) model (see Taylor, 1994) and their various extensions, a new generation of conditional volatility models has been advocated recently by Taylor and Xu (1997) and Andersen and Bollerslev (1998), i.e. models making use of the realized volatility. The origin of this concept is not so recent as it would seem at first sight. Merton (1980) already mentioned that, provided data sampled at a high frequency are available, the sum of squared realizations can be used to estimate the variance of an *i.i.d.* random variable. Taylor and Xu (1997) and Andersen and Bollerslev (1998) (among others) show that daily realized volatility may be constructed simply by summing up intraday squared returns. Assuming that a day can be divided in N equidistant periods and if $y_{i,t}$ denotes the intradaily return of the i^{th} interval of day t , it follows that the daily volatility for day t can be written as:

$$\left[\sum_{i=1}^N y_{i,t} \right]^2 = \sum_{i=1}^N y_{i,t}^2 + 2 \sum_{i=1}^N \sum_{j=i+1}^N y_{j,t} y_{j-i,t}. \quad (6.1)$$

If the returns have mean zero and are uncorrelated, $\sum_{i=1}^N y_{i,t}^2$ is a consistent (see Andersen, Bollerslev, Diebold, and Labys, 2001) and unbiased estimator⁵ of the daily variance σ_t^2 . Because all squared returns on the right side of this equation are observed when intraday data (at equidistant periods) are available, $\left[\sum_{i=1}^N y_{i,t} \right]^2$ is called the daily *realized* volatility.

By summing high-frequency squared returns we may then obtain an “error free”⁶ measure of the daily volatility. However, choosing a very high sampling

⁵Areal and Taylor (2000) show that even if this estimator is consistent and unbiased, it has not the least variance when N is finite. These authors propose to weight the intraday squared returns by a factor proportional to the intraday activity. This deflator may be obtained easily by applying Taylor and Xu’s (1997) variance multiplier or the Flexible Fourier Function (FFF) of Andersen and Bollerslev (1997). Due to the strong similarity of the results with the “non weighted squared returns”, we will not report the results using Areal and Taylor’s (2000) approach.

⁶The theory of quadratic variation reveals that, under suitable conditions, realized volatility is not only an unbiased ex-post estimator of daily return volatility, but also asymptotically free of measurement error (Andersen, Bollerslev, Diebold, and Labys, 2001).

frequency (30-seconds, 1-minute, etc.) may introduce a bias in the variance estimate due to market microstructure effects (bid-ask bounces, price discreteness or non-synchronous trading). As a trade off between these two biases, Andersen, Bollerslev, Diebold, and Labys (2001) propose the use of 5-minute returns to compute daily realized volatility. Using the FTSE-100 stock market index (on the period 1990-2000), Oomen (2001) shows that the realized volatility measure increases when the sampling interval decreases while the summation of the cross terms in Eq. (6.1) decreases. Comparing the average daily realized volatility and the autocovariance bias factor, Oomen (2001) argues that the optimal sampling frequency for his dataset suggests using 25-minute returns. For our two datasets, a sampling frequency of about 15-minute was found to be optimal.⁷ By way of illustration, we also present results for 5-minute returns.

Although the empirical work on realized volatility is still in its infancy, some stylized facts have already been ascertained and we highlight these with our datasets.

- First, the unconditional distribution of the realized volatility is highly skewed and kurtosed. On the other hand, the unconditional distribution of the logarithmic realized volatility is nearly gaussian, while standard tests reject the normality assumption. Figures 6.1 and 6.2 display the level and the unconditional distribution of the logarithmic realized volatility of the CAC40 and SP500 stock indexes based on 15-minute returns. From Figure 6.2, both series appear slightly skewed (the unconditional skewness are respectively 0.62 and 0.38) and kurtosed (the unconditional kurtosis are respectively equal to 4.25 and 3.37).
- Secondly, the (logarithmic) realized volatility appears to be fractionally integrated. Indeed, Figure 6.3 displays the first 200 autocorrelations of the logarithmic realized volatility of the CAC40 and SP500 stock indexes based on 15-minute returns. This figure shows that a shock on volatility dies out very slowly, which is neither in accordance with an ARMA structure (which implies an exponential decay) nor with a unit root process (ADF tests, not reported to save space, all clearly reject the unit root assumption). This is

⁷To find the optimal sampling frequency, Oomen (2001) proposes to plot both the sum of squared intra-daily returns and the autocovariance bias factor versus the sampling frequency. The “optimal” sampling frequency is chosen as the highest available frequency for which the autocovariance bias term has disappeared.

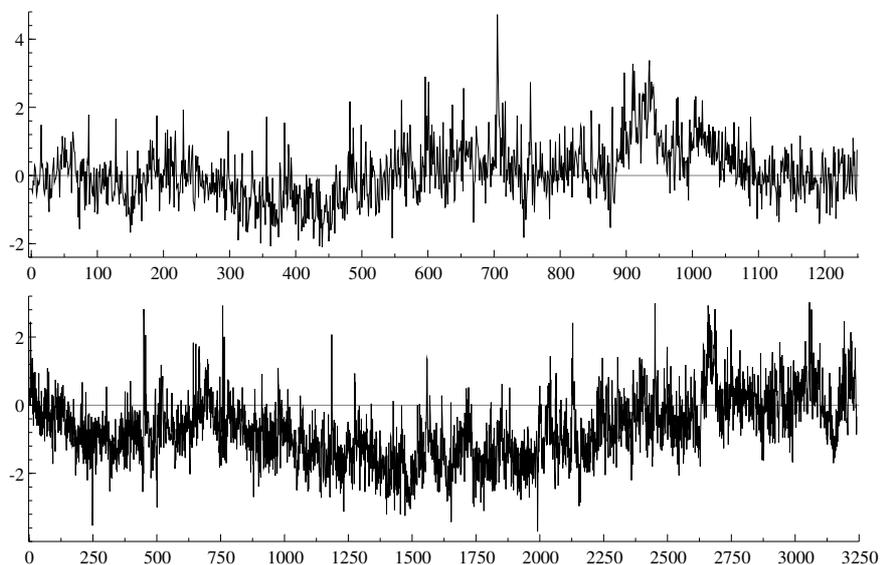


Figure 6.1: Logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.

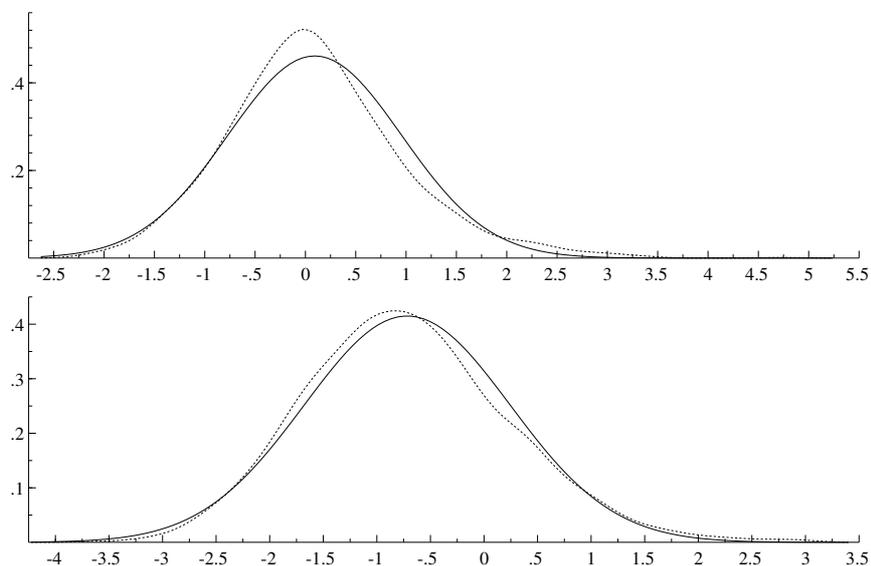


Figure 6.2: Density estimates (dashed line) and corresponding normal density (solid line) for the logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.

in line with the previous findings of Ding, Granger, and Engle (1993) and Baillie, Bollerslev, and Mikkelsen (1996) (among others) who suggest the modelling of conditional variance of high frequency financial data by the use of an (Asymmetric) Power GARCH (APARCH) or Fractionally Integrated GARCH (FIGARCH) models.

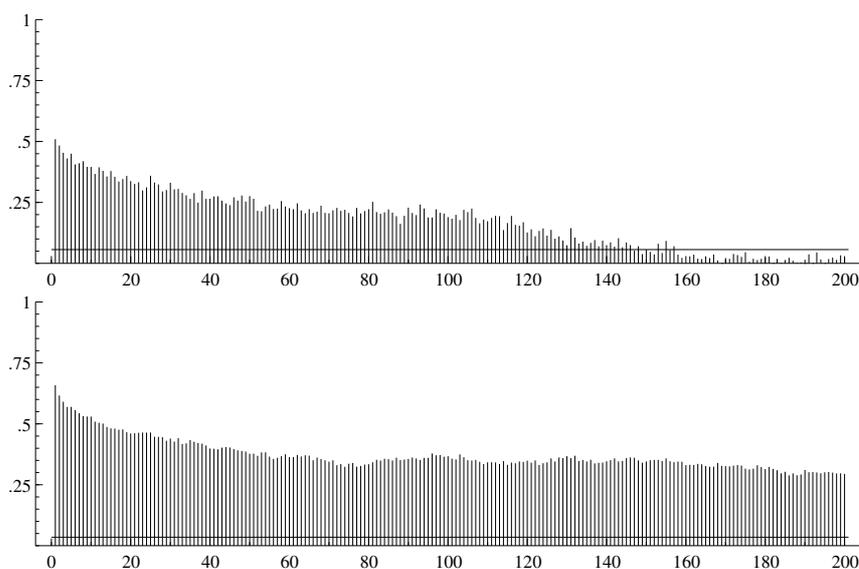


Figure 6.3: First 200 autocorrelations for the logarithmic realized volatility of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns. The horizontal lines show the upper limit 95% Bartlett confidence bands.

To gain a first insight in the degree of persistence of a shock on the (logarithmic) realized volatility, we computed the Geweke and Porter-Hudak (1983) (GPH) log-periodogram⁸ estimate for the fractional integration parameter d_a . If $d_a \in (0, 1/2)$, the process is stationary, has a long memory and is said to be persistent. If $d_a \in (-1/2, 0)$, the process has a short memory and is said to be antipersistent.⁹ The estimated d are equal to 0.437 (0.038) and 0.430 (0.026) respectively for the CAC40 and SP500 stock indexes based on

⁸The number of low frequency periodogram points used in the estimation is set to $T^{\frac{4}{5}}$, see Hurvich, Deo, and Brodsky (1998).

⁹Furthermore, if $d_a \leq -1/2$, the process is non invertible and if $d_a \geq 1/2$, the process is not stationary but mean reverting if $d_a < 1$.

15-minute¹⁰ returns (standard errors are given in parentheses). Thus d_a is fairly close to the “typical value” of 0.4 (see Andersen, Bollerslev, Diebold, and Labys, 2001, Ebens, 1999 among others) and just significantly lower than 0.5 at the 5% critical level, suggesting that these series might be covariance-stationary.

- Finally, according to Ebens (1999) who analyzes the Dow Jones Industrial portfolio over the January 1993 to May 1998 period, the (logarithmic) realized volatility of stock indexes are non-linear in returns. To show this, consider the following Least-Squares (LS) regression: $\ln RV_t = c_0 + c_1 y_{t-1} + c_2 y_{t-1}^- + u_t$, where $\ln RV_t$ is the logarithm of the realized volatility, y_t is the daily return on day t , y_t^- is equal to 0 when $y_t > 0$ and is equal to y_t when $y_t < 0$ and u_t is a white noise. Figure 6.4 displays the fitted values of these LS regressions (solid lines) for the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns as well as a nonparametric estimation (dashed lines).¹¹ These graphs suggest that a negative shock on the returns is more likely to be associated with a high volatility (the next day) than for a positive shock.¹² This feature is also well known for ARCH type models and is known as the *leverage effect*¹³ (see Black, 1976; French, Schwert, and Stambaugh, 1987; Pagan and Schwert, 1990 and Zakoian, 1994).

6.3 Two competing models

Realized volatility was reviewed in the preceding section and we can now introduce a model for the daily VaR based on this measure. Subsection 6.3.2 is devoted to this topic. As the goal of the chapter is to compare the performance of an ARCH type model directly applied to the daily data with the performance of a model

¹⁰Results for the 5-minute returns are very similar and are thus not reported.

¹¹Quite similar to Ebens (1999), the nonparametric regression estimates are obtained using the Nadaraya-Watson estimator with the Epanechnikov kernel while the bandwidth parameters are determined using cross-validation scores. The plot regions are restricted to returns in the -5 to 5 interval, even if all the sample size was used when estimating this nonparametric regression.

¹²The R^2 of these LS regressions are respectively 11.5 and 17.5%, which is very similar to the ones reported by Ebens (1999).

¹³Past negative (resp. positive) shocks have a different impact on current realized volatility than past positive shocks.

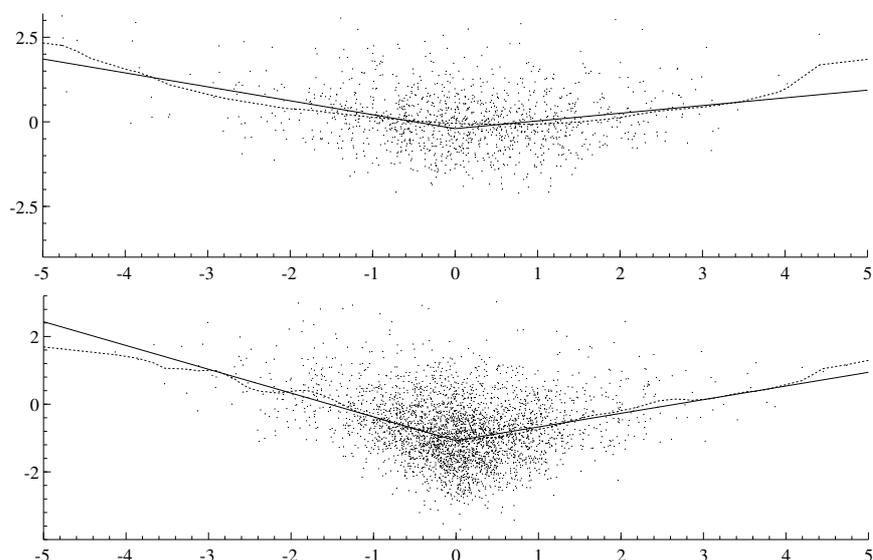


Figure 6.4: Regression lines for the logarithmic realized volatility (y-axis) of the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns against the previous (i.e. one day before) returns (x-axis).

based on the realized volatility, we also need to characterize the skewed Student APARCH model for the daily data. This is done in Subsection 6.3.1.

In both cases the link between the forecasted one-day-ahead volatility and the one-day-ahead VaR is immediate. Indeed, both models are parametric conditional models for volatility and the corresponding VaR measures are easily computed as the product of the square root of the conditional volatility and the quantile at $\alpha\%$ of the underlying distribution for the standardized error term.¹⁴ Thus, for example, if the forecasted volatility at time $t - 1$ is $\hat{\sigma}_t^2$ and one assumes a normal distribution for the error term, then the forecasted one-day-ahead VaR in $t - 1$ is equal to $n_\alpha \hat{\sigma}_t$, with n_α being the left quantile at $\alpha\%$ for the normal distribution.

6.3.1 The skewed Student APARCH model

To model daily returns y_t , with $t = 1, \dots, T$, we use the AR(3)-APARCH(1,1) model given in Eq. (5.1)-(5.4). Based on information criteria and standard serial

¹⁴In this chapter we consider a forecast for the demeaned VaR which only depends on the level of the volatility.

correlation tests, the AR(3)-APARCH(1,1) specification was found to be adequate in describing our two series. In order to save space, we only report the results concerning the more parsimonious specification.

In VaR applications, the choice of an appropriate distribution is an important issue. As in Chapters 2, 3 and 5, we use the (standardized) skewed Student distribution.

Because of the direct relationship between the VaR and the quantile in parametric VaR models, the one-day-ahead VaRs for long and short positions are given by $skst_{\alpha,v,\xi}\hat{\sigma}_t$ and $skst_{1-\alpha,v,\xi}\hat{\sigma}_t$, with $skst_{\alpha,v,\xi}$ being the left quantile at α for the skewed Student distribution with v degrees of freedom and asymmetry coefficient ξ ; $skst_{1-\alpha,v,\xi}$ is the corresponding right quantile (see Eq. (5.5)). The quantile function of the skewed Student has been derived in Eq. (5.5). As formally defined in the previous chapter, the long side of the daily VaR is defined as the VaR level for traders having long positions in the relevant equity index: this is the “usual” VaR where traders incur losses when negative returns are observed. Correspondingly, the short side of the daily VaR is the VaR level for traders having short positions, i.e. traders who incur losses when stock prices increase.

6.3.2 Forecasting realized volatility

Regarding the realized volatility, the main findings of Section 6.2 are that the logarithmic realized volatility is approximately normal, appears fractionally integrated and correlated with past negative shocks. To take these properties into account, let us consider the following ARFIMAX(0, d ,1) model (initially developed by Granger, 1980 and Granger and Joyeux, 1980 among others):

$$(1 - L)^{d_a}(\ln RV_t - \mu_0 - \mu_1 y_{t-1} - \mu_2 y_{t-1}^-) = (1 + \theta_1 L)u_t \quad (6.2)$$

$$(1 - L)^{d_a} = \sum_{k=0}^{\infty} \frac{\Gamma(d_a+1)}{\Gamma(k+1)\Gamma(d_a-k+1)} L^k \quad ,$$

where L is the lag operator, μ_0 , μ_1 , μ_2 , θ_1 and d_a are parameters to be estimated, u_t is an *i.i.d.* random process with mean 0 and variance σ_u , $\ln RV_t$ is the logarithm of the realized volatility computed from the intraday returns observed for day t , y_t is the daily return on day t , y_t^- takes the value 0 when $y_t > 0$ and the value y_t when $y_t < 0$. Note that to determine the orders of this ARFIMA model we rely on the AIC criterion.

Estimation of Eq. (6.2) is carried out by exact maximum likelihood (Sowell, 1992) under the normality assumption using ARFIMA 1.0 (see Ooms and Doornik, 1998 and Doornik and Ooms, 1999) and conditional sum-of-squares maximum likelihood (Hosking, 1981) using G@RCH 2.0 (see Appendix A).¹⁵ Due to the strong similarity between the outcomes of the two estimation procedures, we only report the results obtained with the first method.

When $u_t \sim N(0, \sigma_u^2)$, we have by definition that $\exp(u_t) \sim \log N(0, \sigma_u^2)$ (where $\log N$ denotes the log-normal distribution). Thus, the conditional realized volatility (or in-sample one-step-ahead forecast of the volatility) is computed according to:

$$\hat{R}V_{t|t-1} = \exp\left(\ln RV_t - \hat{u}_{t|t-1} + \frac{1}{2}\hat{\sigma}_u^2\right), \quad (6.3)$$

where $\hat{u}_{t|t-1}$ denotes the estimated value of u_t by Eq. (6.2) and $\hat{\sigma}_u^2$ is the estimated variance of u_t in the same equation.

To compute a one-day-ahead forecast for the VaR of the daily returns y_t using the conditional realized volatility, we specify the following AR(3) model:

$$y_t^* = y_t/\sigma_t^* \quad (6.4)$$

$$y_t^* = \mu^* \left(1 - \sum_{i=1}^3 \psi_i^*\right) + \sum_{j=1}^3 \psi_j^* y_{t-j}^* + \varepsilon_t^* \quad (6.5)$$

$$\varepsilon_t^* \sim D(0, \sigma^{2,*}, \kappa^*), \quad (6.6)$$

where now $\sigma_t^* = \sqrt{\hat{R}V_{t|t-1}}$ and μ^* , ψ_1^* , ψ_2^* , ψ_3^* , $\sigma^{2,*}$ and κ^* are parameters to be estimated. κ^* stands for a vector of parameters determining the shape of the density $D(\cdot)$, while $\sigma^{2,*}$ is the variance of ε_t^* . This specification is almost identical to the one introduced in Subsection 6.3.1, but now the conditional volatility for the daily returns is equal to the conditional realized volatility $\hat{R}V_{t|t-1}$. As in Subsection 6.3.1, an adequate distribution for $D(\cdot)$ should be selected. The recent empirical literature has stressed that the normal distribution is a good candidate for $D(\cdot)$ when $\sigma_t^* = \sqrt{RV_t}$, i.e. when one uses realized volatility computed at the end of day t (or ex-post realized volatility). Because we wish to forecast the one-day-ahead VaR, $\sigma_t^* = \sqrt{\hat{R}V_{t|t-1}}$ is substituted to $\sigma_t^* = \sqrt{RV_t}$ in our

¹⁵The finite sample properties of the conditional sum-of-squares maximum likelihood have been investigated by Chung and Baillie (1993).

framework. In Section 6.4, we show that this invalidates the choice of the normal distribution as an adequate distribution for $D(\cdot)$. Therefore, we suggest the use of the skewed Student distribution. For reason of comparison we also present results for the normal distribution.¹⁶ In both cases, the one-day-ahead (demeaned) VaR for long and short positions are given as the product of the quantile at $\alpha\%$ for each distribution with $\sqrt{\hat{R}V_{t|t-1}}$.

6.3.3 Assessing the VaR performance of the models

Using a procedure that is now standard in the VaR literature, we assess the models' performance by first computing their empirical failure rate (both for the left and right tails of the distribution of returns) and then performing the Kupiec LR test presented in Section 5.3.3.

6.4 Empirical application

In this section, we report estimation results for the two models presented in Section 6.3. We first focus on the skewed Student APARCH model which is applied to the daily returns; the second model uses the intraday returns via the computation of the realized volatility. Both models are used to forecast the one-day-ahead VaR for the two stock indexes and their performance is assessed by comparing their empirical failure rate with the theoretical threshold.

6.4.1 VaR, daily returns and the skewed Student APARCH

Our first setting uses daily data only and computes the one-day-ahead daily VaR using these daily observations. The skewed Student APARCH and corresponding one-day-ahead VaR were defined in Subsection 6.3.1. Tables 6.1 (estimated parameters) and 6.2 (assessment of the one-day-ahead VaR) report estimation results when this model is applied to the CAC40 and SP500 daily returns. To simplify the reading of the tables we only report the results concerning the conditional variance equation and the skewed Student density.¹⁷

¹⁶Note that if $D(\cdot)$ is the normal density, then κ^* is a null vector, while the choice of the skewed Student distribution for $D(\cdot)$ implies that $\kappa^* = (\ln(\xi^*), v^*)$.

¹⁷Table 6.1 reports robust standard errors in the sense that the estimates are obtained by approximate QML for a skewed Student pseudo-likelihood.

Table 6.1: Skewed Student APARCH

	CAC40 (daily returns)	SP500 (daily returns)
ω	0.023 (0.013)	0.006 (0.002)
α_1	0.042 (0.015)	0.053 (0.009)
α_n	0.452 (0.193)	0.539 (0.105)
β_1	0.940 (0.018)	0.954 (0.009)
$\ln(\xi)$	-0.075 (0.042)	-0.029 (0.024)
v	12.849 (4.391)	5.462 (0.504)
δ	1.775 (0.568)	0.955 (0.157)
Q_{20}^2	14.75	17.36
$\alpha_1 E(z - \gamma z)^\delta + \beta_1$	0.975	0.991

Estimation results for the volatility specification of the skewed Student APARCH model. Robust standard errors are reported in parentheses. Q_{20}^2 is the Ljung-Box Q-statistic of order 20 computed on the squared standardized residuals.

According to the estimated coefficients for the skewed Student APARCH,

- β_1 is close to 1 but significantly different from 1 for both indexes, which indicates a high degree of volatility persistence.¹⁸ Furthermore both APARCH models are stationary in the sense that $\alpha_1 E(|z| - \gamma z)^\delta + \beta_1$ is lower than 1.
- δ is close to 2 for the CAC40 and close to 1 for the SP500: the APARCH models the conditional variance for the CAC40 and the conditional standard deviation for the SP500;
- γ is significantly positive: negative returns lead to higher subsequent volatility than positive returns (asymmetry in the conditional variance);
- v is much larger for the CAC40 than for the SP500: daily returns defined on the U.S. data display a much larger kurtosis and exhibit fatter tails than returns for the French data;

¹⁸Tse (1998) extended the APARCH by including a pure long memory feature (FIAPARCH). LR tests between the APARCH and the FIAPARCH clearly reject the FIAPARCH specification.

Table 6.2: VaR results for the CAC40 and SP500
(models using daily data)

α	5%	2.5%	1%	0.5%	0.25%
VaR for long positions (CAC40)					
RiskMetrics	0.065	0.005	0	0	0
Skewed Student APARCH	0.764	0.170	0.483	0.609	0.498
VaR for long positions (SP500)					
RiskMetrics	0.335	0.001	0	0.003	0
Skewed Student APARCH	0.682	0.995	0.543	0.962	0.332
VaR for short positions (CAC40)					
RiskMetrics	0.665	0.336	0.671	0.297	0.021
Skewed Student APARCH	0.928	0.336	0.879	0.762	0.948
VaR for short positions (SP500)					
RiskMetrics	0.297	0.152	0.035	0.008	0.060
Skewed Student APARCH	0.625	0.906	0.010	0.024	0.010

P-values for the null hypotheses $f_l = \alpha$ (i.e. failure rate for the long trading positions is equal to α , top of the table) and $f_s = \alpha$ (i.e. failure rate for the short trading positions is equal to α , bottom of the table). α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%. The RiskMetrics and skewed Student APARCH models are estimated on the daily returns (i.e. no use is made of the intraday returns).

- $\ln(\xi)$ is negative in both specifications, albeit not significant from zero for the SP500 and barely significant for the CAC40.¹⁹
- the APARCH dynamical structure succeeds in taking into account all the dynamical structure exhibited by the volatility as the Ljung-Box Q_{20} on the squared standardized residuals is not significant at the 5% level for both models.

For the skewed Student APARCH model, the p-values for the null hypothesis $f_l = \alpha$ (VaR for the left tail of the distribution of returns) and $f_s = \alpha$ (VaR for the right tail of the distribution of returns) given in Table 6.2 confirm that this volatility model succeeds in correctly forecasting the one-day-ahead VaR for most of the probability levels α . Indeed, the p-values are larger than 0.05 for all configurations except the VaR for short positions on the SP500 (with α ranging from 0.25% to 1%). Broadly speaking these results are similar to those of the previous chapter reported for five stock market indexes.

6.4.2 VaR, intraday returns and daily realized volatility

In our second framework we explicitly use the intradaily (5- and 15-minute) returns to compute the daily realized volatility. We first estimate an ARFIMAX(0,d,1) model on the logarithmic realized volatility $\ln RV_t$ as in Eq. (6.2). In a second step, we standardize the daily returns y_t by the one-day-ahead forecast of the realized volatility $\hat{RV}_{t|t-1}$ as in Eq. (6.4) and compute the one-day-ahead VaR using an AR(3) model on the $y_t^* = y_t/\sigma_t^*$. As explained below, the choice of the distribution for $D(\cdot)$ is of paramount importance. Table 6.3 presents estimation results for the ARFIMA specification.

- First, the ARFIMA specification seems to be adequate in modelling the dynamics of $\ln RV_t$. Indeed, the Ljung-Box statistics indicate that all serial correlation in the error term has been removed (at the conventional levels of significance). Parameter d is well above 0 but is not significantly lower than 0.5, indicating

¹⁹This indicates that, at least for the U.S. data, there is no real need for a skewed Student APARCH; nevertheless, as this specification encompasses the simpler Student APARCH, we stick with the more general model (owing to the large number of observations, the loss of degrees of freedom is minimal).

Table 6.3: Asymmetric ARFIMA

	CAC40		SP500	
	5-minute	15-minute	5-minute	15-minute
μ_0	-0.019 (0.913)	-0.016 (0.729)	-0.457 (1.758)	-0.565 (1.120)
μ_1	0.027 (0.023)	0.029 (0.026)	-0.007 (0.017)	-0.016 (0.020)
μ_2	-0.188 (0.040)	-0.187 (0.035)	-0.190 (0.028)	-0.215 (0.034)
θ_1	-0.345 (0.045)	-0.341 (0.053)	-0.237 (0.022)	-0.287 (0.030)
d_a	0.478 (0.025)	0.463 (0.034)	0.492 (0.010)	0.480 (0.019)
σ^2	0.357	0.444	0.289	0.399
Q_{20}	20.0	15.6	22.8	15.9

Estimation results for the logarithm of the realized volatility (defined on 5- and 15-minute returns) using an ARFIMAX(0,d,1) specification: $(1 - L)^{d_a}(\ln RV_t - \mu_0 - \mu_1 y_{t-1} - \mu_2 y_{t-1}^-) = (1 + \theta_1 L)u_t$. Standard errors are reported in parentheses. Q_{20} is the Ljung-Box Q-statistic of order 20 computed on the residuals.

that, in contrast to the GPH test of Subsection 6.2.2, the logarithm of the realized volatility is not covariance-stationary;²⁰

- μ_1 and μ_2 are respectively non significant and significantly positive: negative returns lead to higher subsequent volatility than positive returns (asymmetry in the conditional variance similar to the APARCH model).

Estimation results for the skewed Student AR(3) model are presented in Table 6.4. As in Table 6.1 we do not report the results of the conditional mean in order to save space. As indicated by the Ljung-Box Q_{20}^2 on the standardized residuals of this model, the $y_t^* = y_t / \sqrt{\hat{R}V_{t|t-1}}$ do not display time dependence in volatility. This justifies the use of a skewed Student AR(3) model without ARCH effects. Of course, this is expected as the time dependence in volatility has been captured by the previous ARFIMA model on the dynamics of $\ln RV_t$. In the usual ARCH framework, the $y_t^* = y_t / \sqrt{\hat{R}V_{t|t-1}}$ would play the role of standardized residuals.

²⁰However as argued by Andersson (2000), one has to be careful with the notion of long memory because “(surprisingly) negative moving average parameters (θ_1 is significantly below 0 for both indexes), which alone make no memory contribution, absorb a substantial amount of memory induced by fractional integration.”

This is somewhat true as we do standardize the returns by the square root of forecasted realized volatility.

While the recent literature has stressed that ex-post standardized returns have an almost normal distribution (see Andersen, Bollerslev, Diebold, and Labys, 2000), this is certainly not true for ex-ante standardized returns. The estimated parameters $\ln(\xi^*)$ and v^* reported in Table 6.4 suggest that the ex-post standardized returns of the CAC40 are slightly skewed and kurtosed while the SP500 is kurtosed but symmetric. These results are in line with those reported in Table 6.1 (skewed Student APARCH on daily returns).²¹ Furthermore, assessing the VaR performance of a *normal* model (i.e. choosing the normal distribution for $D(\cdot)$ instead of the skewed Student distribution) for the ex-ante standardized returns gives the results shown in the first line of each cell of Table 6.5:

- for the left tail of the distribution of returns (long VaR), the p-values for the null hypothesis $f_t = \alpha$ are smaller than 0.05 when α is below 1%: the empirical failure rate is significantly higher than α for low VaR levels;
- for the right tail of the distribution of returns (short VaR), the performance of the model is satisfactory;
- there are no real differences between the results for the 5- and 15-minute returns.

However, using the skewed Student distribution instead gives much better results (second line of each cell of Table 6.5). For the CAC40 data, all p-values are larger than 0.05, both for the long and short VaR. For the SP500 data, all p-values are larger than 0.05 except for the short VaR at level $\alpha = 1\%$ and $\alpha = 0.25\%$.

²¹Note that one has to be careful when computing the empirical skewness and the kurtosis on the raw data. Indeed, Table 6.4 also reports these statistics (lines 1 and 2 for both series). For instance, the empirical skewness of the 5-minute (ex-post) standardized returns of the CAC40 and SP500 equal respectively -0.198 and -1.093. To test the departure from normality, it is common to use the t-test $sk/\sqrt{6/T}$ where sk is the empirical skewness and T the number of observations. Based on the result of this test one could be tempted to conclude that the SP500 is highly skewed while the CAC40 is hardly skewed (which contradicts the results obtained with the skewed-Student density, see lines 4 and 5 of Table 6.4). However, as shown by De Ceuster and Trappers (1992) and Peiró (1999), this test is not appropriate when the series is fat-tailed. For a sample size of 2000 observations, De Ceuster and Trappers (1992) tabulate that the 95% confidence intervals of the skewness of Student-t distributed observations with a kurtosis of 3.5 and 18 are respectively $(-0.131; 0.127)$ and $(-0.814; 0.787)$, i.e. the higher the kurtosis, the larger the confidence bands of the skewness.

Table 6.4: Ex-ante standardized returns
(w.r.t. forecasted realized volatility)

5-minute returns		
	CAC40	SP500
Skewness	-0.198	-1.093
Kurtosis	3.537	17.861
$\sigma^{2,*}$	1.073	1.067
$\ln(\xi^*)$	-0.078 (0.042)	-0.020 (0.024)
v^*	14.516 (5.384)	6.055 (0.618)
Q_{20}^2	15.87	3.17
15-minute returns		
	CAC40	SP500
Skewness	-0.167	-1.107
Kurtosis	3.441	18.247
$\sigma^{2,*}$	1.024	1.106
$\ln(\xi^*)$	-0.073 (0.041)	-0.022 (0.024)
v^*	15.708 (6.414)	5.987 (0.606)
Q_{20}^2	15.86	3.01

Descriptive statistics (skewness and kurtosis) and estimation results ($\sigma^{2,*}$, $\ln(\xi^*)$ and v^*) for the skewed Student AR(3) model on the ex-ante standardized returns with respect to the daily realized volatility computed on 5- and 15- minute intraday returns. Q_{20}^2 is the Ljung-Box Q-statistic of order 20 computed on the squared standardized residuals.

Table 6.5: VaR results for the CAC40 and SP500
(models using intraday data and daily realized volatility on 5- and 15-minute returns)

α	5%	2.5%	1%	0.5%	0.25%
VaR for long positions (CAC40)					
Normal	0.323, 0.173	0.087, 0.122	0.026, 0.048	0.018, 0.007	0.021, 0.146
Skewed Student	0.541, 0.265	0.607, 0.607	0.333, 0.482	0.927, 0.927	0.630, 0.499
VaR for long positions (SP500)					
Normal	0.086, 0.032	0.572, 0.816	0.051, 0.035	0.001, 0	0.001, 0
Skewed Student	0.800, 0.800	0.432, 0.917	0.775, 0.948	0.840, 0.657	0.974, 0.694
VaR for short positions (CAC40)					
Normal	0.217, 0.336	0.250, 0.338	0.470, 0.897	0.495, 0.760	0.949, 0.949
Skewed Student	0.461, 0.722	0.444, 0.444	0.470, 0.897	0.760, 0.760	0.499, 0.499
VaR for short positions (SP500)					
Normal	0.059, 0.040	0.575, 0.499	0.948, 0.807	0.963, 0.765	0.332, 0.200
Skewed Student	0.514, 0.739	0.917, 0.740	0.010, 0.052	0.170, 0.051	0.039, 0.010

P-values for the null hypothesis $f_l = \alpha$ (i.e. failure rate for the long trading positions is equal to α , top of the table) and $f_s = \alpha$ (i.e. failure rate for the short trading positions is equal to α , bottom of the table). α is equal successively to 5%, 2.5%, 1%, 0.5% and 0.25%. In each cell, the first number (resp. second number, after the comma) corresponds to the case where the daily realized volatility is estimated on the 5- (resp. 15-) minute returns. Normal and skewed Student refer to the statistical distribution used to model the daily realized volatility.

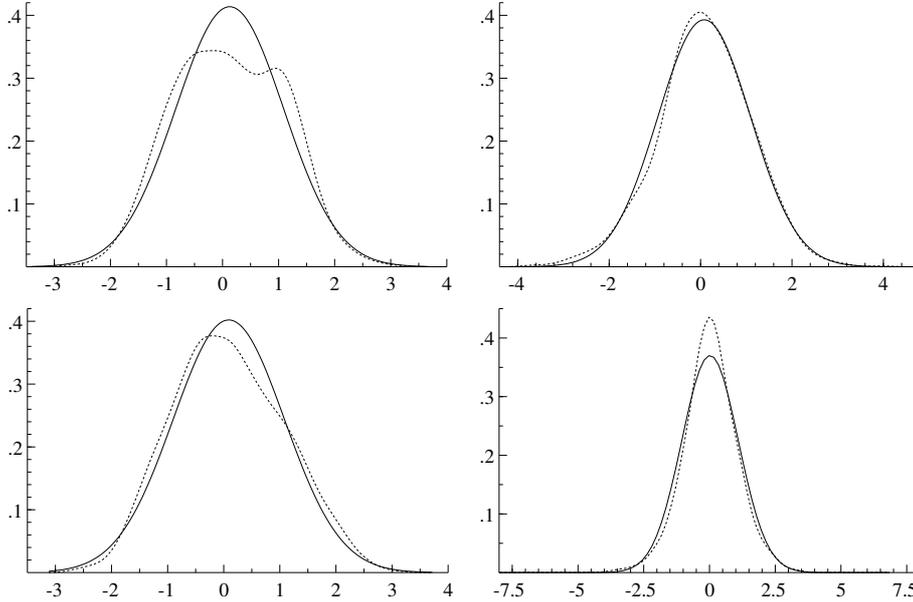


Figure 6.5: The graphs display the density distributions, i.e. empirical (dashed lines) vs normal (solid lines), for the daily returns standardized with respect to the square root of the ex-post (left panel) and the ex-ante (right panel) daily realized volatility computed for the CAC40 (top panel) and SP500 (bottom panel) stock indexes based on 15-minute returns.

Thus the switch from the normal distribution to the skewed Student distribution yields a significant improvement in the VaR performance of the model.

Finally we also give density plots (empirical vs the normal distribution) for the ex-ante and ex-post standardized returns in Figure 6.5. While the tails of the ex-post standardized returns closely track those of the normal distribution, ex-ante standardized returns feature fat tails, especially for the U.S. data. Estimation results and descriptive statistics given in Table 6.4 tell the same story.

6.4.3 Which model is best?

The evidence presented in the two preceding subsections indicates that using an APARCH model with daily data or a two step approach relying on the new concept of realized volatility leads very similar results in terms of VaR. It should be emphasized that to have accurate VaR forecasts, one needs to specify correctly the

full conditional density with *both* methods.

This implies that previous results given in the empirical literature must be qualified. For example, Ebens (1999) concludes his paper by stating that the GARCH model underperforms (when volatility must be forecasted) with respect to the model based on the daily realized volatility. However, the author uses a “simple” GARCH model which neither really accounts for the long memory property observed in the realized volatility nor the fat-tails or asymmetry of the returns (even after standardization). Indeed, when estimating the more simple RiskMetrics VaR model on daily returns (the RiskMetrics model is tantamount to an IGARCH model with pre-specified coefficients, under the additional assumption of normality), we have the VaR results given in Table 6.2: its one-day-ahead forecasting performance is rather poor, especially when α is small.²² With a more “sophisticated” model on the other hand (the skewed Student APARCH model in this chapter), VaR results are much better.

Interestingly and as pointed out in the previous subsection by comparing the results obtained with the normal and skewed Student distributions for the ex-ante standardized returns, the same conclusion is true for the more complex model based on the combination of intraday returns and realized volatility.

6.5 Conclusion

In this chapter we show how to compute a daily VaR measure for two stock indexes (CAC40 and SP500) using the one-day-ahead forecast of the daily realized volatility. The daily realized volatility is equal to the sum of the squared intraday returns over a given day and thus uses intraday information to define an aggregated daily volatility measure. While the VaR forecasts which use this method perform adequately over our sample, we also show that a more simple model based solely on daily returns delivers good results too. Indeed, while the VaR specification based on an ARFIMAX(0, d_a ,1)-skewed Student model for the daily realized volatility provides adequate one-day-ahead VaR forecasts, it does not really improve on the performance of a VaR model based on the skewed Student APARCH model and estimated using daily data. Thus, for the two financial assets considered in an

²²Although the results are not reported in the chapter, we also estimated a normal GARCH(1,1) model and its performance was not much better than the RiskMetrics specification.

univariate framework, the two methods seem to be rather equivalent. Another important conclusion of this chapter is that daily returns standardized by the square root of the one-day-ahead forecast of the daily realized volatility are not normally distributed.

At this stage, one of the most immediate and promising extension of these techniques is to consider corresponding multivariate volatility models to forecast the VaR of a portfolio of financial assets. Multivariate models of the ARCH type are not easy to implement as they often require the estimation of a large number of parameters. Furthermore, these parameters are present in the latent volatility specification and this is one of the main difficulty of the problem. Therefore, multivariate realized volatility models should provide a much easier way to correctly model variances and correlations across financial assets as they assume that volatility is observed. This paves the way for the use of “usual” multivariate models (VAR, ECM) directly applied to realized volatility and correlations.

Chapter 7

Central Bank Intervention and Exchange Rate Volatility: Evidence from a Switching Regime Analysis

7.1 Introduction

Since the beginning of the 90's, the release of high frequency data by several major central banks has led to a renewed interest in the empirical assessment of the effect of direct interventions on the short run evolution of foreign exchange rates. In particular, the empirical literature investigated whether direct purchases and sales made by the central bank on the foreign exchange market could be effective in moving the nominal exchange rate in one direction or another. These sought-after dynamics have been implicitly defined in two well known major international agreements: the 1985 Plaza Agreement that favored central bank cooperation in order to induce a sharp depreciation of the US dollar (USD hereafter) and the 1987 Louvre Agreement that emphasized the need to decrease excess exchange rate volatility. More recently, the interest for direct interventions on the foreign exchange market has been fostered at the European level by the sharp depreciation of the Euro against the major currencies, i.e. the USD and the Japanese Yen (YEN hereafter) and, to a lesser extent, its relatively high volatility. In September 2000, the European Central Bank directly intervened in support of the Euro in coordination with the major other central banks (the Federal Reserve, the Bank of Japan, the Bank of Canada and the Bank of England). This was followed by three

official unilateral interventions carried out in November 2000. Recently, central bank interventions have also been used extensively as an instrument by the Bank of Japan to depreciate the YEN, in order to support its expansive monetary policy.

In the 80's, the inference of the empirical literature was mainly based on the use of quarterly variations of official reserves as proxies to the direct interventions of central banks on the foreign exchange markets. The public release of daily data regarding these direct interventions by the Federal Reserve, the Bundesbank and the Swiss bank (among others) has nevertheless allowed the study of the short-run impact on exchange rates or interest rates. More recently, the Bank of Japan also decided to publish (ex-post) the official interventions made since April 1991. Accordingly, the econometric techniques using these data have been adjusted to account for some of the key features associated with such high frequency financial data (conditional heteroskedasticity for instance).

The results of the empirical literature on foreign exchange rate interventions seem quite surprising. General speaking, there is only some weak evidence that interventions can affect the level of the exchange rate (Baillie and Osterberg, 1997a).¹ When some effects are however detected, net purchases of a particular currency appear to be associated with a subsequent depreciation of this currency (Almekinders and Eijffinger, 1993; Dominguez and Frankel, 1993; Baillie and Osterberg, 1997a and Beine, Bénassy-Quéré, and Lecourt, 2002), suggesting leaning-against-the-wind phenomena.² Regarding the second moment of the distribution of returns, the main findings of the literature emphasize a significant increase of volatility subsequent to the foreign exchange rate interventions. This last effect is extensively documented in the previously quoted papers and also by Connoly and Taylor (1994), Dominguez (1998) and Baillie and Humpage (1992) that use an ex-post characterization of volatility (ARCH and subsequent developments). Focusing on some ex-ante measure of volatility leads to the same conclusion (Bonser-Neal and Tanner, 1996 for instance). All in all, these reported effects raise some doubts on the efficiency of such an instrument, at least in the very short run.

As far as the methodological part of the study is concerned, most of the empirical analyzes use an ARCH-type specification to model the heteroskedasticity

¹Although Baillie and Osterberg(1997b) find some effects on the risk premium in the forward market.

²Leaning-against-the-wind refers to an intervention aiming at reverting the evolution of a particular currency.

observed on these series at a high-frequency basis. For instance, Baillie and Osterberg (1997a,b) as well as Dominguez (1998) use GARCH models while Beine, Bénassy-Quéré, and Lecourt (2002) allow for long memory in the conditional variance through a FIGARCH specification. To study the impact of central bank interventions, explanatory variables are usually added in the conditional mean and/or the conditional variance equations.

In this chapter, we propose an alternative approach to the GARCH specification (Bollerslev, 1986) and the single-regime framework that are commonly used in the empirical literature on the effectiveness of central bank interventions in the foreign exchange markets. In contrast with earlier analysis, we allow for regime-dependent frameworks to assess the impact of direct interventions. More specially, and following the approach proposed by Hamilton (1994), we assume that the evolution of the spot exchange rates depends on a latent regime variable whose dynamics is driven by a first-order Markov switching process (this generalizes the static mixture of normal distributions presented in Chapter 2). Then, in the spirit of Filardo (1994) or Diebold, Lee, and Weinbach (1994), the probabilities of switching from one regime to another depend on strongly exogenous variables, in our case central bank interventions.

Compared to single-regime GARCH type models, one important advantage of such an approach is that it explicitly allows for different outcomes of central bank interventions with respect to the initial state of the economy. For instance, central bank purchases can lead to an increase in volatility when the markets are calm, but not if the market is in a state of high volatility. Similarly, the effect on the level of exchange rate could be different depending on whether the dollar is depreciating or appreciating. The economic rationale is as follows. The literature tends to favor the signalling channel as the prevailing channel of transmission of central bank interventions on foreign exchange rates. As pointed out by Dominguez (1998), according to the intervention signalling hypothesis, the expected effect of an intervention depends on whether its associated signal is unambiguous and consistent with the official goals of these operations. As indicated in Dominguez (1999), the motivations of the FED include influencing trend movements in exchange rates and calming disorderly markets. Therefore, depending on the prevailing state of the market, the signal of an intervention will be ambiguous or not and the effect on the two first moments of exchange rate changes will be different. Our results dealing

with the effects of the central bank interventions on exchange rate volatility turn out to be consistent with this idea.

In this chapter, different Markov switching models are estimated and a selected specification is then used for the study of the DEM/USD exchange rate over the 1985-1995 period. Some evidence is also provided for the YEN/USD in order to assess to which extent our results are only valid for the DEM. Due to data availability, the analysis of the YEN is performed over a shorter period, 1991-1995. It is found that this regime-switching framework fits the data rather well on the one hand, and compares very well with usual GARCH specifications when investigating the respective out-of-sample forecasting properties on the other hand. One of our main conclusions is that official central bank interventions explain a significant part of the observed switches between volatility regimes. Our results lead us to challenge the previous conclusions according to which central bank interventions cannot have any stabilizing influence on the short run dynamics of exchange rates.

The chapter is organized as follows. Section 7.2 investigates the relevance of several statistical models and presents some evidence in favor of a regime-switching model. Section 7.3 is devoted to the analysis of the effects of central bank interventions. Section 7.4 concludes.

7.2 Regime-dependent frameworks

This section introduces the Markov Switching model on which our analysis is based. A comparison with the traditional GARCH model is carried out in order to justify such a regime dependent model. Some statistical model selection search within this class of models is also conducted so that a preferred model can be chosen and extended to time-varying transition probabilities.

7.2.1 Regime-dependent models versus single regime (G)ARCH models

Most of the statistical models used in the literature to study the impact of foreign exchange rate interventions are single-regime models in the sense that the parameters are assumed to be constant over the whole sample. In this chapter, we introduce a more flexible framework by allowing the value of parameters to depend on the prevailing regime. Our data set consists of weekly returns of spot exchange

rates $y_t = 100 \times \ln(p_t/p_{t-1})$, where p_t denotes the number of units of the foreign currency (the DEM or the YEN) per unit of USD. The data has been provided by the Bank of International Settlements. These are mid-day spot exchange rates quoted at Frankfurt at 2:00 pm (DEM) and at Tokyo at 10 am (YEN) in local time. In contrast with the previous literature, we use weekly data rather than daily data. Indeed, it is unclear (and controversial) what is the exact horizon of the central bank interventions. As reported by Neely (2000), an important proportion of central banks believe that the full effect of the intervention is seen over a few days or more. This suggests that the weekly frequency is relevant, at least from the point of view of the central banks. Furthermore, it was implicit that the Plaza and the Louvre agreements focused on lower frequencies than the daily one which is usually considered in the literature. For the DEM, the data ranges from the first week of 1985 to the last one of 1995, yielding 573 observations. This period turns out to include most central bank operations undertaken on the foreign exchange market during the 80's and the 90's. It also corresponds to the period subsequent to the two major agreements in this field, namely the Plaza (September 1985) and the Louvre (February 1987) agreements. For the YEN, given the availability of official central bank interventions of the Bank of Japan, the investigation period ranges from April 1991 to December 1995; this amounts to 272 observations.

To a certain extent, some substitutions are possible between ARCH and regime-switching modelling.³ Although the variance is constant *within* each regime in the latter model, the estimated conditional variance of this model is allowed to vary over time due to the evolution of the probabilistic assessment of being in the first or the second regime. In turn, this suggests that a two-regime model with a constant variance may be an alternative candidate to single-regime (G)ARCH-type models traditionally used in the empirical assessment of central bank interventions. As a starting point, we estimate a two-regime model with shifts allowed both in the conditional mean and variance. Such a framework is proposed by Hamilton (1994). Bollen, Gray, and Whaley (2000) have recently shown that such a model fits the exchange rate data rather well on the one hand and tends to outperform the usual GARCH model on the other hand. In the two-regime case, one has:

³Kim and Kon (1999), Granger and Hyung (1999) or Beine and Laurent (2001) have recently provided some specific evidence on the strong interaction between structural change (captured for instance through regime switching models) and volatility persistence.

$$y_t | \Omega_t \sim \Delta(\mu_1, \sigma_1^2) \text{ if } s_t = 1 \quad (7.1)$$

$$y_t | \Omega_t \sim \Delta(\mu_2, \sigma_2^2) \text{ if } s_t = 2, \quad (7.2)$$

where Ω_t denotes the information set at time t and Δ the Gaussian distribution function. In this framework, the dynamics of y_t is assumed to depend on an unobserved random variable s_t that can take on the values 1 or 2. This unobserved variable is then supposed to follow a first-order Markov process of the type:

$$p_1 = \text{Prob}(s_t = 1 | s_{t-1} = 1) \quad (7.3)$$

$$p_2 = \text{Prob}(s_t = 2 | s_{t-1} = 2). \quad (7.4)$$

In turn, these transition probabilities can be collected in the following P matrix:

$$P = \begin{bmatrix} p_1 & 1 - p_2 \\ 1 - p_2 & p_2 \end{bmatrix}. \quad (7.5)$$

Because of the persistence of each regime (a stylized fact of Markov switching models applied to empirical finance, see for instance Kim and Nelson, 1999) captured by p_1 and p_2 , the model accounts for the volatility clustering observed in the data. Persistence and thus the relevance of the Markov-Switching approach require p_1 and p_2 to be significantly higher than 0.5. This contrasts with single-regime (G)ARCH approaches in which the evolution of the conditional variance is driven by volatility innovations and past values of variances.

Nevertheless, as reported by Bollen, Gray, and Whaley (2000), this two-regime framework imposes some restrictions that can be too strong to capture the dynamics of exchange rates. In particular, since the switching process involves both the mean and the variance, a particular combination of the level of returns and variance of exchange rates is enforced within each regime. For instance, if $\mu_1 > \mu_2$ and $\sigma_1^2 < \sigma_2^2$, the first regime necessarily associates patterns of low volatility with patterns of high returns (appreciation of the USD), while the second regime captures high volatility episodes associated with phases of USD depreciation. Such a restriction can be rejected by the data and thus needs to be tested statistically. As analyzed by Bollen, Gray, and Whaley (2000), the model may be generalized to include independent shifts in the mean and in the variance. In this case, one

has to define two latent variables, $s_{\mu,t}$ and $s_{\sigma,t}$, relative respectively to the mean and to the variance process. As before, each of these two variables is governed by a first-order Markov process. The transition probabilities are denoted by $p_{1,\mu}$ and $p_{2,\mu}$ for the mean process and $p_{1,\sigma}$ and $p_{2,\sigma}$ for the variance one, respectively. This corresponds to a four-regime model with a new latent variable s_t ($s_t = 1, 2, 3, 4$) taking values depending on the mean and variance regimes:

$$y_t | \Omega_t \sim \Delta(\mu_1, \sigma_1^2) \text{ if } s_t = 1 \quad (7.6)$$

$$y_t | \Omega_t \sim \Delta(\mu_2, \sigma_1^2) \text{ if } s_t = 2 \quad (7.7)$$

$$y_t | \Omega_t \sim \Delta(\mu_1, \sigma_2^2) \text{ if } s_t = 3 \quad (7.8)$$

$$y_t | \Omega_t \sim \Delta(\mu_2, \sigma_2^2) \text{ if } s_t = 4. \quad (7.9)$$

In this case, one ends up with a (4×4) matrix of transition probabilities (see for details Bollen, Gray, and Whaley, 2000 or Ravn and Sola, 1995).

The markov-switching regimes are estimated by the Expected Maximum Likelihood (EML) procedure (see Appendix C). In short, the EML estimation relies on the maximisation of the log-likelihood function $\sum_{t=1}^T [Ln(\Delta(y_t | \Omega_t))]$ which is computed from the sum of the log-likelihood values conditional upon each regime:⁴

$$Ln(\Delta(y_t | \Omega_t)) = Ln \left[\sum_{i=1}^S (\Delta(y_t | \Omega_t, s_t = i) \Pr(s_t = i | \Omega_t)) \right], \quad (7.10)$$

where S denotes the total number of regimes (1, 2 or 4 in our analysis) and T is the sample size. One has to be cautious in assessing the relevance of the two-regime model against either the one-regime model or the four-regime model since the standard conditions are not fulfilled to carry out usual likelihood ratio tests (LRT). Several solutions have been proposed (see for instance Hansen, 1992), including the adjustment of critical values proposed by Garcia (1998) for a set of specific two-regime models. When these adjusted critical values are not available, several features, like results from the usual diagnostic tests (Ljung-Box or information criteria for instance) or the forecasting performances, should be computed.

⁴For the estimation of the smoothed probabilities $\Pr(s_t = i | \Omega_t)$, we rely on the algorithm developed by Kim (1994). Similar results have also been obtained with the alternative procedure developed by Gray (1996).

7.2.2 Results and comparison with GARCH model

Before using the Markov switching model to tackle the issue of central bank interventions, the different competing specifications should be compared and assessed and a preferred model should be selected. Tables 7.1 and 7.2 present the results obtained from the various Markov-Switching specifications.

Table 7.1 indicates that the model with two dependent regimes is validated by the data. The one-regime model [model (1)] is clearly rejected in favor of the two-regime model with a switching mean [model (3)] using the χ^2 adjusted critical values provided by Garcia (1998) for this specific model. Indeed, the LRT amounts to 32.672, well above the critical value at the 99% confidence level (17.52). Comparing the four-regime model [model (4)] with model (3), a LRT clearly rejects the hypothesis of independence between mean and variance regime, but once again, because of the identification issue of some parameters under the null hypothesis, one cannot discriminate between these models on these grounds.⁵ Nevertheless, information criteria and other standard diagnostics tend to favor the two-regime model. Basically, the same result holds for the YEN: model (2) with a switching variance and a constant mean turns out to be the preferred model. Another way to discriminate between these regime-switching models but also to compare them with the standard single-regime GARCH model is to investigate their relative out-of-sample forecasting properties. This is done in the next sub-section and will confirm that the four-regime model is clearly dominated.

From the results of model (3), it is also obvious that the estimated models capture volatility regimes rather than mean regimes, which is quite consistent with the literature on Markov-Switching models applied to exchange rates. The first regime is basically the high volatility regime with a variance σ_1^2 roughly three times larger than the one in the second regime (σ_2^2).⁶ By contrast, the two unconditional means do not significantly differ across regimes, neither for the DEM nor the YEN. Restricting the mean to be constant leads to model (2) that can be compared to

⁵It should also be noticed that as emphasized by Garcia (1998), unadjusted critical values tend in general to be too *low*. Therefore, it should be expected that using adjusted critical values would also lead to the rejection of the four-regime model in favor of the two-regime model.

⁶Notice that Tables 7.1 and 7.2 report the estimated standard errors. In turn, this suggests that the variables introduced to explain the transition probabilities in model (3) should be mainly variables thought to influence exchange rate volatility and not the returns. In particular, one should use absolute values of central bank interventions.

Table 7.1: Markov switching models: DEM (1985-1995)

	(1)	(2)	(3)	(4)
μ_1	-0.1381 (0.0685)	-0.1513 (0.0642)	-0.0241 (0.2366)	0.3858 (0.4141)
μ_2	-	-	-0.1758 (0.0790)	-0.3646 (0.1410)
σ_1	1.6417 (0.1033)	2.3853 (0.3025)	2.3872 (0.3106)	2.3624 (0.2962)
σ_2	-	1.2997 (0.1020)	1.2997 (0.1020)	1.2427 (0.1014)
$p_1/p_{1,\mu}$	-	0.8395 (0.0752)	0.8394 (0.0678)	0.8576 (0.1714)
$p_2/p_{2,\mu}$	-	0.9466 (0.0325)	0.9473 (0.0339)	0.9420 (0.0331)
$p_{1,\sigma}$	-	-	-	0.8381 (0.0760)
$p_{2,\sigma}$	-	-	-	0.9446 (0.0333)
Q_{20}	25.8038	25.4923	25.5471	26.8022
Q_{20}^2	31.9849	17.9924	18.1727	18.2120
SIC	3.849	3.826	3.837	3.855
$Log-Lik$	-1096.594	-1080.408	-1080.264	-1079.185

Robust standard errors of maximum likelihood estimates are in parentheses. SIC is the Schwarz information criterion (divided by the sample size) and $Log-Lik$ refers to the log-likelihood value at maximum. Model (1) has constant mean and variance. In Model (2), only the variance switches. In Model (3), the mean and variance switch simultaneously while in Model (4) they can switch independently.

Table 7.2: Markov switching models: YEN (1991-1995)

	(1)	(2)	(3)	(4)
μ_1	-0.1467 (0.0992)	-0.1415 (0.0899)	-0.1822 (0.3884)	0.3736 (0.6928)
μ_2	-	-	-0.1358 (0.1108)	-0.2037 (0.2133)
σ_1	1.6364 (0.2276)	2.3943 (0.3180)	2.3896 (0.3307)	2.3846 (0.3062)
σ_2	-	1.3135 (0.0822)	1.3126 (0.0861)	1.2991 (0.0944)
$p_1/p_{1,\mu}$	-	0.9481 (0.0455)	0.9481 (0.0455)	0.8011 (0.1508)
$p_2/p_{2,\mu}$	-	0.9818 (0.0156)	0.9816 (0.0167)	0.9751 (0.0762)
$p_{1,\sigma}$	-	-	-	0.9469 (0.0473)
$p_{2,\sigma}$	-	-	-	0.9813 (0.0159)
Q_{20}	21.0343	17.1436	17.1834	17.1239
Q_{20}^2	29.3048	10.6958	10.6951	11.8250
SIC	3.860	3.845	3.865	3.906
$Log-Lik$	-519.422	-508.881	-508.872	-508.835

Note: see Table 7.1.

model (3); this restriction is supported by a LRT, which implies that model (2) is finally our preferred model for assessing the impact of interventions on both the mean and variability of exchange rate returns.

Interestingly, the Ljung-Box statistics at lag 20 for the residuals (Q_{20}) and the squared residuals (Q_{20}^2) suggest that the Markov-Switching models are supported by the data. In particular, allowing for a switching variance accounts for the heteroskedasticity present in the data without using the GARCH specification. By contrast, the model does not require a switch in the mean to account for the autocorrelation in the data, as suggested by the Q_{20} statistics for model (2). To illustrate this point and to compare these non-nested specifications, one may investigate the out-of-sample forecasting properties of each model.

7.2.3 Forecasting Performance

We compare the out-of-sample variance forecasts of five volatility models: the GARCH (1,1), the random walk (RW) and three regime switching models (two-regime with constant mean, two-regime with varying mean and four-regime models). The models are estimated for the DEM/USD⁷ using the first 521 observations (up to 1994) with the rest of the data (52 points) left for post-sample forecast evaluation. Variance forecasts at 1, 4 and 8 weeks horizons are constructed for each model.

The volatility forecasts should be compared with the realized variance over the forecast period. The usual measure for the observed volatility in the literature is the square of the returns or the absolute returns (Pagan and Schwert, 1990). However, in a recent paper dealing with daily volatility, Andersen and Bollerslev (1998) have shown that this measure is not fully relevant and have proposed an alternative measure. This new measure uses cumulated squared intradaily returns, also called “realized volatility”, which is a more precise measure of the daily volatility. In our analysis, the realized volatility is defined as:

$$\sigma_t^2 = \sum_{i=1}^5 y_{i,t}^2, \quad (7.11)$$

where $y_{i,t}^2$ is the squared return on day i of week t . For the two-regime and four-regime MS models, the volatility forecast is of course a function of the regime

⁷This experience is not conducted for the YEN/USD due to the small sample size.

probabilities.⁸

To compare the forecasting performances of the different models, we use the following criteria:

- the Root Mean Squared forecast Error (RMSE) generally used in the volatility forecast literature;
- the Heteroskedastic Mean Average Error (HMAE) of Andersen, Bollerslev, and Lange (1999) which is adjusted for ARCH effects;
- the Logarithmic Loss Function (LL) of Pagan and Schwert (1990) as well as Bollerslev, Engle, and Nelson (1994), which stresses the influence of low volatility periods.

The forecast horizon has been set to 1, 4 and 8 weeks. Summary statistics are given in Table 7.3, respectively in panels A, B and C.

Results in Table 7.3 show that the two-regime model with constant mean often leads to a reduction of the variance forecasts errors relative to others models. Such a result is obtained for each forecast length, at least using one criterion. Exceptions are the HMAE and the LL criteria at the one-week horizon and the RMSE criterion at the eight-week horizon. As a whole, it comes out that our preferred model compares very well with the GARCH(1,1) model. More importantly, in almost all cases, the two-regime model clearly outperforms the four-regime model.⁹ This may be due to the fact that the uncertainty regarding the estimates of the mean parameters is quite important in the four regime model. Thus, this makes legitimate the use of the two-regime with constant mean model compared to a GARCH (1,1) specification or to the four-regime model and tends to support the findings drawn from the estimations reported in Tables 7.1 and 7.2. Figure 7.1 plots the conditional variances implied by model (2) and by a GARCH specification. It is seen that, both models give rise to similar episodes of high and low volatility.

⁸See Appendix C for further details.

⁹Except for the HMAE criteria at four-week horizon.

Table 7.3: Variance forecasts for the models

A. One week horizon					
	<i>Two-regime constant mean</i>	<i>Two-regime</i>	<i>Four-regime</i>	<i>GARCH(1,1)</i>	<i>Random Walk</i>
RMSE	4.097	4.099	4.109	4.164	5.135
HMAE	0.401	0.403	0.411	0.388	0.392
LL	1.264	1.273	1.331	1.157	1.461
B. Four week horizon					
	<i>Two-regime constant mean</i>	<i>Two-regime</i>	<i>Four-regime</i>	<i>GARCH(1.1)</i>	<i>Random Walk</i>
RMSE	4.519	4.522	4.750	4.533	5.281
HMAE	0.384	0.385	0.372	0.413	0.392
LL	1.226	1.227	1.275	1.361	1.479
C. Eight week horizon					
	<i>Two-regime constant mean</i>	<i>Two-regime</i>	<i>Four-regime</i>	<i>GARCH(1.1)</i>	<i>Random Walk</i>
RMSE	4.697	4.703	5.191	4.642	5.497
HMAE	0.364	0.365	0.464	0.397	0.391
LL	1.180	1.184	2.050	1.369	1.526

Bold figures highlight the minimal forecast error.

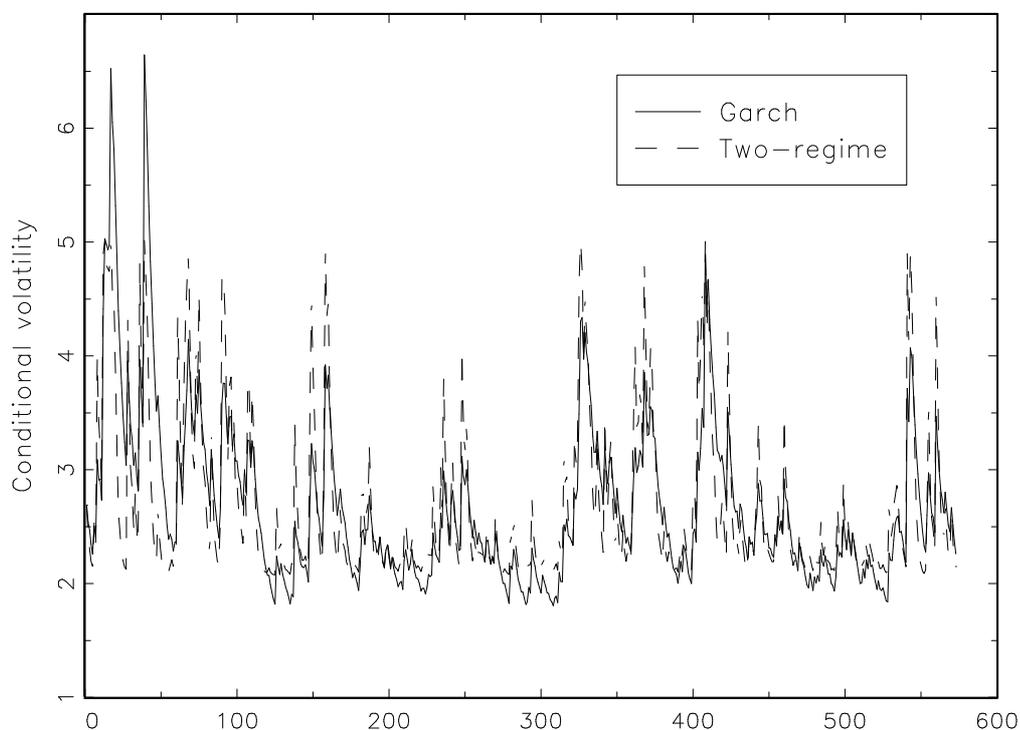


Figure 7.1: Conditional variances: GARCH vs. two-regime model.

7.3 The impact of central bank interventions

7.3.1 The TVTP model

As explained in Section 7.2.1, the change over time of the probabilities of being in one particular regime is in the Markov switching framework the only driving force of the dynamics of the conditional mean and variance of the exchange rate returns. Within each regime, these mean and variance remain constant. Up to now, the transition probabilities of remaining in a particular regime only depend on the previous state of the economy, i.e. the volatility level of past week. To study the impact of central bank interventions on the dynamics of exchange rate returns, we follow Filardo (1994) and Diebold, Lee, and Weinbach (1994) and extend the constant transition probability assumption (see Eq. (7.5)) by conditioning the transition probabilities on exogenous variables (in our case central bank interventions) through a logistic specification. For instance, in the two-regime model

similar to model (2) that involves only volatility regimes, one has:

$$\begin{aligned} p_{1,t} &= Prob(s_t = 1 \mid s_{t-1} = 1, |x_{t-1}|) \\ &= 1 - \left[1 + \exp(\eta_{1,0} + \sum_{i=1}^k \eta_{1,i} |x_{i,t-1}|) \right]^{-1} \end{aligned} \quad (7.12)$$

$$\begin{aligned} p_{2,t} &= Prob(s_t = 2 \mid s_{t-1} = 2, |x_{t-1}|) \\ &= 1 - \left[1 + \exp(\eta_{2,0} + \sum_{i=1}^k \eta_{2,i} |x_{i,t-1}|) \right]^{-1}, \end{aligned} \quad (7.13)$$

where x_t is a matrix of k explanatory variables, i.e. $x_t = (x_{1,t}, \dots, x_{k,t})$. In our framework, these explanatory variables are of course the central bank interventions. In the subsequent estimations we use $k = 1$ when dealing with coordinated interventions and $k = 2$ with unilateral interventions.

We use model (2) and also introduce interventions as explanatory variables of the conditional mean of exchange rate returns. This implies that we allow only for linear effects on the returns:

$$y_t = \mu + \sum_{i=1}^k \varpi_i x_{i,t-1} + \varepsilon_t. \quad (7.14)$$

By contrast, since interventions influence the transition probabilities of volatility regimes, they should be introduced in a non-linear way in the conditional variance specification. Filardo (1998) provides the necessary conditions to ensure that the estimation of TVTP with a ML procedure is possible and relevant. According to the main condition of Filardo (1998), the explanatory variables should be conditionally uncorrelated with the latent regime variable (s_t). Thus one should check that the central bank interventions are not caused in a systematic way by the level of exchange rate volatility. From an econometric point of view, this is similar to the well-known simultaneous bias problem which has been investigated in the literature of central bank interventions. In this respect, evidence presented in the literature is rather mixed: regarding the mean, central banks tend to lean against the wind (Almekinders and Eijffinger, 1993; Dominguez, 1998; Baillie and Osterberg, 1997b and Beine, Bénassy-Quéré, and Lecourt, 2002). In other terms, it is the tendency to depreciate rather than the mere previous change in the level that matters. Concerning volatility, the results appear rather mixed. Baillie and

Osterberg (1997a) find that volatility caused interventions on the 1985-1991 period. Nevertheless, using another measure of conditional variance over the same period, Beine, Bénassy-Quéré, and Lecourt (2002) find less evidence according to which volatility levels motivate the intervention of the major central banks, at least for the DEM. As a whole, it turns out that the condition of non-causality from the current state of the market to the central bank interventions is far from being fulfilled. As a result, one should use one-week lagged interventions ($|x_{i,t-1}|$) rather than the contemporaneous ones ($|x_{i,t}|$) in the TVTP in order to ensure that such a simultaneous bias does not occur. Given that we work with volatility regimes, both specifications are used to assess the robustness of the results.¹⁰ Before proceeding to the ML estimation, we describe the central bank intervention data.

7.3.2 The intervention data

Our data consists of weekly official central bank interventions of the Federal Reserve (FED) and the Bundesbank (BB) on the DEM/USD market over the 1985-1995 period and the interventions of the Federal Reserve (FED) and the Bank of Japan (BoJ) on the YEN/USD market over the 1991-1995 period. As in Bonser-Neal and Tanner (1996), Dominguez and Frankel (1993) or Dominguez (1998), we distinguish between the nature of these interventions.

First, we use discrete variables focusing on the number of official intervention days rather than on the (cumulated) amounts of daily interventions. Basically, this allows us to assess the influence of the presence of the banks in the markets, and emphasizes the signalling channel of interventions rather than the basic portfolio effect. Table 7.4 provides the number of (official) intervention days for each central bank.¹¹ The number of coordinated interventions is also given. Two interventions are said to be coordinated if they happen on the same day and in the same direction. For the DEM, we take FED interventions at day $t-1$ but Bundesbank interventions at time t in order to account for time lags between the markets. For the YEN, we

¹⁰This is especially important in the DEM case. For the YEN, all results emphasize some causality from exchange rate volatility to interventions (see Beine, Bénassy-Quéré, and Lecourt, 2002 for details). Not lagging these interventions would definitely result in endogeneity biases.

¹¹Table 7.4 provides the number of official and reported interventions. Reported interventions are obtained from reports extracted from the financial newspapers (we are grateful to K. Bonser-Neal for providing the reported interventions on the DEM market over the 1985-1991 period). Given the important discrepancy between reported and official interventions (see for instance the reported interventions for the YEN), we prefer to focus on official interventions.

Table 7.4: Official and reported central bank interventions, number of days

	Observations	Total number of daily interventions (DEM/USD, 1985-1995)		Coordinated
		FED	BB	
Official	2868	215	264	97
Reported	2868	184	161	-
		Total number of daily interventions (YEN/USD, 1991-1995)		
		FED	BoJ	
Official	1445	16	159	15
Reported	1445	15	22	-

consider FED and BoJ interventions at day $t - 1$.¹²

Because the number of coordinated interventions is large, one may expect that the weekly intervention data will be highly correlated. Table 7.5 confirms that, in the case of the DEM, the correlation between interventions measured through discrete variables, both in levels¹³ and in absolute value (used in the conditional volatility specification) is very high.¹⁴

Such a high correlation would give rise to multicollinearity problems and poor estimates of standard errors. To account for this problem, we isolate unilateral interventions, i.e. interventions made by a single central bank on a particular day. The cross correlations between these adjusted interventions given in Table 7.6 show that the correlations have dramatically decreased and thus multicollinearity should not be a problem in our estimations. We run two types of regressions with discrete variables: the first one relies only on the unilateral interventions while the second one uses only the coordinated interventions. This distinction makes sense from an economic point of view as some authors have argued that the effect of coordinated interventions has more strength than the one obtained by unilateral ones (see among others Catte, Galli, and Rebecchini, 1992; Dominguez and Frankel, 1993

¹²The German market is six hours ahead of the US market and lags the Japanese market by 8 hours.

¹³In this case, the variable is trinomic: -1 indicates that the bank is selling dollars, 0 means that the bank does not intervene and 1 that the bank is buying dollars.

¹⁴Similar results are also obtained for the YEN (although the problem seems less important given the lower proportion of coordinated interventions. These results are not reported in order to save space.

Table 7.5: Cross correlations between central bank interventions

Discrete variables (DEM/USD, 1985-1995)							
		Levels			Absolute values		
		BB	FED	Coord	BB	FED	Coord
Levels	BB	1	0.647	0.769	-	-	-
	FED		1	0.770	-	-	-
	Coord			1	-	-	-
Absolute values	BB				1	0.594	0.755
	FED					1	0.751
	Coord						1

Table 7.6: Cross correlations between central bank interventions

Discrete variables (unilateral) (DEM/USD, 1985-1995)							
		Levels			Absolute values		
		BB	FED	Coord	BB	FED	Coord
Levels	BB	1	0.208	0.346	-	-	-
	FED		1	0.253	-	-	-
	Coord			1	-	-	-
Absolute values	BB				1	0.113	0.289
	FED					1	0.183
	Coord						1

and Weber, 1996).

7.3.3 The results

Tables 7.7 and 7.8 report the estimation results for the DEM and the YEN respectively. In both cases, the two-regime specification with a constant conditional mean is used. In these models, central bank interventions enter linearly the conditional mean equation. The official central bank interventions are modelled using dummy variables giving the number of intervention days over a particular week. For both currencies, we study the effect of coordinated and unilateral interventions.¹⁵

¹⁵However, in the case of the YEN, it is impossible to consider the effect of unilateral interventions of the FED, given that there is only one occurrence over the considered period (see also Table 7.4). This unilateral intervention occurred on the May 24th, 1993.

7.3. THE IMPACT OF CENTRAL BANK INTERVENTIONS

Table 7.7: Central bank interventions, DEM (1985-1995)

	Discrete variables, official interventions		
	Coordinated	Coordinated (no lag)	Unilateral
μ	-0.1475 (0.0663)	-0.1627 (0.0656)	-0.1760 (0.0696)
ϖ_1 [Coord/BB]	-0.0971 (0.0964)	-0.1256 (0.1438)	-0.2229 (0.0898)
ϖ_2 [FED]	-	-	0.1398 (0.0919)
σ_1	2.3100 (0.3186)	2.2150 (0.2000)	2.3644 (0.4295)
σ_2	1.2771 (0.0975)	1.1834 (0.1173)	1.2848 (0.0927)
$\eta_{1,0}$	2.1562 (1.1973)	1.5501 (0.6535)	2.3029 (1.4194)
$\eta_{1,1}$ [Coord/BB]	-1.9778 (0.9440)	-2.8840 (1.2807)	-0.5312 (0.4362)
$\eta_{1,2}$ [FED]	-	-	-0.4257 (0.5804)
$\eta_{2,0}$	3.3381 (0.9334)	2.4328 (0.7621)	3.5558 (0.8762)
$\eta_{2,1}$ [Coord/BB]	-2.1755 (0.8356)	-15.2901 (2.3933)	-0.3774 (0.3960)
$\eta_{2,2}$ [FED]	-	-	-0.5916 (0.4626)
p_1	0.8961 (0.1113)	0.8249 (0.0944)	0.9091 (0.1172)
p_2	0.9651 (0.0305)	0.9193 (0.0565)	0.9722 (0.0236)
Q_{20}	26.1531	27.1567	25.8283
Q_{20}^2	18.5289	19.9721	18.5653
SIC	3.854	3.849	3.873
$Log-Lik$	-1078.676	-1077.274	-1074.855

Robust standard errors of maximum likelihood estimates are in parentheses. SIC is the Schwarz information criterion (divided by the sample size) and $Log-Lik$ refers to the log-likelihood value at maximum. $y_t = \mu + \sum_{i=1}^k \varpi_i x_{i,t-1} + \varepsilon_t$, $p_{s,t} = 1 - [1 + \exp(\eta_{s,0} + \sum_{i=1}^k \eta_{s,i} |x_{i,t-1}|)]^{-1}$, $p_s = 1 - [1 - \exp(\eta_{s,0})]^{-1}$ and $s = 1, 2$. For coordinated interventions, $x_{1,t}$ stands for the number of official intervention days; for unilateral interventions, $x_{1,t}$ and $x_{2,t}$ stand respectively for the number of official intervention days of the Bundesbank [BB] and of the Federal Reserve [FED]. Column labelled “Coordinated (no lag)” refers to estimations of $p_{s,t}$ based on $|x_{i,t}|$ rather than $|x_{i,t-1}|$.

Table 7.8: Central bank interventions, YEN (1991-1995)

Discrete variables, official interventions		
	Coordinated	Unilateral
μ	-0.1539 (0.0896)	-0.1661 (0.1520)
ϖ_1 [BoJ/Coord]	-0.6426 (0.4304)	0.0122 (0.0903)
σ_1	2.7379 (0.5599)	2.5407 (0.9431)
σ_2	1.3464 (0.0802)	1.3117 (0.1232)
$\eta_{1,0}$	2.1913 (1.8748)	2.7183 (2.7587)
$\eta_{1,1}$ [BoJ/Coord]	-12.9226 (1.6359)	-1.4253 (13.7202)
$\eta_{2,0}$	3.8605 (0.8717)	3.6524 (2.6839)
$\eta_{2,1}$ [BoJ/Coord]	-3.0663 (1.6830)	-0.5383 (0.3429)
p_1	0.8995 (0.1695)	0.9380 (0.1602)
p_2	0.9794 (0.0176)	0.9747 (0.0661)
Q_{20}	18.0481	17.8551
Q_{20}^2	20.1797	19.2777
SIC	3.892	3.898
$Log-Lik$	-506.868	-507.742

Robust standard errors of maximum likelihood estimates are in parentheses. *Log-Lik* refers to the log-likelihood value at maximum. $y_t = \mu + \sum_{i=1}^k \varpi_i x_{i,t-1} + \varepsilon_t$, $p_{s,t} = 1 - [1 + \exp(\eta_{s,0} + \sum_{i=1}^k \eta_{s,i} |x_{i,t-1}|)]^{-1}$, $p_s = 1 - [1 - \exp(\eta_{s,0})]^{-1}$ and $s = 1, 2$. For coordinated interventions, $x_{1,t}$ stands for the number of official intervention days; for unilateral interventions, $x_{1,t}$ stands for the number of official intervention days of the Bank of Japan [BoJ].

Basically, our model is in agreement with the literature as far as the conditional mean of exchange rate returns is concerned. This is not surprising since the basic specification (i.e. linear impacts of the interventions) is consistent with the previously adopted approaches: the Bundesbank purchases of dollars lead to a subsequent depreciation of the USD, which is also documented in Almekinders and Eijffinger (1993), Dominguez and Frankel (1993), Baillie and Osterberg (1997a) and Beine, Bénassy-Quéré, and Lecourt (2002). Baillie and Humpage (1992) interpret this result as a smoothing effect, suggesting that the depreciation might have been even sharper without such an intervention. The FED interventions do not give similar results, at least on the 1985-1995 period.¹⁶ The results for the YEN suggest that coordinated interventions or unilateral operations of the BoJ have a limited impact on the exchange rate returns.

Our results present a quite different view regarding the effects of interventions on exchange rate volatility. In contrast with the single regime GARCH framework, our regime-dependent specification allows us to account explicitly for the initial state of the market in which a specific intervention occurs. Almost all regression results of Tables 7.7 and 7.8 clearly show that when the market is in the low volatility state, central bank interventions tend to increase volatility (see estimates of $\eta_{2,i}$ ($i = 1, 2$)). For instance, when $\eta_{2,1}$ is significantly negative, this means that coordinated interventions tend to reduce the probability of remaining in the low volatility regime and thus tends to increase exchange rate volatility. Our results also suggest that the unilateral interventions had less power than coordinated ones in “moving” the markets. This tends to be consistent with the main results of the literature.

Nevertheless, it is also found that, when the market is quite volatile (i.e. when the high volatility regime prevails), direct coordinated interventions can have a stabilizing impact. In the second column of Tables 7.7 and 7.8 (labelled “Coordinated”), the $\eta_{1,1}$ parameter is negative and significant at the 5% level. To a certain extent, such a result is fairly new in the literature.¹⁷ Furthermore, it holds

¹⁶Beine, Bénassy-Quéré, and Lecourt (2002) obtain different results across sub-periods concerning the effects of the FED interventions on the conditional mean. While the full period (1985-1995) is associated with positive signs (albeit not always significant), the estimations relative to the 1985-1991 sub-period yield negative signs (net purchases associated to a depreciation).

¹⁷Note that this dampening effect of central bank intervention is also found by Murray, Zelmer, and McManus (1996). They show that this effect is specific to some circumstances (including the size of the intervention) but do not make any distinction concerning the prevailing level of

for *both* pairs of currencies. As suggested by the results reported in the third column of Table 7.7 (labelled “Coordinated (no lag)”), this stabilizing impact is robust to the choice of the one-week lagging procedure whose goal is to account for the potential endogeneity problem.¹⁸ Quite interestingly, this stabilizing impact occurs in the case of coordinated interventions only when the high volatility regime prevails. It should be stressed that such a result is fully consistent with the signalling approach presented in Dominguez (1998) who shows that an intervention can reduce exchange rate volatility only if such an intervention is credible and its associated signal is unambiguous. If the intervention occurs in the high volatility regime, the objective of reducing exchange rate volatility is best understood by the market, especially subsequent to the Louvre Agreement which was made public in 1987. By contrast, when the market is less volatile, the signal associated to the intervention is more ambiguous and the resulting effect on exchange rate volatility is definitely positive, a case clearly identified in the signalling approach. Another interpretation involves the traded amounts on the market. Indeed, volatility and traded volumes on the market are often related (see for instance MacDonald (2000) on this point). Furthermore, trading volumes reflect the amount of information processed by the market. This could suggest that the way central bank interventions affect the behavior of market participants depends on market activity and the amount of information flows. These findings are also in agreement with the recent results of Mundaca (2001) in the special case of the interventions carried out by the bank of Norway. Indeed, Mundaca (2001) shows that the direct interventions carried out by the Bank of Norway were stabilizing when they occurred while the exchange rate was moving around the central parity of the currency band rather than near the weakest edge of this band, and thus when the objective was to decrease exchange rate volatility rather than to support the level of the exchange rate.

Moreover, it should be noticed that the size of these effects can be substantial. For example, in the case of the DEM, if both central banks intervene once on a particular week in a concerted way whereas the market is in the high volatility regime, the probability of remaining the next week in this regime drops from

volatility.

¹⁸In contrast with the DEM, for the YEN, previous empirical evidence emphasizes this simultaneity problem even on the volatility side.

89.62% to 54.4%; in other words, the expected number of weeks of high volatility in this market drops from 9.62 weeks (more than two months) to 2.19 weeks.¹⁹ *Ceteris Paribus*, when both banks intervene three times during the same week, the probability of remaining in a high volatility regime falls below 3%. These computations of course assume that the marginal effect of one additional intervention during a particular week is constant on the logistic scale. When two concerted interventions occur the same week on the DEM/USD market, the probability of remaining in the high volatility regime amounts to 14.19%. This probability is less than 1% when four coordinated interventions are made in the same week. In our dataset, we observe respectively 4 weeks with 4 concerted interventions, 7 weeks with 3 concerted ones and 14 weeks with 2 coordinated interventions.

Our results also shed an interesting light on the results found in the literature. As illustrated by Baillie and Osterberg (1997 a,b), all studies emphasize either a positive impact or no effect of interventions on exchange rate volatility. Single regime specifications cannot account for the initial state of the market. As a result, the estimates of the effect of the central bank interventions tend to correspond to an average effect. Because the occurrences of the low volatility regime are more frequent (i.e. $\hat{p}_1 < \hat{p}_2$ or equivalently $\hat{\eta}_{1,0} < \hat{\eta}_{2,0}$ for both exchange rates), single regime estimates tend to be driven by the effects related to this regime. Our results confirm that these impacts are definitely positive. Next to this, it is found that the effect of coordinated interventions differs drastically from the effect of unilateral interventions. While coordinated interventions influence the volatility patterns of the DEM and the YEN exchange rates, unilateral interventions do not seem to be effective in “moving” the markets. These results are in agreement with the results obtained by several authors including Catte, Galli, and Rebecchini (1992), Dominguez and Frankel (1993) or Weber (1996).

7.4 Conclusion

In this chapter we study the impact of weekly central bank interventions on the level and the volatility of the DEM/USD and YEN/USD exchange rate returns. In contrast to the usual literature which favors GARCH-type specifications, we rely

¹⁹The $\eta_{1,0}$ and $\eta_{2,0}$ parameters are expressed on the logistic scale. Given p_{ii} , the expected value of the number of periods with prevailing regime i is equal to $\frac{1}{1-p_{ii}}$.

on a regime dependent approach. Because of this new feature, the interventions can have different outcomes depending on the prevailing state of the market. Our estimations suggest that the dynamics of both series is mainly driven by volatility regimes (a high and a low volatility regime). Thanks to out-of-sample forecasting experiments, it is shown that this specification compares very well with GARCH models and thus offers a relevant statistical alternative to the usual methodology presented in the literature.

Our results partly confirm the positive impact of central bank interventions on exchange rate volatility emphasized in the literature. Nevertheless, it is found for *both* the DEM and the YEN that when the market is highly volatile and when market participants expect the central banks to intervene, concerted interventions can have a stabilizing effect. This new result in the empirical literature is consistent with the signalling approach to central bank interventions on the foreign exchange market. It is also consistent with the 1987 Louvre Agreement objective of decreasing excess volatility of exchange rate through direct coordinated interventions. Such a result also sheds an interesting light on previous results obtained with a “single regime” specifications. By not taking into account the volatility regime in which the interventions occur, these models tend to favor the impact observed in the most prevailing state of the market, i.e. the low volatility one.

Regarding economic policy issues, our results have two important implications. First, they confirm previous results according to which coordinated rather than unilateral interventions lead to large effects in the currency market. Second, our findings suggest that the signal sent to market participants through central bank interventions and hence its impact on exchange rates crucially depends on the current state of the market and the perceived motivation to intervene. This supports a more transparent intervention policy by central banks.

Chapter 8

Conclusion

Modelling high-frequency financial time-series is far from being obvious. These data have several properties that make the use of traditional regression tools inappropriate. Indeed, it is well known that exchange rate returns and stock returns (among others), recorded on an intra-daily, daily and weekly basis are in general serially correlated and often heteroscedastic, fat-tailed and even skewed.

When building a model, it is thus of primary importance to account for these stylized facts. In this respect, many researchers follow Engle (1982) and choose an ARMA specification for the conditional mean and an ARCH-type model for the conditional variance.

To estimate the resulting models, it is convenient to maximize the associated log-likelihood function. Consequently, one has to make an additional assumption about the distribution of the innovation process. Even if this hypothesis is unrealistic in practice, the normality assumption may be justified by the fact that the Gaussian QML estimator is consistent assuming that the conditional mean and the conditional variance are correctly specified (Weiss, 1986; Bollerslev and Wooldridge, 1992). The price to pay for this nice property is that this method is not efficient, the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera, 1991).

There is no doubt that searching for a more suitable distribution is crucial to gain in efficiency. However, the other side of the coin, is that wrongly assuming that the innovations are, for instance, Student-t distributed (when they are skewed) will provide biased estimates of the conditional mean and conditional variance. As a consequence, we have to be very cautious with the choice of the density. It is

thus important to check its appropriateness for the data to be analyzed.

What do we learn from this thesis ?

The main objective of this thesis was to find a conditional density able to replicate the stylized facts enumerated above. In this respect, we have proposed to extend the skewed Student density of Fernández and Steel (1998) in two directions.

First, from a technical point of view, we have reexpressed this density to have mean zero and unit variance innovations. We have shown that the main advantages of this technique are threefold:

- the skewed Student is easy to implement because its pdf, cdf and inverse cdf are linked to the corresponding functions of its symmetric counterpart (which are available in most statistical packages) and the score vector of this density is fairly simple to obtain (which can provide more accurate estimations and highly speed-up the estimation procedure);
- the additional parameters have a clear interpretation;
- and more importantly, it is validated by the data for all the series we have investigated. For instance, we have shown that the use of the skewed Student density is very promising in Value-at-Risk applications. Indeed, unlike the normal and Student densities, the skewed Student (coupled with an AR-APARCH model) provided fairly accurate Value-at-Risk forecasts for the investigated series.

One possible cause for the fact that, in general, estimated residuals from an ARCH-type model still have large excess kurtosis and excess skewness is that few observations on returns are so-called Additive Outliers (AO), which are not captured by a standard ARCH model. Using and extending the approach proposed by Franses and Ghijsels (1999) to detect and correct the AO in a GARCH model, we have shown that for a sample of 2000 daily observations of the NASDAQ stock index, more than 70 have been characterized as AO in the variance. We have also shown that this large number of AO is primary responsible for the excess kurtosis but not for the skewness.

Second, we have proposed a multivariate generalization of this family of skewed densities. This offers a practical and flexible solution to introduce skewness in multivariate symmetrical distributions. Applying this procedure to the multivariate Student density leads to a “multivariate skewed Student” density, for which each marginal has a different asymmetry coefficient. Similarly, when applied to the product of independent univariate Student densities, it provides a “multivariate

skewed density with independent Student components” for which each marginal has a different asymmetry coefficient and number of degrees of freedom. Combined with a multivariate GARCH model, this new family of distributions is potentially useful for modelling stock returns. In an application on the NASDAQ and the DAX, on a daily basis, these densities were found to outperform their symmetric competitors (the multivariate normal and Student).

Third, the notion of realized volatility has been introduced recently in the literature by Taylor and Xu (1997) and Andersen and Bollerslev (1998). According to these authors the realized volatility, computed as an aggregated measure of volatility defined on intraday returns, offers an “error free” measure of the daily volatility. Interestingly, when one uses the realized volatility instead of the conditional variance produced by a parametric ARCH-type model, the normality assumption on the innovation process is supported. It is thus natural to question the relevance of the approach adopted in this thesis. Does the use of the realized volatility invalidate the choice of a skewed Student density ? The answer is obviously not. Indeed, the realized volatility is not a forecast but a realization of the observed volatility. When using a parametric model to produce a forecast of the realized volatility, the results obtained are very similar to the ones given by the standard ARCH approach. Using 5-minutes returns of the CAC40 and the SP500 we have shown how to compute a daily VaR measure based on a one-day-ahead forecast of the realized volatility. When relying on the normality assumption this technique was found to produce very bad VaR forecasts, similar to the ones produced by an APARCH model on daily data with a Gaussian log-likelihood. However, when coupled with a skewed Student density, the realized volatility produced satisfactory results that are nearly equivalent to the ones given by a skewed Student APARCH model. Our main results can be summarized in one sentence: yes, an (adequate) ARCH type model can deliver accurate VaR forecasts and this model performs as well as a competing VaR model based on the realized volatility. The key issue is to use a model that clearly recognizes the full features of the empirical data such as a high kurtosis and skewness in the observed returns (a skewed Student density for instance).

Fourth, it is widely accepted that there is no consensus in the literature in favor of a leading ARCH model (see Palm, 1996 among others). Instead, there is a large number of alternative specifications. Which model to select for our

data ? A GARCH, a FIGARCH, an APARCH, ... ? If a GARCH model is appropriate, do we prefer a GARCH(1,1), a GARCH(1,2), ... ? To answer these questions, a researcher is likely to estimate several candidate models, with different lag orders and perhaps different log-likelihood functions. The most challenging part of this thesis was to develop a package dedicated to the estimation and forecast of several of the most popular univariate ARCH-type models. The package, called G@RCH, has been developed with the Ox 3.0 matrix programming language of Doornik (1999) and offers a friendly dialog-oriented interface similar to the well known software PcGive. It is free of charge and can be used on several platforms, including Windows, Unix, Linux and Solaris. For most of the specifications, it is generally very fast and its main characteristic is its ease of use. To investigate the numerical accuracy of several econometric softwares, including our package, we have compared the estimation results of a GARCH (1,1) with respect to a benchmark, i.e. the results provided by Fiorentini, Calzolari, and Panattoni (1996) for the same dataset. Indeed, these authors estimate this model relying on the analytic hessian (and provide a FORTRAN procedure to replicate their results). To conclude, even if G@RCH uses numerical scores in the estimation procedure, it gives very satisfactory results, unlike EViews for instance, which is found to give the worst numerical accuracy.

The last contribution of this thesis was to investigate the effect of central bank interventions on the weekly returns and volatility of the DEM/USD and YEN/USD exchange rate returns (at weekly frequency). In contrast with previous analyzes, we allowed for regime-dependent specifications (an extension of the normal mixture presented in the third chapter) and investigated whether official interventions may explain the observed volatility regime switches. The estimation results shed an interesting light on the conclusions given in the literature. It is found that depending on the prevailing volatility level, coordinated central bank interventions can have either a stabilizing or a destabilizing effect. Our results lead us to challenge the usual view that such interventions are necessarily associated with increases in volatility.

Appendix A

G@RCH 2.0: An Ox Package for Estimating and Forecasting Various ARCH Models

A.1 Introduction

Well known statistical packages such as Eviews, Gauss, Matlab, Microfit, PcGive, Rats, SAS, S-Plus and TSP provide various options to estimate sophisticated econometric models in very different areas such as cointegration, panel data, limited dependent model, etc. It has been shown at the beginning of this thesis that to fully account for the characteristics of high-frequency financial returns we need to specify a model in which the conditional mean and the conditional variance may be time-varying. It is common to use an ARMA structure in the first conditional moment and an ARCH-type model for the second conditional moment. It has also been shown that relying on a non-normal assumption for the innovation process sometimes provides much more efficient estimates (at least asymptotically) than the Gaussian QML estimator.

A researcher is thus facing the problem of the specification choice. Which model to select ? And which selection criterion to use ? It is not our goal to answer these questions. However, it is almost sure that this researcher is going to estimate several candidate models, with different lag orders and perhaps different log-likelihood functions.

The aim of this appendix is to provide a package dedicated to the estimation and forecast of various univariate ARCH-type models. Contrary to the software's

mentioned above, G@RCH 2.0 is only concerned with ARCH-type models (Engle, 1982), including some recent contributions in this field such as the GARCH (Bollerslev, 1986), EGARCH (Nelson, 1991), GJR (Glosten, Jagannathan, and Runkle, 1993), APARCH (Ding, Granger, and Engle, 1993), Integrated GARCH (IGARCH, see Engle and Bollerslev, 1986) but also FIGARCH (Baillie, Bollerslev, and Mikkelsen, 1996a and Chung, 1999), Hyperbolic GARCH (HYGARCH, see Davidson, 2001), Fractionally Integrated EGARCH (FIEGARCH, see Bollerslev and Mikkelsen, 1996) and Fractionally Integrated APARCH (FIAPARCH, see Tse, 1998) specifications of the conditional variance and an AR(FI)MA specification of the conditional mean (Baillie, Chung, and Tieslau, 1996, Tschernig, 1995, Teyssière, 1997, Lecourt, 2000 or Beine, Laurent, and Lecourt, 2000). This package provides a lot of features, including two standard errors estimation methods (Approximate Maximum Likelihood and Approximate Quasi-Maximum Likelihood) for four distributions (normal, Student- t , GED or skewed Student- t). Moreover, explanatory variables can enter the mean and/or the variance equations. Finally, h -step-ahead forecasts of both the conditional mean and variance are available as well as many misspecification tests (Nyblom, SBT, Pearson goodness-of-fit, Box-Pierce,...).

Our package has been developed with the Ox 3.0 matrix programming language of Doornik (1999).¹ It can be used on several platforms, including Windows, Unix, Linux and Solaris. For most of the specifications, it is generally very fast and its main characteristic is its ease of use. G@RCH 2.0 may be downloaded from the web site <http://www.egss.ulg.ac.be/garch/>.

Two versions of the program are available and called the “Light Version” and the “Full Version”, respectively. The “Full Version” offers a friendly dialog-oriented interface similar to PcGive and some graphical features by using OxPack, a GiveWin batch client module. This version requires a registered version of Ox and GiveWin.

The “Light Version” is launched from a simple Ox file. It does not take advantage of the OxPack extension (no dialog-oriented interface and no graphs) and can therefore be used with a unregistered version of Ox. This version thus simply requires any Ox executable and a text editor.

This appendix is structured as follows: in Section A.2, we propose an overview

¹For a comprehensive review of this language, see Cribari-Neto and Zarkos (2001).

of the package's features, with the presentation of the different specifications of the conditional mean and conditional variance. Comments on estimation procedures (parameters constraints, distributions, tests, forecasts, numerical accuracy of the package and a comparison with other softwares) are introduced in Section A.3. Then a user guide is provided for both versions of G@RCH 2.0 in Section A.4 with an application using the CAC40 stock index. Finally, Section A.5 concludes.

A.2 Features of the package

This section proceeds to describe the models implemented in G@RCH 2.0 and gives some technical details. Our attention will be first devoted to review the specifications of the conditional mean equation. Then, some of the most recent contributions in the ARCH modelling framework will be presented.

A.2.1 Mean equation

Let us consider an univariate time series y_t . If Ω_{t-1} is the information set at time $t - 1$, we can define its functional form as:

$$y_t = E(y_t|\Omega_{t-1}) + \varepsilon_t, \tag{A-1}$$

where $E(\cdot|\cdot)$ denotes the conditional expectation operator and ε_t is the disturbance term (or unpredictable part), with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t\varepsilon_s) = 0, \forall t \neq s$.

This is the mean equation which has been studied and modelled in many ways. Two of the most famous specifications are the Autoregressive (AR) and Moving Average (MA) models. Mixing these two processes and introducing n_1 deterministic or strongly exogenous variables in the equation, we obtain this ARMAX(n, s) process,

$$\begin{aligned} \Psi(L)(y_t - \mu_t) &= \Theta(L)\varepsilon_t \\ \mu_t &= \mu + \sum_{i=1}^{n_1} \delta_i x_{i,t}, \end{aligned} \tag{A-2}$$

where L is the lag operator, $\Psi(L) = 1 - \sum_{i=1}^n \psi_i L^i$ and $\Theta(L) = 1 + \sum_{j=1}^s \theta_j L^j$. To start the recursion, it is convenient to set the initial conditions as $\varepsilon_t = 0$ for all $t \leq \max\{p, q\}$.

Several studies have shown that the dependent variable (interest rate returns, exchange rate returns, etc.) may exhibit significant autocorrelation between observations widely separated in time. In such a case, we can say that y_t displays long memory, or long-term dependence and is best modelled by a fractionally integrated ARMA process (so called ARFIMA process) initially developed in Granger (1980) and Granger and Joyeux (1980) among others.² The ARFIMA(n, d_a, s) is given by:

$$\Psi(L)(1-L)^{d_a}(y_t - \mu_t) = \Theta(L)\varepsilon_t, \quad (\text{A-3})$$

where the operator $(1-L)^{d_a}$ accounts for the long memory of the process and is defined as:

$$\begin{aligned} (1-L)^{d_a} &= \sum_{k=0}^{\infty} \frac{\Gamma(d_a+1)}{\Gamma(k+1)\Gamma(d_a-k+1)} L^k \\ &= 1 - d_a L - \frac{1}{2}d_a(1-d_a)L^2 - \frac{1}{6}d_a(1-d_a)(2-d_a)L^3 - \dots \\ &= 1 - \sum_{k=1}^{\infty} c_k(d_a)L^k, \end{aligned} \quad (\text{A-4})$$

with $0 < d_a < 1$, $c_1(d_a) = d_a$, $c_2(d_a) = \frac{1}{2}d_a(1-d_a)$, ... and $\Gamma(\cdot)$ denoting the Gamma function (see Baillie, 1996, for a survey on this topic). The truncation order of the infinite summation is set to $t-1$.

It is worth noting that Doornik and Ooms (1999) recently provided an Ox package for estimating, forecasting and simulating ARFIMA models. However, in opposition to our package, they assume that the conditional variance is constant over time.

A.2.2 Variance equation

The ε_t term in Eq. (A-1)-(A-3) is the innovation of the process. About twenty years ago, Engle (1982) defined as an Autoregressive Conditional Heteroscedastic (ARCH) process, all ε_t of the form:

$$\varepsilon_t = z_t \sigma_t, \quad (\text{A-5})$$

²ARFIMA models have been combined with an ARCH-type specification by Baillie, Chung, and Tieslau (1996), Tschernig (1995), Teyssi re (1997), Lecourt (2000) and Beine, Laurent, and Lecourt (2000).

where z_t is an independently and identically distributed (*i.i.d.*) process with $E(z_t) = 0$ and $Var(z_t) = 1$. By definition, ε_t is serially uncorrelated with a mean equal to zero, but its conditional variance equals σ_t^2 and, therefore, may change over time, contrary to what is assumed in the standard regression model.

The models provided by our program are all ARCH-type.³ They differ on the functional form of σ_t^2 but the basic principles are the same. Besides the traditional ARCH and GARCH models, we focus mainly on two kinds of models: the asymmetric models and the fractionally integrated models. The former are defined to take account of the so-called “leverage effect” observed in many stock returns, while the latter allows for long-memory in the variance. Early evidence of the “leverage effect” can be found in Black (1976), while persistence in volatility is a common finding of many empirical studies; see Bera and Higgins (1993), Bollerslev, Chou, and Kroner (1992) or Palm (1996) for an excellent survey on ARCH models.

ARCH model

The ARCH (q) model can be expressed as:

$$\begin{aligned} \varepsilon_t &= z_t \sigma_t \\ z_t &\sim i.i.d. D(0, 1) \\ \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2, \end{aligned} \tag{A-6}$$

where $D(\cdot)$ is a probability density function with mean 0 and unit variance (it will be defined below).

The ARCH model can thus describe volatility clustering. Indeed, the conditional variance of ε_t is an increasing function of the square of the shock that occurred in $t - 1$. Consequently, if ε_{t-1} was large in absolute value, σ_t^2 and thus ε_t is expected to be large (in absolute value) as well. Notice that even if the conditional variance of an ARCH model is time-varying ($\sigma_t^2 = E(\varepsilon_t^2 | \Omega_{t-1})$), the unconditional variance of ε_t is constant and, provided that $\omega > 0$ and $\sum_{i=1}^q \alpha_i < 1$,

³For stochastic volatility models, see Koopman, Shepard, and Doornik (1998).

we have:

$$\sigma^2 \equiv E(\varepsilon_t^2) = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i}. \quad (\text{A-7})$$

Note also that the ARCH model can explain part of the excess kurtosis that we observe in financial time series. As shown by Engle (1982) for the ARCH(1) case under the normality assumption, the kurtosis of ε_t is indeed equal to $\frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}$. The kurtosis is thus finite if $\alpha_1 < \frac{1}{3}$ and larger than 3 (the kurtosis of a standard normal distribution) if $\alpha_1 > 0$.

The computation of σ_t^2 in Eq. (A-6) depends on past (squared) residuals (ε_t^2), that are not observed for $t = 0, -1, \dots, -q + 1$. To initialize the process, the unobserved squared residuals have been set to their sample mean.

In the rest of the appendix, ω is assumed fixed. If n_2 explanatory variables are introduced in the model, $\omega_t = \omega + \sum_{i=1}^{n_2} \omega_i x_{i,t}$ with an exception for the exponential models (EGARCH and FIEGARCH) where $\omega_t = \omega + \ln \left(1 + \sum_{i=1}^{n_2} \omega_i x_{i,t} \right)$.

Finally, σ_t^2 has obviously to be positive for all t . Sufficient conditions to ensure that the conditional variance in Eq. (A-6) is positive are given by $\omega > 0$ and $\alpha_i \geq 0$. However, these conditions are not necessary as shown by Nelson and Cao (1992). Furthermore, when explanatory variables enter the ARCH equation, these positivity constraints are not valid anymore (even if the conditional variance still has to be non-negative).

GARCH model

Early empirical evidence has shown that a high ARCH order has to be selected to catch the dynamics of the conditional variance (thus involving the estimation of numerous parameters). The Generalized ARCH (GARCH) model of Bollerslev (1986) is an answer to this issue. It is based on an infinite ARCH specification and it allows to reduce the number of estimated parameters by imposing non-linear restrictions on them. The GARCH (p, q) model can be expressed as:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2. \quad (\text{A-8})$$

Using the lag or backshift operator L , the GARCH (p, q) model is:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2,$$

with $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$ and $\beta(L) = \beta_1 L + \beta_2 L^2 + \dots + \beta_p L^p$.

If all the roots of the polynomial $|1 - \beta(L)| = 0$ lie outside the unit circle, we have:

$$\sigma_t^2 = \omega [1 - \beta(L)]^{-1} + \alpha(L) [1 - \beta(L)]^{-1} \varepsilon_t^2, \quad (\text{A-10})$$

which may be seen as an ARCH(∞) process since the conditional variance linearly depends on all previous squared residuals. In this case, the conditional variance of y_t can become larger than the unconditional variance given by:

$$\sigma^2 \equiv E(\varepsilon_t^2) = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i - \sum_{j=1}^p \beta_j},$$

if past realizations of ε_t^2 are larger than σ^2 (Palm, 1996).

As in the ARCH case, some restrictions are needed to ensure $\sigma_t^2 > 0$ to be positive for all t . Bollerslev (1986) shows that imposing $\omega > 0$, $\alpha_i \geq 0$ (for $i = 1, \dots, q$) and $\beta_j \geq 0$ (for $j = 1, \dots, p$) is sufficient for the conditional variance to be positive. In practice, the GARCH parameters are often estimated without the positivity restrictions. Nelson and Cao (1992) argued that imposing all coefficients to be nonnegative is too restrictive and that some of these coefficients are found to be negative in practice while the conditional variance remains positive (by checking on a case-by-case basis). Consequently, they relaxed this constraint and gave sufficient conditions for the GARCH(1, q) and GARCH(2, q) cases based on the infinite representation given in Eq. (A-10). Indeed, the conditional variance is strictly positive provided $\omega [1 - \beta(1)]^{-1} > 0$ is positive and all the coefficients of the infinite polynomial $\alpha(L) [1 - \beta(L)]^{-1}$ in Eq. (A-10) are nonnegative. The positivity constraints proposed by Bollerslev (1986) can be imposed during the estimation (see A.3.1). If not, these constraints, as well as the ones implied by the ARCH(∞) representation, will be tested a posteriori and reported in the output.

EGARCH model

The Exponential GARCH (EGARCH) model is introduced by Nelson (1991). Bollerslev and Mikkelsen (1996) propose to re-express the EGARCH model as

follows:

$$\ln \sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 + \alpha(L)]g(z_{t-1}). \quad (\text{A-11})$$

The value of $g(z_t)$ depends on several elements. Nelson (1991) notes that, “to accommodate the asymmetric relation between stock returns and volatility changes (...) the value of $g(z_t)$ must be a function of both the magnitude and the sign of z_t ”.⁴ That is why he suggests to express the function $g(\cdot)$ as

$$g(z_t) \equiv \underbrace{\gamma_1 z_t}_{\text{sign effect}} + \underbrace{\gamma_2 [|z_t| - E|z_t|]}_{\text{magnitude effect}}. \quad (\text{A-12})$$

$E|z_t|$ depends on the assumption made on the unconditional density of z_t . For the normal distribution, $E(|z_t|) = \sqrt{\frac{2}{\pi}}$. For the skewed Student distribution, $E(|z_t|) = \frac{4\xi^2}{\xi+1} \frac{\Gamma(\frac{1+\nu}{2})\sqrt{\nu-2}}{\sqrt{\pi}(\nu-1)\Gamma(\frac{\nu}{2})}$, where $\xi = 1$ for the symmetric Student. For the GED, we have $E(|z_t|) = \lambda 2^{\frac{1}{\nu}} \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})}$. ξ, ν and λ concern the shape of the non-normal densities and will be defined in Section A.3.2.

Note that the use of a \ln transformation of the conditional variance ensures that σ_t^2 is always positive.

GJR model

This popular model is proposed by Glosten, Jagannathan, and Runkle (1993). Its generalized version is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^q (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (\text{A-13})$$

where S_t^- is a dummy variable that takes the value “0” (respectively “1”) when ε_t is positive (negative).

In this model, it is assumed that the impact of ε_t^2 on the conditional variance σ_t^2 is different when ε_t is positive or negative. The TGARCH model of Zakoian (1994) is very similar to the GJR but models the conditional standard deviation instead of the conditional variance. Finally, Ling and McAleer (2002) has proposed, among other stationarity conditions for GARCH models, the conditions of existence of the second and fourth moment of the GJR.

⁴Note that with the EGARCH parameterization of Bollerslev and Mikkelsen (1996), it is possible to estimate an EGARCH $(p, 0)$ since $\ln \sigma_t^2$ depends on $g(z_{t-1})$, even when $q = 0$.

APARCH model

We have shown in Chapter 2 that the additional features introduced by the APARCH model seem justified at least for modelling the NASDAQ (on a daily basis). This model has been introduced by Ding, Granger, and Engle (1993). The APARCH (p, q) model can be expressed as:

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta, \quad (\text{A-14})$$

where $\delta > 0$ and $-1 < \gamma_i < 1$ ($i = 1, \dots, q$).

This model couples the flexibility of a varying exponent with the asymmetry coefficient (to take the “leverage effect” into account). The APARCH includes seven other ARCH extensions as special cases:⁵

- The ARCH of Engle (1982) when $\delta = 2$, $\gamma_i = 0$ ($i = 1, \dots, p$) and $\beta_j = 0$ ($j = 1, \dots, p$).
- The GARCH of Bollerslev (1986) when $\delta = 2$ and $\gamma_i = 0$ ($i = 1, \dots, p$).
- Taylor (1986)/Schwert (1990)’s GARCH when $\delta = 1$, and $\gamma_i = 0$ ($i = 1, \dots, p$).
- The GJR of Glosten, Jagannathan, and Runkle (1993) when $\delta = 2$.
- The TARARCH of Zakoian (1994) when $\delta = 1$.
- The NARCH of Higgins and Bera (1992) when $\gamma_i = 0$ ($i = 1, \dots, p$) and $\beta_j = 0$ ($j = 1, \dots, p$).
- The Log-ARCH of Geweke (1986) and Pentula (1986), when $\delta \rightarrow 0$.

The properties of the APARCH model have been studied recently by He and Teräsvirta (1999a, 1999b). Following Ding, Granger, and Engle (1993), provided that $\omega > 0$ and $\sum_{i=1}^q \alpha_i E(|z| - \gamma_i z)^\delta + \sum_{j=1}^p \beta_j < 1$, a stationary solution for Eq. (A-14) exists and is:

$$E(\sigma_t^\delta) = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i E(|z| - \gamma_i z)^\delta - \sum_{j=1}^p \beta_j}.$$

⁵Complete developments leading to these conclusions are available in Ding, Granger, and Engle (1993).

Notice that if we set $\gamma = 0$, $\delta = 2$ and z_t has zero mean and unit variance, we have the usual stationarity condition of the GARCH(1,1) model ($\alpha_1 + \beta_1 < 1$). However, if $\gamma \neq 0$ and/or $\delta \neq 2$, this condition depends on the assumption made on the innovation process.

Ding, Granger, and Engle (1993) derived a closed form solution to $\kappa_i = E(|z| - \gamma_i z)^\delta$ in the gaussian case. We have shown in Chapter 2 that for the standardized skewed Student:⁶

$$\kappa_i = \left\{ \xi^{-(1+\delta)} (1 + \gamma_i)^\delta + \xi^{1+\delta} (1 - \gamma_i)^\delta \right\} \frac{\Gamma(\frac{\delta+1}{2})\Gamma(\frac{v-\delta}{2})(v-2)^{\frac{1+\delta}{2}}}{(\xi+\frac{1}{\xi})\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}.$$

For the GED, we can show that:

$$\kappa_i = \frac{[(1+\gamma_i)^\delta + (1-\gamma_i)^\delta] 2^{\frac{\delta-v}{v}} \Gamma(\frac{\delta+1}{v}) \lambda_v^\delta}{\Gamma(\frac{1}{v})}.$$

Note that ξ, v and λ_v concern the shape of the non-normal densities and will be defined in Section A.3.2.

IGARCH model

In many high-frequency time-series applications, the conditional variance estimated using a GARCH(p, q) process has the following property:

$$\sum_{j=1}^p \beta_j + \sum_{i=1}^q \alpha_i \approx 1.$$

If $\sum_{j=1}^p \beta_j + \sum_{i=1}^q \alpha_i < 1$, the process (ε_t) is second order stationary, and a shock to the conditional variance σ_t^2 has a decaying impact on σ_{t+h}^2 , when h increases, and is asymptotically negligible. Indeed, let us rewrite the ARCH(∞) representation of the GARCH(p, q), given in Eq. (A-10), as follows:

$$\sigma_t^2 = \omega^* + \lambda(L)\varepsilon_t^2, \tag{A-15}$$

where $\omega^* = \omega [1 - \beta(L)]^{-1}$, $\lambda(L) = \alpha(L) [1 - \beta(L)]^{-1} = \sum_{i=1}^{\infty} \lambda_i L^i$ and λ_i are lag coefficients depending nonlinearly on α_i and β_i . For a GARCH(1,1), $\lambda_i = \alpha_1 \beta_1^{i-1}$. Recall that this model is said to be second order stationary provided that $\alpha_1 + \beta_1 <$

⁶For the symmetric Student density, $\xi = 1$.

1 since it implies that the unconditional variance exists and equals $\frac{\omega}{1-\alpha_1-\beta_1}$. As shown by Davidson (2001), the amplitude of the GARCH(1,1) is measured by $S = \sum_{i=1}^{\infty} \lambda_i = \alpha_1/(1-\beta_1)$, which determines “how large the variations in the conditional variance can be” (and hence the order of the existing moments). This concept is often confused with the memory of the model that determines “how large shocks to the volatility take to dissipate”. In this respect, the GARCH(1,1) model has a geometric memory $\rho = 1/\beta_1$, where $\lambda_i = O(\rho^{-i})$.

In practise, we often find $\alpha_1 + \beta_1 = 1$. In this case, we are confronted to an Integrated GARCH (IGARCH) model.

Recall that the GARCH(p, q) model can be expressed as an ARMA process. Using the lag operator L , we can rearrange Eq. (A-8) as:

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)](\varepsilon_t^2 - \sigma_t^2).$$

When the $[1 - \alpha(L) - \beta(L)]$ polynomial contains a unit root, i.e. the sum of all the α_i and the β_j is one, we have the IGARCH(p, q) model of Engle and Bollerslev (1986). It can then be written as:

$$\phi(L)(1-L)\varepsilon_t^2 = \omega + [1 - \beta(L)](\varepsilon_t^2 - \sigma_t^2), \quad (\text{A-16})$$

where $\phi(L) = [1 - \alpha(L) - \beta(L)](1-L)^{-1}$ is of order $[\max\{p, q\}-1]$.

We can rearrange Eq. (A-16) to express the conditional variance as a function of the squared residuals. After some manipulations, we have its ARCH(∞) representation:

$$\sigma_t^2 = \frac{\omega}{[1 - \beta(L)]} + \{1 - \phi(L)(1-L)[1 - \beta(L)]^{-1}\}\varepsilon_t^2. \quad (\text{A-17})$$

For this model, $S = 1$ and thus the second moment does not exist. However, this process is still short memory. To show that Davidson (2001) consider an IGARCH(0,1) model defined as $\varepsilon_t = \sigma_t z_t$ and $\sigma_t^2 = \varepsilon_{t-1}^2$. This process is often wrongly compared to a random walk since the long-range forecast $\sigma_{t+h}^2 = \varepsilon_t^2$, for any h . However, $\varepsilon_t = z_t|\varepsilon_{t-1}|$ which means that the memory of a large deviation persists for only one period.

Fractionally integrated models

Volatility tends to change quite slowly over time, and, as shown in Ding, Granger, and Engle (1993) among others, the effects of a shock can take a considerable time

to decay.⁷ Therefore, the distinction between I(0) and I(1) processes seems to be far too restrictive. Indeed, the propagation of shocks in an I(0) process occurs at an exponential rate of decay (so that it only captures the short-memory), while for an I(1) process the persistence of shocks is infinite. In the conditional mean, the ARFIMA specification has been proposed to fill the gap between short and complete persistence, so that the short-run behavior of the time-series is captured by the ARMA parameters, while the fractional differencing parameter allows for modelling the long-run dependence.⁸

To mimic the behavior of the correlogram of the observed volatility, Baillie, Bollerslev, and Mikkelsen (1996) (hereafter denoted BBM) introduce the Fractionally Integrated GARCH (FIGARCH) model by replacing the first difference operator of Eq. (A-17) by $(1 - L)^d$.

The conditional variance of the FIGARCH (p, d, q) is given by:

$$\sigma_t^2 = \underbrace{\omega[1 - \beta(L)]^{-1}}_{\omega^*} + \underbrace{\{1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\}}_{\lambda(L)}\varepsilon_t^2, \quad (\text{A-18})$$

or $\sigma_t^2 = \omega^* + \sum_{i=1}^{\infty} \lambda_i L^i \varepsilon_t^2 = \omega^* + \lambda(L)\varepsilon_t^2$, with $0 \leq d \leq 1$. It is fairly easy to show that $\omega > 0, \beta_1 - d \leq \phi_1 \leq \frac{2-d}{2}$ and $d(\phi_1 - \frac{1-d}{2}) \leq \beta_1(\phi_1 - \beta_1 + d)$ are sufficient to ensure that the conditional variance of the FIGARCH $(1, d, 1)$ is positive almost surely for all t . Setting $\phi_1 = 0$ gives the condition for the FIGARCH $(1, d, 0)$. Once again, these conditions are verified after the estimation and printed in the output.

Davidson (2001) notes the interesting and counterintuitive fact that the memory parameter of this process is $-d$, and is increasing as d approaches zero, while in the ARFIMA model the memory increases when d_a increases. According to Davidson (2001), the unexpected behavior of the FIGARCH model may be due less to any inherent paradoxes than to the fact that, embodying restrictions appropriate to a model in levels, it has been transplanted into a model of volatility. The main

⁷In their study of the daily S&P500 index, they find that the squared returns series has positive autocorrelations over more than 2,500 lags (or more than 10 years!).

⁸See Bollerslev and Mikkelsen (1996, p.158) for a discussion on the importance of non-integer values of integration when modelling long-run dependencies in the conditional mean of economic time series.

characteristic of this model is that it is not stationary when $d > 0$. Indeed,

$$\begin{aligned}
(1-L)^d &= \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} L^k \\
&= 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \dots \\
&= 1 - \sum_{k=1}^{\infty} c_k(d)L^k, \tag{A-19}
\end{aligned}$$

where $c_1(d) = d, c_2(d) = \frac{1}{2}d(1-d)$, etc. By construction, $\sum_{k=1}^{\infty} c_k(d) = 1$ for any value of d , and consequently, the FIGARCH belongs to the same “knife-edge-nonstationary” class represented by the IGARCH ($S = 1$).⁹ To test whether this nonstationarity feature holds, Davidson (2001) proposes a generalized version of the FIGARCH and calls it the HYperbolic GARCH. The HYGARCH is given by Eq. (A-18), when $\lambda(L)$ is replaced by $1 - [1 - \beta(L)]^{-1}\phi(L) \{1 + \alpha [(1-L)^d - 1]\}$. Note that we report $\ln(\alpha)$ and not α . The $c_k(d)$ coefficients are thus weighted by α . Interestingly, the HYGARCH nests the FIGARCH when $\alpha = 1$ (or equivalently when $\ln(\alpha) = 0$) and if the GARCH component observes the usual covariance stationarity restrictions, then this process is stationary with $\alpha < 1$ (or equivalently when $\ln(\alpha) < 0$) (see Davidson, 2001 for more details).

Chung (1999) underscores some little drawbacks in the BBM model: there is a structural problem in the BBM specification since the parallel with the ARFIMA framework of the conditional mean equation is not perfect, leading to difficult interpretations of the estimated parameters. Indeed the fractional differencing operator applies to the constant term in the mean equation (ARFIMA) while it does not in the variance equation (FIGARCH). Chung (1999) proposes a slightly different process:

$$\phi(L)(1-L)^d (\varepsilon_t^2 - \sigma^2) = [1 - \beta(L)](\varepsilon_t^2 - \sigma_t^2), \tag{A-20}$$

where σ^2 is the unconditional variance of ε_t .

If we keep the same definition of $\lambda(L)$ as in Eq. (A-18), we can formulate the conditional variance as:

$$\sigma_t^2 = \sigma^2 + \{1 - [1 - \beta(L)]^{-1}\phi(L)(1-L)^d\} (\varepsilon_t^2 - \sigma^2)$$

⁹Note that the hyperbolic memory of the FIGARCH is measured by the parameter d , such that $\lambda_i = O(i^{-1-d})$. The memory is thus increasing as d approaches 0 unlike for the ARFIMA model. See Davidson (2001) on this point.

or

$$\sigma_t^2 = \sigma^2 + \lambda(L) (\varepsilon_t^2 - \sigma^2). \quad (\text{A-21})$$

$\lambda(L)$ is an infinite summation which, in practice, has to be truncated. BBM propose to truncate $\lambda(L)$ at 1000 lags (this truncation order has been implemented as the default value in our package, but it may be changed by the user) and initialize the unobserved ε_t^2 at their unconditional moment. Contrary to BBM, Chung (1999) proposes to truncate $\lambda(L)$ at the size of the information set ($t - 1$) and to initialize the unobserved $(\varepsilon_t^2 - \sigma^2)$ at 0 (this quantity is small in absolute values and has a zero mean), see Chung (1999) for more details.

The idea of fractional integration has been extended to other GARCH types of models, including the Fractionally Integrated EGARCH (FIEGARCH) of Bollerslev and Mikkelsen (1996) and the Fractionally Integrated APARCH (FIAPARCH) of Tse (1998).¹⁰

Similarly to the GARCH(p, q) process, the EGARCH(p, q) of Eq. (A-11) can be extended to account for long memory by factorizing the autoregressive polynomial $[1 - \beta(L)] = \phi(L)(1 - L)^d$ where all the roots of $\phi(z) = 0$ lie outside the unit circle. The FIEGARCH (p, d, q) is specified as follows:

$$\ln(\sigma_t^2) = \omega + \phi(L)^{-1} (1 - L)^{-d} [1 + \alpha(L)]g(z_{t-1}). \quad (\text{A-22})$$

Finally, the FIAPARCH (p, d, q) model can be written as:¹¹

$$\sigma_t^\delta = \omega + \left\{ 1 - [1 - \beta(L)]^{-1} \phi(L) (1 - L)^d \right\} (|\varepsilon_t| - \gamma\varepsilon_t)^\delta. \quad (\text{A-23})$$

A.3 Estimation methods

A.3.1 Parameters constraints

When numerical optimization is used to maximize the log-likelihood function with respect to the vector of parameters Ψ , the inspected range of the parameter space is $]-\infty; \infty[$. The problem is that some parameters might have to be constrained in

¹⁰Notice that the GJR has not been extended to the long-memory framework. It is however nested in the FIAPARCH class of models.

¹¹When using the BBM option in G@RCH for the FIEGARCH and FIAPARCH, $(1 - L)^d$ and $(1 - L)^{-d}$ are truncated at some predefined value (see above). It is also possible to truncate this polynomial at the information size at time t , i.e. $t - 1$.

a smaller interval. For instance, the leverage effect parameter γ of the APARCH model must lie between -1 and 1. To impose these constraints one could estimate Ψ^* (which ranges from $-\infty$ to $+\infty$) instead of Ψ where Ψ is recovered using the non-linear function: $\Psi = x(\Psi^*)$. In our package, $x(\cdot)$ is defined as:

$$x(\Psi^*) = Low + \frac{Up - Low}{1 + e^{-\Psi^*}}, \quad (\text{A-24})$$

where Low is the lower bound and Up the upper bound (i.e. in our example, $Low = -1$ and $Up = 1$).

So, applying unconstrained optimization of the log-likelihood function with respect to Ψ is equivalent to applying constrained optimization with respect to Ψ^* . Therefore, the optimization process of the program results in $\hat{\Psi}^*$ with the covariance matrix being noted $Cov(\hat{\Psi}^*)$. The estimated covariance of the parameters of interest $\hat{\Psi}$ is:

$$Cov(\hat{\Psi}) = \left(\frac{\partial x(\hat{\Psi}^*)}{\partial \Psi^*} \right) Cov(\hat{\Psi}^*) \left(\frac{\partial x(\hat{\Psi}^*)}{\partial \Psi^*} \right)'. \quad (\text{A-25})$$

In our case, we have $Cov(\hat{\Psi}) = Cov(\hat{\Psi}^*) \frac{\exp(-\hat{\Psi}^*)(Up-Low)}{[1+\exp(-\hat{\Psi}^*)]^2}$. Note that, in G@RCH 2.0, lower and upper bounds of the parameters can be easily modified by the user in the file *startingvalues.txt*.

A.3.2 Distributions

Four distributions are available in our program: the usual Gaussian, the Student- t , the Generalized Error Distribution (GED) and the skewed Student distribution.

The GARCH models are estimated using an approximate Maximum Likelihood (ML) approach. It is quite evident from Eq. (A-6) (and all the following equations of Section A.2) that the recursive evaluation of this function is conditional on unobserved values. The ML estimation is therefore not perfectly exact. To solve the problem of unobserved values, we have set these quantities to their unconditional expected values or sample mean.

If we express the mean equation as in Eq. (A-1) and $\varepsilon_t = z_t \sigma_t$, the Gaussian, Student and skewed Student log-likelihood functions are given respectively in Eq. (2.5), (2.13) and (2.31).

The GED log-likelihood function of a normalized random variable is given by:

$$L_{GED} = \sum_{t=1}^T \left[\ln(v/\lambda_v) - 0.5 \left| \frac{z_t}{\lambda_v} \right|^v - (1 + v^{-1}) \ln(2) - \ln \Gamma(1/v) - 0.5 \ln(\sigma_t^2) \right], \quad (\text{A-26})$$

where $0 < v < \infty$ and

$$\lambda_v \equiv \sqrt{\frac{\Gamma\left(\frac{1}{v}\right) 2^{-\frac{2}{v}}}{\Gamma\left(\frac{3}{v}\right)}}.$$

In principal, the gradient vector and the hessian matrix can be obtained numerically or by evaluating its analytic expressions. We have shown in Chapter 3 that using analytical scores can highly speed-up ML estimation and improve the numerical accuracy. However, due to the high number of possible models and distributions, we use numerical techniques to approximate the derivatives of the log-likelihood function with respect to the parameter vector.

A.3.3 Tests

In addition to the possibilities offered by GiveWin (ACF, PACF, QQ-plots...), several tests are provided:

- Four Information Criteria (divided by the number of observations):¹²
 - Akaike = $-2\frac{LogL}{n} + 2\frac{k}{n}$;
 - Hannan-Quinn = $-2\frac{LogL}{n} + 2\frac{k \ln[\ln(n)]}{n}$;
 - Schwartz = $-2\frac{LogL}{n} + 2\frac{\ln(k)}{n}$;
 - Shibata = $-2\frac{LogL}{n} + \ln\left(\frac{n+2k}{n}\right)$.
- The value of the skewness and the kurtosis of the standardized residuals (\hat{z}_t) of the estimated model, their t -tests and p -values. Moreover, the Jarque-Bera normality test (Jarque and Bera, 1987) is also reported.
- The Box-Pierce statistics at lag l^* for both standardized, i.e. $BP(l^*)$, and squared standardized, i.e. $BP^2(l^*)$, residuals. Under the null hypothesis of no autocorrelation, the statistics $BP(l^*)$ and $BP^2(l^*)$ are respectively $\chi^2(l^* - m - l)$ and $\chi^2(l^* - p - q)$ distributed (see McLeod and Li, 1983).

¹² $LogL$ = log likelihood value, n is the number observations and k the number of estimated parameters.

- The Engle LM ARCH test (Engle, 1982) to test for the presence of ARCH effects in a series.
- The diagnostic test of Engle and Ng (1993) to investigate possible misspecification of the conditional variance equation. The Sign Bias Test (SBT) examines the impact of positive and negative return shocks on volatility not predicted by the model under construction. The negative Size Bias Test (resp. positive Size Bias Test) focuses on the different effects that large and small negative (resp. positive) return shocks have on volatility, which is not predicted by the volatility model. Finally, a joint test for these three tests is also provided.
- The adjusted Pearson goodness-of-fit test. See Chapter 2 for more details.
- The Nyblom test (Nyblom, 1989 and Lee and Hansen, 1994) to check the constancy of parameters over time. See Hansen (1994) for an overview of this test.

A.3.4 Forecasts

Estimating a model can be useful to try to understand the mechanism that produces the series of interest. It can also suggest a solution to an economic problem. Is it the only game in town? Certainly not. Indeed, the main purpose of building and estimating a model with financial data is to produce a forecast. GARCH 2.0 also provides forecasting tools. Actually, forecasts of both the conditional mean and the conditional variance are available as well as several forecast error measures.

Forecasting the conditional mean

Our first goal is to give the optimal h -step-ahead predictor of y_{t+h} given the information we have up to time t .

For instance, for the following AR(1) process,

$$y_t = \mu + \psi_1(y_{t-1} - \mu) + \varepsilon_t.$$

The optimal¹³ h -step-ahead predictor of y_{t+h} , i.e. $\hat{y}_{t+h|t}$, is its conditional ex-

¹³By optimal, we mean optimal under expected quadratic loss, or in a mean square error sense.

pectation at time t (given the estimated parameters $\hat{\mu}$ and $\hat{\psi}_1$):

$$\hat{y}_{t+h|t} = \hat{\mu} + \hat{\psi}_1(\hat{y}_{t+h-1|t} - \hat{\mu}), \quad (\text{A-29})$$

where $\hat{y}_{t+i|t} = y_{t+i}$ for $i \leq 0$.

For the AR(1), the optimal 1-step-ahead forecast equals $\hat{\mu} + \hat{\psi}_1(\hat{y}_t - \hat{\mu})$. For $h > 1$, the optimal forecast can be obtained recursively or directly as $\hat{y}_{t+h|t} = \hat{\mu} + \hat{\psi}_1^h(\hat{y}_t - \hat{\mu})$.

In the general case of an ARFIMA(n, d_a, s) as given in Eq. (A-3), the optimal h -step-ahead predictor of y_{t+h} is:

$$\begin{aligned} \hat{y}_{t+h|t} &= \left[\hat{\mu}_{t+h|t} + \sum_{k=1}^{\infty} \hat{c}_k(\hat{y}_{t+h-k} - \hat{\mu}_{t+h|t}) \right] \\ &+ \sum_{i=1}^n \hat{\psi}_i \left\{ \hat{y}_{t+h-i} - \left[\hat{\mu}_{t+h|t} + \sum_{k=1}^{\infty} \hat{c}_k(\hat{y}_{t+h-i-k} - \hat{\mu}_{t+h|t}) \right] \right\} \\ &+ \sum_{j=1}^s \hat{\theta}_j(\hat{y}_{t+h-j} - \hat{y}_{t+h-j|t}). \end{aligned} \quad (\text{A-30})$$

Recall that when exogenous variables enter the conditional mean equation, μ becomes $\mu_t = \mu + \sum_{i=1}^{n_1} \delta_i x_{i,t}$ and consequently, provided that the information $x_{i,t+h}$ is available at time t (which is the case for instance if $x_{i,t}$ is a “day-of-the-week” dummy variable), $\hat{\mu}_{t+h|t}$ is also available at time t . When there is no exogenous variable in the ARFIMA model and $n = 1, s = 0$ and $d_a = 0$ ($c_k = 0$), the forecast of the AR(1) process given in Eq. (A-29) can be recovered.

Forecasting the conditional variance

Independently from the conditional mean, one can forecast the conditional variance. In the simple GARCH(p, q) case, the optimal h -step-ahead forecast of the conditional variance, i.e. $\hat{\sigma}_{t+h|t}^2$ is given by:

$$\sigma_{t+h|t}^2 = \hat{\omega} + \sum_{i=1}^q \hat{\alpha}_i \varepsilon_{t+h-i|t}^2 + \sum_{j=1}^p \hat{\beta}_j \sigma_{t+h-j|t}^2,$$

where $\varepsilon_{t+i|t}^2 = \sigma_{t+i|t}^2$ for $i > 0$ while $\varepsilon_{t+i|t}^2 = \varepsilon_{t+i}^2$ and $\sigma_{t+i|t}^2 = \sigma_{t+i}^2$ for $i \leq 0$. Eq. (A-31) is usually computed recursively, even if a closed form solution of $\sigma_{t+h|t}^2$ can be obtained by recursive substitution in Eq. (A-31).

Similarly, one can easily obtain the h -step-ahead forecast of the conditional variance of an ARCH, IGARCH and FIGARCH model. By contrast, for thresholds models, the computation of the out-of-sample forecasts is more complicated. Indeed, for the GJR and APARCH models (as well as for their long-memory counterparts), the assumption made on the innovation process may have an effect on the forecast (especially for $h > 1$).

For instance, for the GJR (p, q) model,

$$\hat{\sigma}_{t+h|t}^2 = \hat{\omega} + \sum_{i=1}^q (\hat{\alpha}_i \varepsilon_{t-i+h|t}^2 + \hat{\gamma}_i S_{t-i+h|t}^- \varepsilon_{t-i+h|t}^2) + \sum_{j=1}^p \hat{\beta}_j \sigma_{t-j+h|t}^2. \quad (\text{A-32})$$

When all the γ_i parameters equal 0, one recovers the forecast of the GARCH model. Otherwise, one has to compute $S_{t-i+h|t}^-$. Note first that $S_{t+i|t}^- = S_{t+i}^-$ for $i \leq 0$. However, when $i > 1$, $S_{t+i|t}^-$ depends on the choice of the distribution of z_t . When the distribution of z_t is symmetric around 0 (for the Gaussian, Student and GED density), the probability that ε_{t+i} will be negative is $S_{t+i|t}^- = 0.5$. If z_t is (standardized) skewed Student distributed with asymmetry parameter ξ and degree of freedom ν , $S_{t+i|t}^- = \frac{1}{1+\xi^2}$ since ξ^2 is the ratio of probability masses above and below the mode.

For the APARCH (p, q) model,

$$\begin{aligned} \hat{\sigma}_{t+h|t}^\delta &= E(\sigma_{t+h}^\delta | \Omega_t) \\ &= E\left(\hat{\omega} + \sum_{i=1}^q \hat{\alpha}_i (|\varepsilon_{t+h-i}| - \hat{\gamma}_i \varepsilon_{t+h-i})^\delta + \sum_{j=1}^p \hat{\beta}_j \sigma_{t+h-j}^\delta \mid \Omega_t\right) \\ &= \hat{\omega} + \sum_{i=1}^q \hat{\alpha}_i E\left[(\varepsilon_{t+h-i} - \hat{\gamma}_i \varepsilon_{t+h-i})^\delta \mid \Omega_t\right] + \sum_{j=1}^p \hat{\beta}_j \sigma_{t+h-j|t}^\delta, \end{aligned} \quad (\text{A-33})$$

where $E\left[(\varepsilon_{t+k} - \hat{\gamma}_i \varepsilon_{t+k})^\delta \mid \Omega_t\right] = \kappa_i \sigma_{t+k|t}^\delta$, for $k > 1$ and $\kappa_i = E(|z| - \gamma_i z)^\delta$ (see Section A.3.2).

For the EGARCH (p, q) model,

$$\begin{aligned} \ln \hat{\sigma}_{t+h|t}^2 &= E(\ln \sigma_{t+h}^2 | \Omega_t) \\ &= E\left\{\hat{\omega} + [1 - \hat{\beta}(L)]^{-1} [1 + \hat{\alpha}(L)] \hat{g}(z_{t+h-1}) \mid \Omega_t\right\} \\ &= [1 - \hat{\beta}(L)] \hat{\omega} + \hat{\beta}(L) \ln \hat{\sigma}_{t+h|t}^2 + [1 + \hat{\alpha}(L)] \hat{g}(z_{t+h-1|t}), \end{aligned} \quad (\text{A-34})$$

where $\hat{g}(z_{t+k|t}) = \hat{g}(z_{t+k})$ for $k \leq 0$ and 0 for $k > 0$.

Finally, the h -step-ahead forecast of the FIAPARCH and FIEGARCH models are obtained in a similar way.

One of the most popular measures to check the forecasting performance of the ARCH-type models is the Mincer-Zarnowitz regression, i.e. ex-post volatility regression:

$$\check{\sigma}_t^2 = a_0 + a_1 \hat{\sigma}_t^2 + u_t, \quad (\text{A-35})$$

where $\check{\sigma}_t^2$ is the ex-post volatility, $\hat{\sigma}_t^2$ is the forecasted volatility and a_0, a_1 are parameters to be estimated. If the model for the conditional variance is correctly specified (and the parameters are known) and $E(\check{\sigma}_t^2) = \hat{\sigma}_t^2$, it follows that $a_0 = 0$ and $a_1 = 1$. The R^2 of this regression is often used as a simple measure of the degree of predictability of the ARCH-type model.

However, $\check{\sigma}_t^2$ is never observed. By default, G@RCH 2.0 uses $\check{\sigma}_t^2 = (y_t - \bar{y})^2$, where \bar{y} is the sample mean of y_t . The R^2 of this regression is often lower than 5% and this could lead to the conclusion that GARCH models produce poor forecasts of the volatility (see, among others, Schwert, 1990, or Jorion, 1996). But, as described in Andersen and Bollerslev (1998), the reason of these poor results is the choice of what is considered as the “true” volatility. G@RCH 2.0 allows to select any series as the “observed” volatility (Obs.-Var., see Figure A.1). The user may then compute the daily realized volatility as the sum of squared intraday returns and use it as the “true” volatility. Actually, Andersen and Bollerslev (1998) show that this measure is a more proper one than squared daily returns. Therefore, using 5-minute returns for instance, the daily realized volatility can be expressed as:

$$\sigma_t^2 = \sum_{k=1}^K y_{k,t}^2, \quad (\text{A-36})$$

where $y_{k,t}$ is the return of the k^{th} 5-minutes interval of the t^{th} day and K is the number of 5-minutes intervals per day.

Finally, to compare the adequacy of the different distributions, G@RCH 2.0 also allows the computation of density forecasts tests developed in Diebold, Gunther, and Tay (1998), that we have briefly reviewed in Chapter 2.¹⁴ An illustration is provided in Section A.4 with some formal tests and graphical tools.

¹⁴For more details about density forecasts and applications in finance, see the special issue of *Journal of Forecasting* (Timmermann, 2000).

A.3.5 Numerical accuracy

McCullough and Vinod (1999) and Brooks, Burke, and Persaud (2001) use the daily German mark/British pound exchange rate data of Bollerslev and Ghysels (1996) to compare the numerical accuracy of GARCH model estimation among several econometric softwares. They choose the GARCH(1,1) model described in Fiorentini, Calzolari, and Panattoni (1996) (hereafter denoted FCP) as the benchmark. In this section, we use the same methodology with the same dataset to check the accuracy of our procedures. Coefficients and standard errors estimates of G@RCH 2.0 are reported in Table A.1 together with the results of McCullough and Vinod (1999) (based on the FORTRAN procedure of FCP and thus entitled “FCP” in the table).

Table A.1: Accuracy of the GARCH procedure

	Coefficient		Standard Errors		Robust Standard Errors	
	G@RCH	FCP	G@RCH	FCP	G@RCH	FCP
μ	-0.006184	-0.006190	0.008462	0.008462	0.009187	0.009189
ω	0.010760	0.010761	0.002851	0.002852	0.006484	0.006493
α_1	0.153407	0.153134	0.026569	0.026523	0.053595	0.053532
β_1	0.805879	0.805974	0.033542	0.033553	0.072386	0.072461

G@RCH 2.0 gives very satisfactory results since the first four digits (at least) are the same as those of the benchmark for all but two estimations. In addition, it competes well compared to other well known econometric softwares. Table A.2 gives indeed the coefficient estimates and the error percentage associated for 5 softwares. G@RCH, PcGive and TSP (these last two softwares use analytical second-order derivatives for the standard GARCH model) clearly outperform Eviews and S-Plus on this specification (if one believes in the benchmark values).

Moreover, to investigate the accuracy of our forecasting procedures, we have run a 8-step ahead forecasts of the model, similar to Brooks, Burke, and Persaud (2001). Table 4 in Brooks, Burke, and Persaud (2001) reports the conditional variance forecasts given by six well-known softwares and the benchmark values. Contrary to E-Views, Matlab and SAS, G@RCH 2.0 hits the benchmarks for all steps to the third decimal (note that GAUSS, Microfit and Rats also do).

Table A.2: GARCH accuracy comparison

	FCP	G@RCH	Eviews	PcGive	TSP	S-Plus
μ	-0.00619	-0.00618	-0.00541	-0.00625	-0.00619	-0.00919
ω	0.010761	0.010760	0.009581	0.010760	0.010761	0.011696
α_1	0.153134	0.153407	0.142284	0.153397	0.153134	0.154295
β_1	0.805974	0.805879	0.821336	0.805886	0.805974	0.800276
μ	-	0.10%	12.58%	0.91%	0.00%	48.41%
ω	-	0.01%	10.96%	0.01%	0.00%	8.69%
α_1	-	0.18%	7.08%	0.17%	0.00%	0.76%
β_1	-	0.01%	1.91%	0.01%	0.00%	0.71%

Finally, Lombardi and Gallo (2001) extend the work of Fiorentini, Calzolari, and Panattoni (1996) to the FIGARCH model of Baillie, Bollerslev, and Mikkelsen (1996) and derive analytic expressions for the second-order derivatives of this model in the Gaussian case. For the same DEM/UKP database as in the previous example, Table A.3 reports the coefficients estimates and their standard errors for our package (using numerical gradients and the BFGS optimization method) and for Lombardi and Gallo (2001) (using analytical gradients and the Newton-Raphson algorithm; results correspond to the columns entitled “LG”).

Table A.3: Accuracy of the FIGARCH procedure

	Coefficient		Standard Errors	
	G@RCH	LG	G@RCH	LG
μ	0.003606	0.003621	0.009985	0.009985
ω	0.015772	0.015764	0.003578	0.003581
α_1	0.198134	0.198448	0.042508	0.042444
β_1	0.675652	0.675251	0.051800	0.051693
d	0.570702	0.569951	0.075039	0.074762

Results show that G@RCH 2.0 provides accurate estimates, even for an advanced model such as the FIGARCH. As expected, it is however more time-consuming than the C code of Lombardi and Gallo (2001)¹⁵ (163 sec. vs 43 sec. using a PIII processor with 450 Mhz).

¹⁵This C code is available at <http://www.ds.unifi.it/~mj1/> in the “software” section. Note that the only configuration available is a FIGARCH (1, d , 1) with a constant in the mean and variance equations and a Gaussian likelihood.

A.3.6 Features comparison

The goal of this section is to compare in the most objective way, the features offered by G@RCH 2.0 with respect to nine other well known econometric softwares, namely PcGive 10 (also programmed in Ox), GAUSS and its Fanpac package, Eviews 4, S-Plus 6 and its GARCH module, Rats 5.0 and its *garch.src* procedure¹⁶, TSP 4.5, Microfit 4, SAS 8.2 and Stata 7. It is not our intention to evaluate a program against another, but we will rather show an overview of what can or cannot be done with these softwares.

The proposed models and options differ widely from one program to another as can be seen in Table A.4. Regarding the range of different univariate models, if many programs propose asymmetric models, very few (G@RCH, S-Plus with the FIGARCH and the FIEGARCH and Fanpac with the FIGARCH) offer long memory models in the variance equation and none (except G@RCH) offers a fractionally integrated specification in the mean. As for the distribution, the choice is often limited to symmetric densities (except G@RCH that provides a skewed Student likelihood). Finally, robust standard errors are proposed in 5 programs out of the 10 we have compared (G@RCH, PcGive, GAUSS Fanpac, Eviews and Stata).

A.4 Application

A.4.1 Data and methodology

To illustrate our G@RCH 2.0 package with a concrete application, we analyze the French CAC40 stock index for the years 1995-1999 (1249 daily observations). It is computed by the exchange as a weighted measure of the prices of its components and is available in the database on an intraday basis with the price index being computed every 15 minutes. For the time period under review, the opening hours of the French stock market were 10.00 am to 5.00 pm, thus 7 hours of trading per day. This translates into 28 intraday returns used to compute the daily realized volatility. Intraday prices are the outcomes of a linear interpolation between the closest recorded prices below and above the time set in the grid. Correspondingly,

¹⁶This file is available at <http://www.estima.com/procindx.htm> for download.

Table A.4: GARCH features comparison

	G@RCH	PcGive	Fanpac	Eviews	S-Plus	Rats	TSP	Microft	SAS	Stata
Version	2.0	10	-	4.0	6	5.0	4.5	4	8.2	7
Conditional mean										
Explanatory variables	+	+	+	+	+	+	+	+	+	+
ARMA	+	+	+	+	+	+	+	+	+	+
ARFIMA	+	-	-	-	-	-	-	-	-	-
ARCH-in-Mean	-	+	+	+	+	+	+	+	+	+
Conditional variance										
Explanatory variables	+	+	+	+	+	+	+	+	+	+
GARCH	+	+	+	+	+	+	+	+	+	+
IGARCH	+	-	+	-	-	+	-	-	+	-
EGARCH	+	+	+	+	+	+	-	+	+	+
GJR	+	+	-	+	+	+	-	-	-	+
APARCH	+	-	-	-	+	-	-	-	-	+
C-GARCH	-	-	-	+	+	-	-	-	-	-
FIGARCH	+	-	+	-	+	-	-	-	-	-
FIEGARCH	+	-	-	-	+	-	-	-	-	-
FIAPARCH	+	-	-	-	-	-	-	-	-	-
HYGARCH	+	-	-	-	-	-	-	-	-	-
Distributions										
Normal	+	+	+	+	+	+	+	+	+	+
Student-t	+	+	+	-	+	+	-	+	+	-
GED	+	+	-	-	+	+	-	-	-	-
Skewed-t	+	-	-	-	-	-	-	-	-	-
Double Exponential	-	-	-	-	+	-	-	-	-	-
Estimation										
MLE	+	+	+	+	+	+	+	+	+	+
QMLE	+	+	+	+	+	-	-	-	-	+

A “+” (resp. “-”) means that the corresponding option is (resp. is not) available for this software. C-GARCH corresponds to the Component GARCH of Engle and Lee (1999).

all returns are computed as the first difference in the regularly time-spaced log prices of the index. Because the exchange is closed from 5h pm to 10h am the next day, the first intraday return is the first difference between the log price at 10h15 and the log price at 5h pm the day before. On the first hand, the intraday data are used to compute the daily realized volatility using Eq. (A-36). On the other hand, daily returns in percentage are defined as 100 times the first difference of the log of the closing prices.¹⁷

The estimation of the parameters is carried out for the 800 observations while forecasting is computed for the last observations.

A.4.2 Using the “Full Version”

Once the installation process is correctly completed following the instructions of the *readme.txt* file, the user may open the database he wants to use in GiveWin (in the example “CAC15.xls”), and then select the OXPack module.

Once our package has been selected, one can launch the **Model/Formulate** menu. The list of all the variables of the database appears in the *Database* section (see Figure A.1). There are four possible statuses for each variable: dependent variable (Y variable), regressor in the mean (Mean), regressor in the variance (Variance) or observed volatility (Obs. Var.). Our program provides estimations for univariate models¹⁸, so only one Y variable per model is accepted. However one can include several regressors in the mean and the variance equations and the same variable can be a regressor in both equations.

Once the OK button is pressed, the **Model/Model Settings** box automatically appears. This box allows to select the specification of the model: AR(FI)MA orders for the mean equation, GARCH orders, type of GARCH model for the variance equation and the distribution (Figure A.2). The default specification is an ARMA(0,0)-GARCH(1,1) with normal errors. In our application, we select an ARMA(1,0)-APARCH(1,1) specification with a skewed Student likelihood.

As explained in Section A.3.1, it is possible to constrain the parameters to range between a lower and an upper bound by selecting the **Bounded Parameters**

¹⁷By definition and using the properties of the log distribution, the sum of the intraday returns is equal to the observed daily return based on the closing prices.

¹⁸The extension of this package to multivariate GARCH models is currently under development.

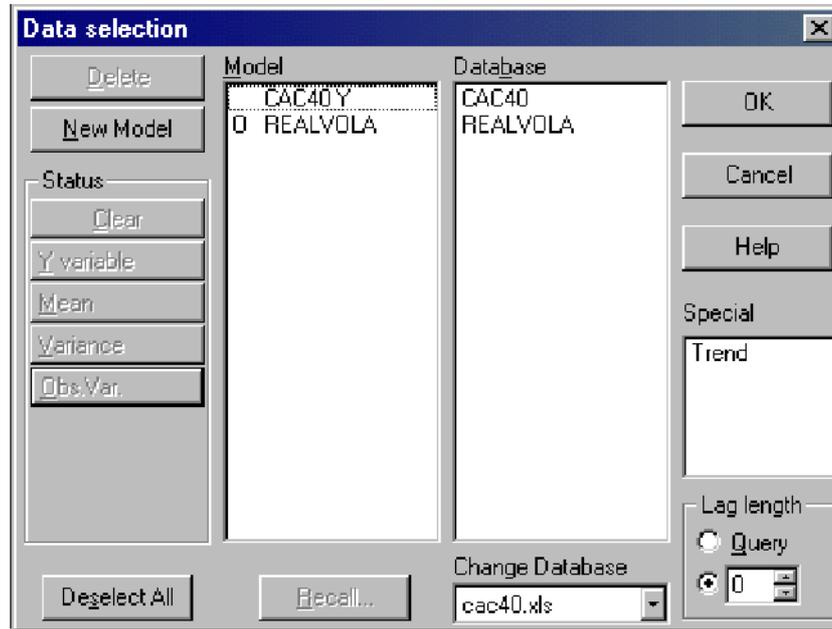


Figure A.1: Selecting the variables

option. The defaults bounds can be changed in the *startingvalues.txt* file.

In the next window, the user is asked to make a choice regarding the starting values (Figure A.3): he might (1) let the program use the predefined starting values¹⁹, (2) enter them manually, element by element, or (3) enter the starting values in a vector form (the required form is “value1;value2;value3”).

Then, the estimation method for standard deviations is selected: ML or QML (with a specified pseudo-likelihood) or both. In this window (see Figure A.4), one may also select the sample and some maximization options (such as the number of iterations between intermediary results printings) when clicking on the *Options* button.

The estimation procedure is then launched and the program comes back to GiveWin. Let us assume that the element-by-element method has been selected. A new window appears (see Figure A.5) with all the possible parameters to be estimated. Depending on the specification, some parameters have a value, other do not. The user should replace only the former, since they correspond to the parameters to be estimated for the specified model.

¹⁹Note that these default values can be modified by the user. Indeed they are stored in the *startingvalues.txt* file installed with the package.

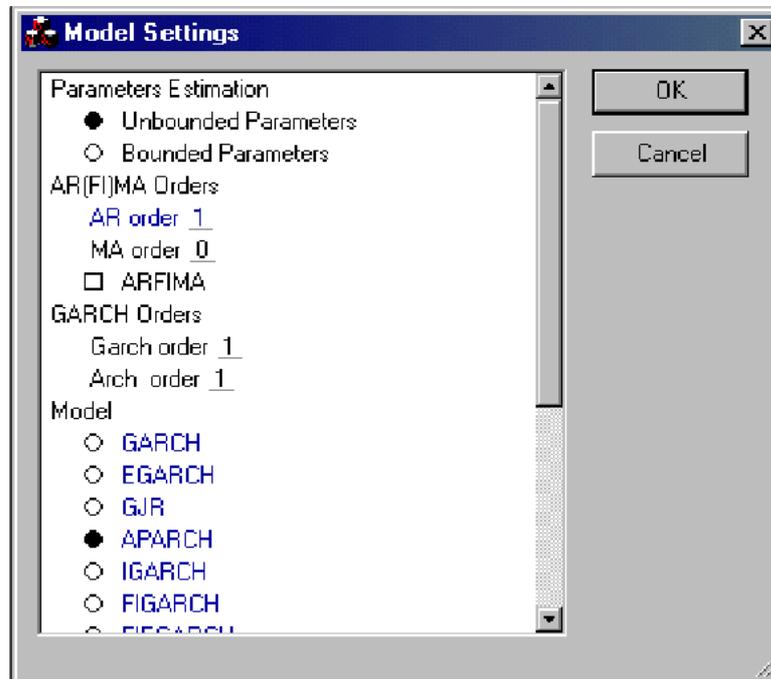


Figure A.2: Model settings

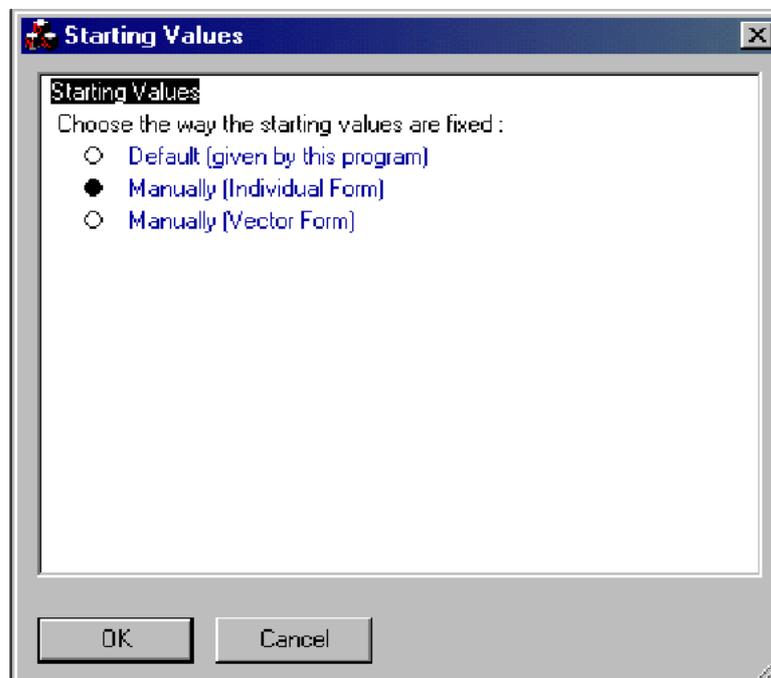


Figure A.3: Selecting the starting values method

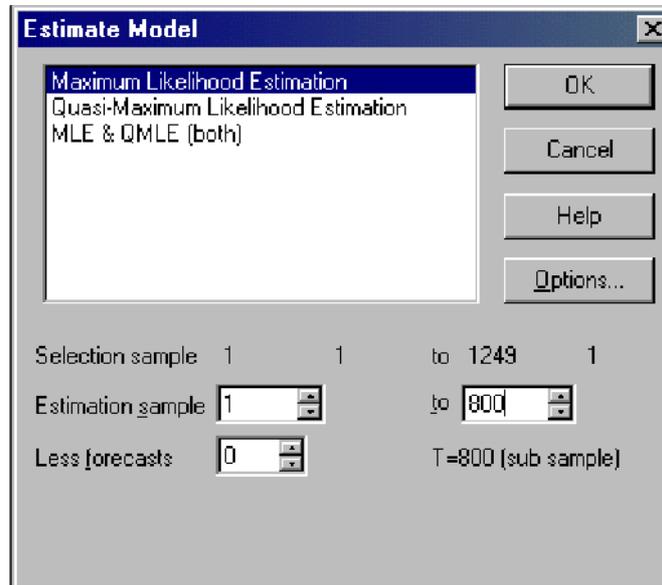


Figure A.4: Standard errors estimation methods

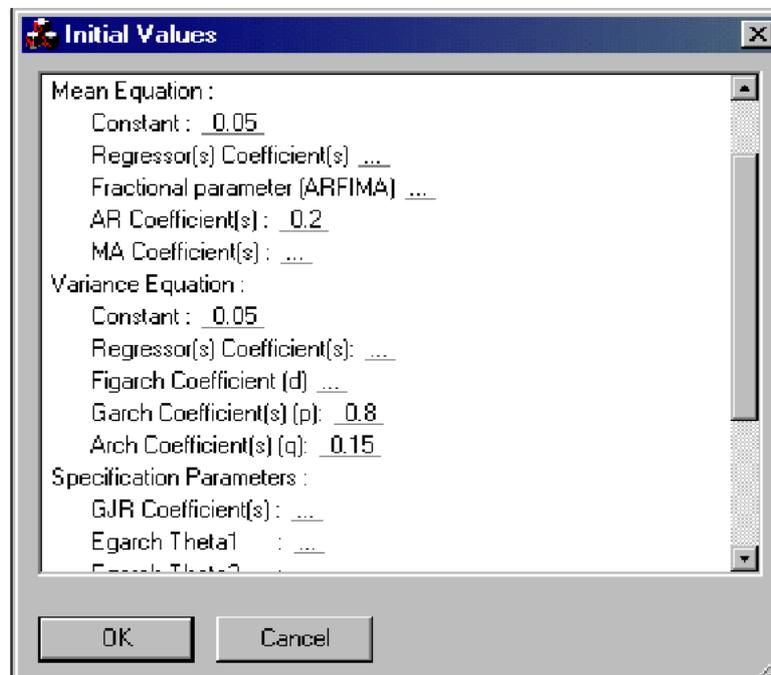


Figure A.5: Entering the starting values

Once this step is completed, the program starts the iteration process. The final output is divided by default in two main parts: first, the model specification reminder; second, the estimated values and other useful statistics of the parameters.²⁰ The output is given in the box “Output 1”.

After the estimation of the model, new options are available in OxPack: `Menu/Tests`, `Menu/Graphic Analysis`, `Menu/Forecasts`, `Menu/Exclusion Restrictions`, `Menu/Linear Restrictions` and `Menu/Store`.

The `Menu/Graphic Analysis` option allows to plot different graphics (see Figure A.6 for details). Just as any other graphs in the GiveWin environment, they can be easily edited (color, size, . . .) and exported in many formats (.eps, .ps, .wmf, .emf and .gwg). Figure A.7 provides the graphs of the squared residuals and the conditional mean with a 95% confidence interval.

The `Menu/Tests` option allows to run different tests (see Section A.3.2 for further explanations). It also allows to print the variance-covariance matrix of the estimated parameters (Figure A.8). The results of these tests are printed in GiveWin. An example of output is reported in the next box (“Output 2”).

²⁰Recall that the estimations are based on the numerical evaluation of the gradients.

```

*****
** SPECIFICATIONS **
*****
Mean Equation: ARMA (1, 0) model.
No regressor in the mean.
Variance Equation : APARCH (1, 1) model.
No regressor in the variance.
The distribution is a Skewed Student distribution, with a tail coefficient of 15.72 and an asymmetry coefficient of
-0.08751.
Strong convergence using numerical derivatives

Maximum Likelihood Estimation

      Coefficient      Std.Error      t-value      t-prob
Cst(M)      0.065337      0.037157      1.758      0.0791
AR(1)      0.004704      0.037117      0.1267      0.8992
Cst(V)      0.017498      0.013488      1.297      0.1949
Beta1      0.947590      0.020193      46.93      0.0000
Alpha1      0.038464      0.017776      2.164      0.0308
Gamma1      0.676364      0.348702      1.940      0.0528
Delta      1.462837      0.533581      2.742      0.0063
Asymmetry   -0.087512      0.054314      -1.611      0.1075
Tail        15.718323      8.087414      1.944      0.0523

No. Observations: 800      No. Parameters: 9
Mean (Y): 0.08103      Variance (Y): 1.27405
Log Likelihood: -1190.521      Alpha[1]+Beta[1]: 0.98605

The sample mean of squared residuals was used to start recursion.
The condition for existence of  $E(\sigma^\delta)$  and  $E(|e^\delta|)$  is observed.
The constraint equals 0.9926 and should be < 1.
Vector of estimated parameters:
0.065337; 0.004704; 0.017498; 0.947590; 0.038464; 0.676364; 1.462837;-0.087512; 15.718323

```

TESTS:

Information Criterium (minimize)
 Akaike 2.998802 Shibata 2.998553
 Schwarz 3.051504 Hannan-Quinn 3.019048

	Statistic	t-value	t-prob
Skewness	-0.2135	2.47	0.0135
Excess Kurtosis	0.4684	2.713	0.006674
Jarque-Bera	13.39	13.39	0.001235

BOX-PIERCE:

H0: No serial correlation \Rightarrow Accept H0 when prob. is High [$Q < \text{Chisq}(\text{lag})$]

Box-Pierce Q-statistics on residuals

\rightarrow P-values adjusted by 1 degree(s) of freedom

Q(10) = 14.47 [0.1064]

Q(20) = 21.67 [0.3012]

Box-Pierce Q-statistics on squared residuals

\rightarrow P-values adjusted by 2 degree(s) of freedom

Q(10) = 9.887 [0.2731]

Q(20) = 16.13 [0.5838]

Diagnostic test based on the news impact curve (EGARCH vs. GARCH)

	Test	Prob
Sign Bias t-Test	0.98838	0.32297
Negative Size Bias t-Test	0.14581	0.88407
Positive Size Bias t-Test	0.62400	0.53263
Joint Test for the Three Effects	5.13914	0.16189

Joint Statistic of the Nyblom test of stability: 2.727

Individual Nyblom Statistics:

Cst(M)	0.72438
AR(1)	0.68524
Cst(V)	0.51505
Beta1	0.42785
Alpha1	0.46229
Gamma1	0.43489
Delta	0.54130
Asymmetry	0.21342
Tail	0.08950

Rem: Asymptotic 1% critical value for individual statistics = 0.75.

Asymptotic 5% critical value for individual statistics = 0.47.

Adjusted Pearson Chi-square Goodness-of-fit test

Lags	Statistic	P-Value(lag-1)	P-Value(lag-k-1)
40	24.9000	0.961261	0.729877
50	26.7500	0.995994	0.946240
60	32.6500	0.997893	0.972622

Rem.: k = # estimated parameters

We do not intend to comment this application in details. However, looking at

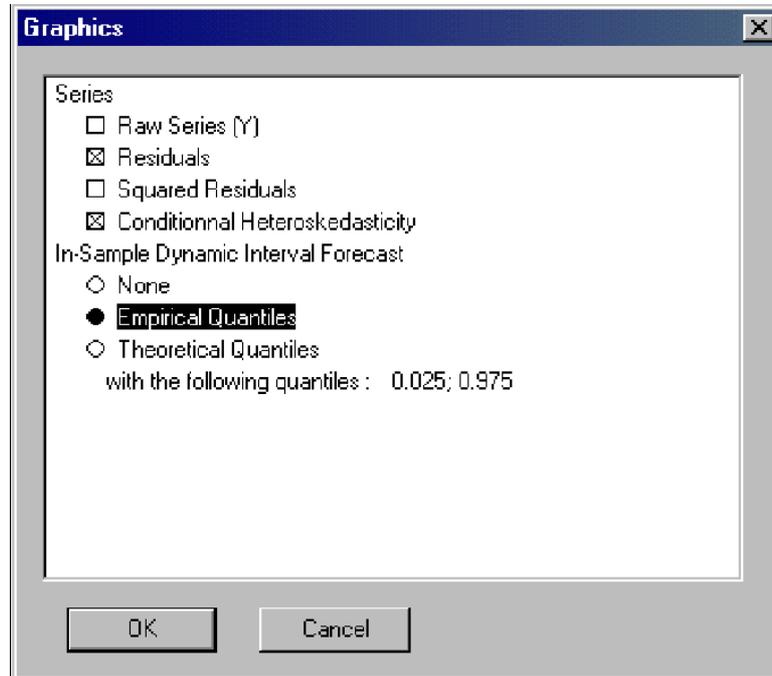


Figure A.6: Graphics menu

these results, one can briefly argue that the model seems to capture the dynamics of the first and second moments of the CAC40 (see the Box-Pierce statistics). Indeed, the Sign Bias tests show that there is no remaining leverage component in the innovations while the Nyblom stability test suggests that the estimated parameters are quite stable during the investigated period. Finally, our model specification is not rejected by the goodness-of-fit tests for various lag lengths.

To obtain the h -step-ahead forecasts, access the menu **Test/Forecast** and set the number of forecasts, pre-sample observations (to be plotted) as well as some other graphical options.

Figure A.9 shows 10 pre-sample observations and the forecasts up to horizon 10 of the conditional mean. The forecasted bands are $\pm 2\hat{\sigma}_{t+h|t}$ (note that the critical value 2 can be modified by the user).

A.4.3 Using the “Light Version”

First, to specify the model you want to estimate, you have to edit *GarchEstim.ox* with any text editor. Yet we recommend OxEdit. It is a shareware that highlights

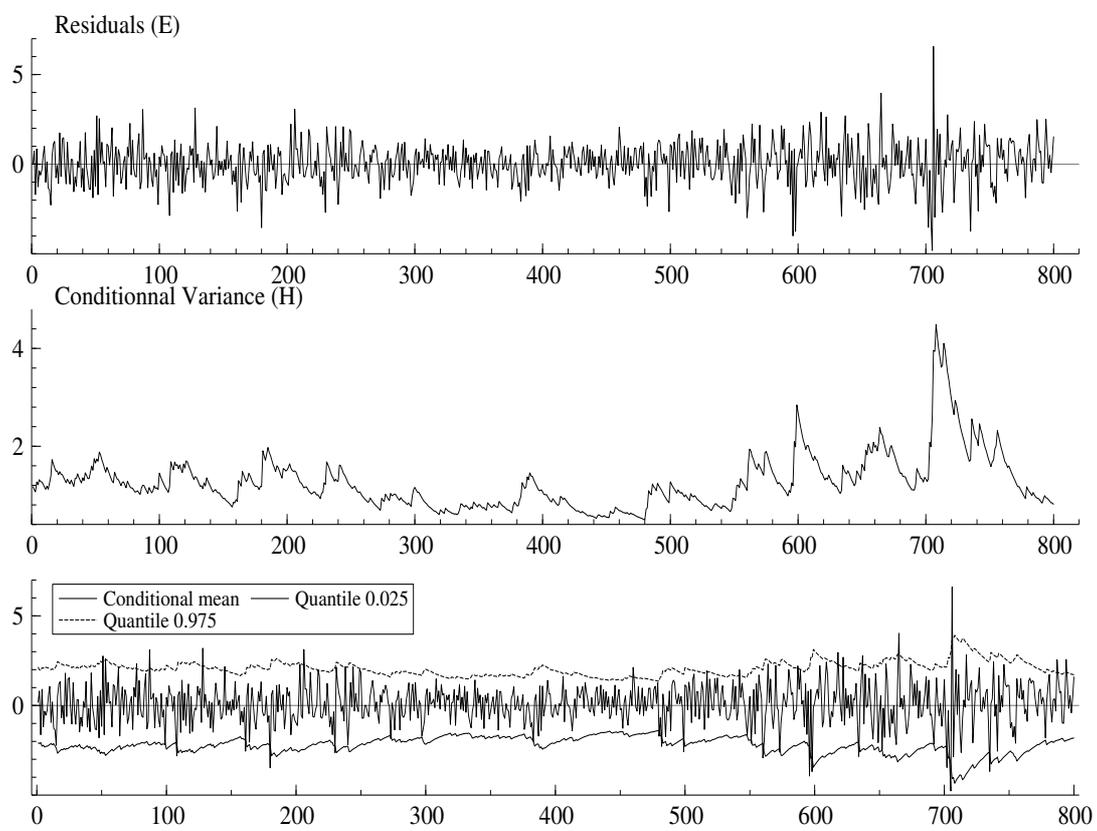


Figure A.7: Graphical analysis

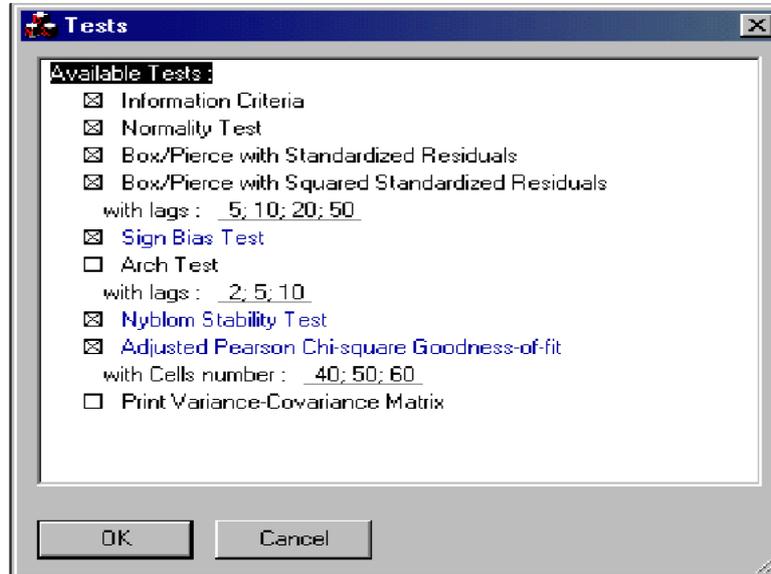


Figure A.8: Tests dialog box

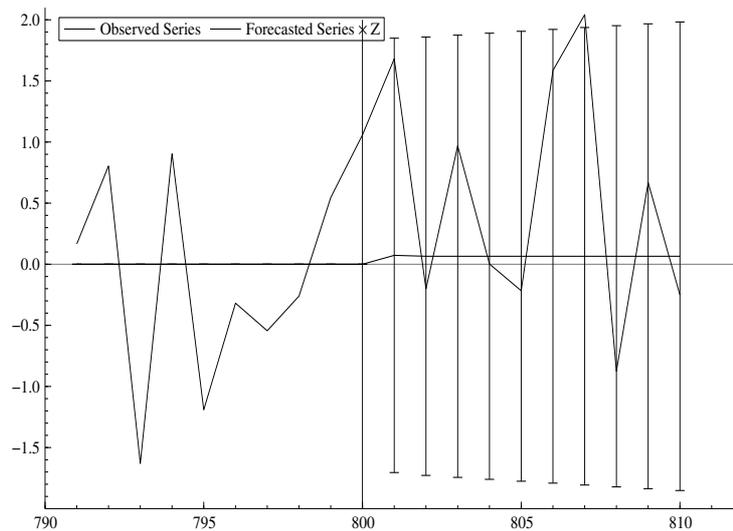


Figure A.9: Forecasts from an AR(1)-APARCH(1,1).

Ox syntax in color (see <http://www.oxedit.com> for more details). An example of the *GarchEstim.ox* file is displayed in the next box.

The *GarchEstim* file consists of five parts:

- the “Data” part deals with the database, the sample and the variables selection;
- the “Specification” part is related to the choice of the model, the lag orders and the shape of the distribution;
- the “Tests & Forecasts” part allows to compute different tests and to parameterize the forecasting part. Note that `BOXPIERCE`, `ARCHLAGS` and `PEARSON` all require a vector of integers corresponding to the lags used in the computation of the statistics;
- the “Output” part includes several options including `MLE` that refer to the computation method of the standard deviations of the estimated parameters, `TESTONLY`, useful when you want to run some tests on the raw series, prior to any estimation and `GRAPHS` and `FOREGRAPHS`, to print graphs for the estimation and the forecasting, respectively;²¹

²¹Graphics will only be displayed when using GiveWin as front-end.

```

                                                                 GarchEstim.ox
#import <packages/garch/garch>
main()
{
    decl garchobj;
    garchobj = new Garch();
    /*** DATA ***/
    garchobj.Load("/data/cac40.xls");
    garchobj.Info();
    garchobj.Select(Y_VAR, {"CAC40",0,0});
    // garchobj.Select(X_VAR, {"NAME",0,0});           // REGRESSOR IN THE MEAN
    // garchobj.Select(Z_VAR, {"NAME",0,0});           // REGRESSOR IN THE VARIANCE
    // garchobj.Select(O_VAR, {"REALVOLA",0,0});        // REALIZED VOLATILITY
    garchobj.SetSelSample(-1, 1, 1000, 1);
    /*** SPECIFICATIONS ***/
    garchobj.CSTS(1,1);           // cst in Mean (1 or 0), cst in Variance (1 or 0)
    garchobj.DISTRI(1);           // 0 for Gauss, 1 for Student, 2 for GED, 3 for Skewed-Student
    garchobj.ARMA_ORDERS(1,0);    // AR order (p), MA order (q).
    garchobj.ARFIMA(0);           // 1 if Arfima wanted, 0 otherwise
    garchobj.GARCH_ORDERS(1,1);   // p order, q order
    garchobj.MODEL(1);           // 1:GARCH 2:EGARCH 3:GJR 4:APARCH 5:IGARCH
    // 6:FIGARCH(BBM) 7:FIGARCH(Chung) 8:FIEGARCH(BBM only)
    // 9:FIAPARCH(BBM) 10: FIAPARCH(Chung) 11: HYGARCH(BBM)

    garchobj.TRUNC(1000);        // Truncation order (only F.I. models with BBM method)
    /*** TESTS & FORECASTS ***/
    garchobj.BOXPIERCE(<10;15;20>); // Lags for the Box-Pierce Q-statistics, <> otherwise
    garchobj.ARCHLAGS(<2;5;10>);    // Lags for Engle's LM ARCH test, <> otherwise
    garchobj.NYBLOM(1);           // 1 to compute the Nyblom stability test, 0 otherwise
    garchobj.PEARSON(<40;50;60>);  // Cells for the adjusted Pearson Chi-square Goodness-of-fit test
    garchobj.FORECAST(0,9,1);     // Arg.1 : 1 to launch the forecasting procedure, 0 otherwise
    // Arg.2 : Number of forecasts
    // Arg.3 : 1 to Print the forecasts, 0 otherwise

    /*** OUTPUT ***/
    garchobj.MLE(1);             // 0 : both, 1 : MLE, 2 : QMLE
    garchobj.COVAR(0);          // if 1, prints variance-covariance matrix of the parameters.
    garchobj.ITER(0);           // Interval of iterations between printed intermediary results
    garchobj.TESTSONLY(0,0);    // Arg.1 : if 1, runs tests for the raw Y series, prior to ...
    // Arg.2 : if 1, runs tests after the estimation.
    garchobj.GRAPHS(0,0,"");    // Arg.1 : if 1, displays graphics of the estimations.
    // Arg.2 : if 1, saves these graphics in a EPS file
    // Arg.3 : Name of the saved file.
    garchobj.FOREGRAPHS(1,0,""); // Same as GRAPHS(p,s,n) but for the graphics of the forecasts.
    /*** PARAMETERS ***/
    garchobj.BOUNDS(1);         // 1 if bounded parameters wanted, 0 otherwise
    garchobj.FixParam(1);       // 1 to fix some parameters to their starting values, 0 otherwise
    garchobj.FixedParam(<0;0;0;0;0;0;1>); // 1 to fix and 0 to estimate the corresponding parameter
    garchobj.DoEstimation(<>);

    // m_vPar = m_clevel | m_vbetam | m_dARFI | m_vAR | m_vMA | m_calpha0 | m_vgammmav | m_dD | m_vbetav |
    // m_valphav | m_vleverage | m_vtheta1 | m_vtheta2 | m_vpsy | m_ddelta | m_cA | m_cV | m_vHY
    garchobj.STORE(0,0,0,0,0,"01",0); // Arg.1,2,3,4,5 : if 1 -> stored. (Res-SqRes-CondV-...
    // Arg.6 : Suffix. The name of the saved series will be...
    // Arg.7 : if 0, saves as an Excel spreadsheet (.xls)...

    delete garchobj;
}

```

- the “Parameters” part consists in four procedures. `BOUNDS` to constraint or not several parameters to range between a lower and an upper bound (see Section A.3.1), `FixParam` to fix some parameters to their starting values, `DoEstimation` that launches the estimation of the model and the `STORE` function allowing to store some series. The argument of the `DoEstimation`

procedure is a vector containing starting values of the parameters in a specified order (but the user can also let the program take default values by entering “<>” as function argument).

Note that the “Light Version” is more than just a replication of the “Full Version” without the graphical interface. Indeed, G@RCH uses the object-oriented programming features of Ox and provides a new class called `Garch`. All the functions of this class can thus be used within an Ox programme. To illustrate the potentiality of our package, we also provide *Forecast.ox*, an example that computes 448 one-step-ahead forecasts of the conditional mean and conditional variance (using the estimated parameters presented in the previous section), computes the Mincer-Zarnowitz regression and performs some out-of-sample density forecast tests as suggested by Diebold, Gunther, and Tay (1998) and summarized in Chapter 2.

The interesting part of *Forecast.ox* is printed in the next box. This code has been used to produce Figure A.10 and the outputs associated with this forecasting experiment (see below).

In the first four panels of Figure A.10, we show the correlograms of $(\hat{\zeta} - \bar{\zeta})^j$, for $j = 1, 2, 3, 4$, where $\hat{\zeta}$ is the probability integral transform (see Section 2.4). This graphical tool has been proposed by Diebold, Gunther, and Tay (1998) to detect potential remaining dependence in the conditional mean, variance, skewness, kurtosis (see Chapter 2 for more details). In our example, it seems that the probability integral transform is independently distributed.

Panel 5 of Figure A.10 also show the histogram (with 30 cells) of $\hat{\zeta}$ with the 95 % confidence bands. From this figure, it is clear that the AR(1)-APARCH(1,1) model coupled with a skewed Student distribution for the innovations performs very well with the dataset we have investigated. This conclusion is reinforced by the Pearson Chi-square goodness-of-fit test printed hereafter that provides a statistical version of the graphical test presented in Figure A.10. Finally, the program performs the Mincer-Zarnowitz regression given in Eq. (A-35) that regresses the observed volatility (in our case the realized volatility) on a constant and a vector of 448 one-step-ahead forecasts of the conditional variance (produced by the APARCH model).²² The results (reported in the box “Output 3”) suggest

²²The realized and one-step-ahead forecasts are plotted in the last panel of Figure A.10.

that the APARCH model gives good forecasts of the conditional variance. Indeed, looking at the estimated parameters of this regression, one can hardly conclude that the APARCH model provides biased forecasts. Moreover, the R^2 of this regression is higher than 40 % (which is similar to findings of Andersen and Bollerslev, 1998 or Blair, Poon, and Taylor, 2000).

```

                                                                    Forecast.ox
#import <packages/garch/garch>
main()
{
  decl garchobj;
  garchobj = new Garch();

  ...

  garchobj.DoEstimation(<>);
  decl number_of_forecasts=448; // number of h_step_ahead forecasts
  decl step=1; // specify h (h-step-ahead forecasts)
  decl T=garchobj.GetcT();
  decl par=garchobj.PAR() [][0];
  println("!!! Please Wait while computing the forecasts !!!");
  decl forc=<>, h, yfor=<>, Hfor=<>;
  decl RV=columns(garchobj.GetGroup(0_VAR));
  decl shape=<>;
  if (garchobj.GetDistri()==1 || garchobj.GetDistri()==2) // Except for the HYGARCH
    shape=par[rows(par)-1];
  else if (garchobj.GetDistri()==3)
    shape=par[rows(par)-2:rows(par)-1];
  for (h=0; h<number_of_forecasts; ++h)
  {
    garchobj.FORECAST(1,step,0);
    garchobj.SetSelSample(-1, 1, T+h, 1);
    garchobj.InitData();
    yfor|=garchobj.GetForcData(Y_VAR, step);
    forc|=garchobj.FORECASTING();
    if (RV==1)
      Hfor|=garchobj.GetForcData(0_VAR, step); // If you use the realized volatility
  }
  decl cd=garchobj.CD(yfor-forc[][0],forc[][1],garchobj.GetDistri(),shape);
  println("Density Forecast Test on Standardized Forecast Errors");
  garchobj.APGT(cd,20|30,rows(par));
  garchobj.AUTO(cd, number_of_forecasts, -0.1, 0.1, 0);
  garchobj.confidence_limits_uniform(cd,30,0.95,1,4);
  if (RV==0)
  {
    DrawTitle(5, "Conditional variance forecast and absolute returns");
    Hfor = (yfor - meanc(yfor)).^2;
  }
  else
    DrawTitle(5, "Conditional variance forecast and realized volatility");
  Draw(5, (Hfor~forc[][1]));
  ShowDrawWindow();
  garchobj.MZ(Hfor, forc, number_of_forecasts);
  garchobj.FEM(forc, yfor~Hfor);

  garchobj.STORE(0,0,0,0,0,"01",0); // Arg.1,2,3,4,5 ...
                                   // Arg.6 : Suffix. ...
                                   // Arg.7 : if 0, ...

  delete garchobj;
}

```

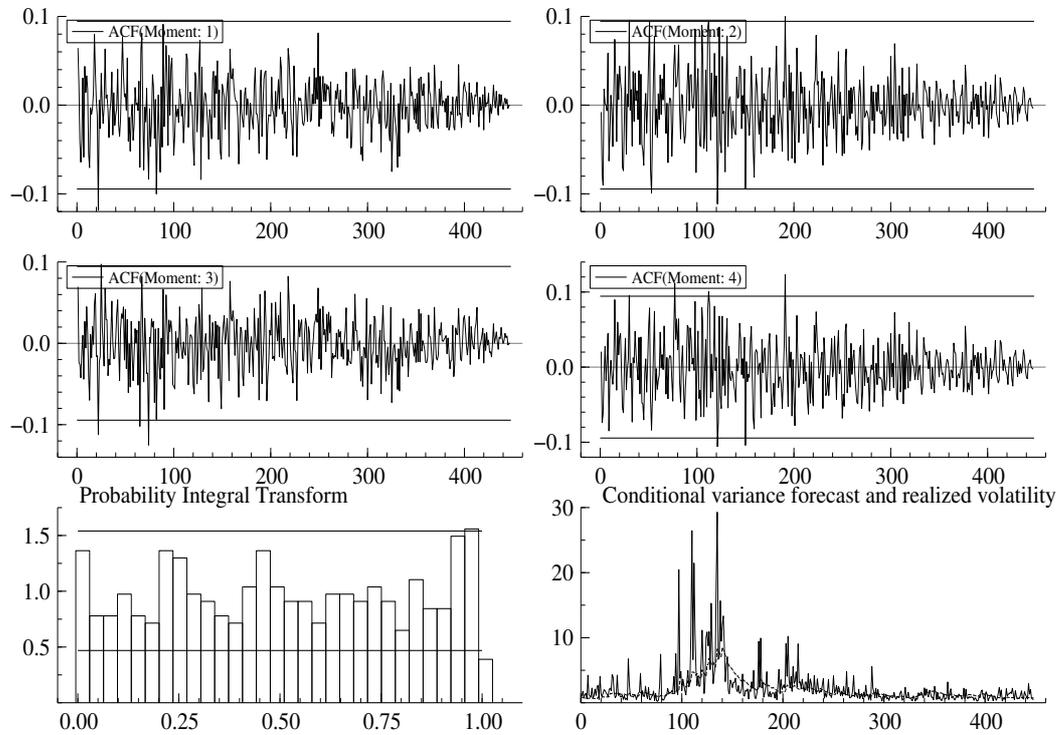


Figure A.10: Density forecast analysis

Output 3

Density Forecast Test on Standardized Forecast Errors Adjusted Pearson Chi-square Goodness-of-fit test

Lags	Statistic	P-Value(lag-1)	P-Value(lag-k-1)
20	21.0179	0.335815	0.020969
30	26.5089	0.598181	0.149654

Rem.: k = number of estimated parameters

Mincer-Zarnowitz regression on the forecasted volatility

	Coefficient	Std.Error	t-value	t-prob
a_0	-0.225818	0.264837	-0.8527	0.3940
a_1	1.370648	0.176086	7.784	0.0000

R^2 : 0.402914

Note: S.E. are Heteroskedastic Consistent (White, 80)

A.5 Conclusions

This appendix documents the software G@RCH 2.0, an Ox package allowing to estimate and to forecast numerous univariate ARCH-type processes including GARCH, EGARCH, GJR, APARCH, IGARCH, FIGARCH, HYGARCH, FIEGARCH and FIAPARCH specifications of the conditional variance. Several features of the program are worth noting since they are unavailable in most of the traditional econometric softwares: the asymmetric and fractionally integrated processes, four distributions (normal, Student- t , GED and skewed Student- t), (editable) parameters bounds, several misspecification tests and h-step-ahead forecasts.

G@RCH 2.0 is free of charge when used for educational or research purposes and can be downloaded at <http://www.egss.ulg.ac.be/garch/>.

Appendix B

Appendix Chapter 2

Additive Outliers (AO) detection and correction

In this Appendix we briefly review and discuss the extension of Franses and Ghijsels (1999) method to deal with additive outliers in an GARCH model.

Consider a univariate time series y_t^* following an ARMA (n, s) process:

$$\Psi(L)(y_t^* - \mu) = \Theta(L)\varepsilon_t^* \quad (\text{B-1})$$

$$\varepsilon_t^* = \sigma_t^* z_t \quad (\text{B-2})$$

where $\Psi(L) = 1 - \psi_1 L - \dots - \psi_n L^n$ and $\Theta(L) = 1 + \theta_1 L + \dots + \theta_s L^s$ are the usual AR and MA lag polynomials of respective orders n and s (with all roots lying outside the unit circle), σ_t^* is a time-varying conditional standard deviation and z_t is an independent and identically distributed (*i.i.d.*) random variable with zero mean.

Let us also assume that σ_t^* is described by an APARCH (p, q) :

$$\sigma_t^{\delta*} = \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}^*| - \gamma_i \varepsilon_{t-i}^*)^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^{\delta*}. \quad (\text{B-3})$$

Eq. (B-3) can be rewritten as follows:

$$\sigma_t^{\delta*} = \omega + \alpha_+(L)(\varepsilon_t^{*+})^\delta + \alpha_-(L)(-\varepsilon_t^{*-})^\delta + \beta(L)\sigma_t^{\delta*}, \quad (\text{B-4})$$

where $\varepsilon_t^{*+} = \max(\varepsilon_t^*, 0)$, $\varepsilon_t^{*-} = \min(\varepsilon_t^*, 0)$, $\alpha_+(L) = \sum_{i=1}^q \alpha_i (1 - \gamma_i) L^i$,

$\alpha_-(L) = \sum_{i=1}^q \alpha_i (1 + \gamma_i) L^i$ and $\beta(L) = \sum_{i=1}^p \beta_i L^i$.

After some manipulations, (B-4) can be re-expressed as an ARMA ($\max_{p,q}, p$) process (where $\max_{p,q}$ is the maximum value with respect to p and q):

$$|\varepsilon_t^*|^\delta = \omega + [\alpha_+(L) + \beta(L)] (\varepsilon_t^{*+})^\delta + [\alpha_-(L) + \beta(L)] (-\varepsilon_t^{*-})^\delta + v_t^* - \beta(L)v_t^*, \quad (\text{B-5})$$

where $v_t^* = |\varepsilon_t^*|^\delta - \sigma_t^{\delta*}$ plays the role of the innovation term of the APARCH model (similar to ε_t^* in Eq.(B-1)).

An additive outlier model in the conditional variance can be represented by:

$$v_t = v_t^* + \varphi I_t(\tau^*), \quad (\text{B-6})$$

where $I_t(\tau^*) = 1$ when $t = \tau^*$, 0 otherwise, while v_t^* is given by Eq. (B-5). Note that we have non a priori knowledge of τ^* .

This process with v_t instead of v_t^* leads to y_t instead of y_t^* . If an ARMA (n, s)-APARCH (p, q) model is fitted to y_t , one obtains \hat{v}_t as:

$$\hat{v}_t = \lambda_+(L)(\varepsilon_t^{*+})^\delta + \lambda_-(L)(-\varepsilon_t^{*-})^\delta - \omega [1 - \beta(L)]^{-1}, \quad (\text{B-7})$$

where $\lambda_+(L) = [1 - \beta(L)]^{-1} [1 - \alpha_+(L) - \beta(L)] = 1 - \lambda_1^+ L - \lambda_2^+ L^2 - \dots$ and $\lambda_-(L) = [1 - \beta(L)]^{-1} [1 - \alpha_-(L) - \beta(L)] = 1 - \lambda_1^- L - \lambda_2^- L^2 - \dots$

For the moment we assume that the values of the parameters, as well as the orders n, s, p and q , are known. For the AO in (B-6), the Eq. (B-7) amounts to:

$$\hat{v}_t = v_t^* + \varphi \lambda_+(L) S_t^+ I_t(\tau^*) + \varphi \lambda_-(L) S_t^- I_t(\tau^*), \quad (\text{B-8})$$

where $S_t^+ = 1$ when $\varepsilon_t > 0$, 0 otherwise and $S_t^- = 1 - S_t^+$.

The expression in (B-8) can be viewed as a regression model for \hat{v}_t , i.e.

$$\hat{v}_t = \varphi x_t + v_t^*, \quad (\text{B-9})$$

where $x_t = 0$ for $t < \tau^*$, $x_t = 1$ for $t = \tau^*$ and $x_{t+k} = -\lambda_k^+$ if $\varepsilon_{\tau^*} > 0$, $-\lambda_k^-$ otherwise for $t < \tau^*$ and $k = 1, 2, \dots$

The impact φ of the AO at time $t = \tau^*$ can then be estimated as:

$$\hat{\varphi}(\tau^*) = \frac{\sum_{t=\tau^*}^T \hat{v}_t x_t}{\sum_{t=\tau^*}^T x_t^2} \quad (\text{B-10})$$

Chang, Tiao, and Chen (1988) suggest to standardize $\hat{\varphi}(\tau^*)$ such that one can test for the significance of an AO. For this standardization we need an estimate

of the variance of the residual process. Preferably, this estimate should not be biased too much because of outliers. Chen and Liu (1993) suggest three methods to estimate a robust error variance. In the empirical section below we will use the so-called omit-one method for computational convenience. This method calculates the error variance from the sample where the observation at $t = \tau^*$ has been deleted. Denoting the estimated error variance, based on the omit-one method, as $\hat{\sigma}_a$, we can construct the standardized statistic:

$$\hat{AO}(\tau^*) = \frac{\hat{\varphi}(\tau^*)\hat{\sigma}_a^{-1}}{\sqrt{\sum_{t=\tau^*}^T x_t^2}} \quad (\text{B-11})$$

When \hat{AO} exceeds some value C , the impact of the AO is said to be significant. Based on extensive simulations, Chen and Liu (1993) advocate setting C equal to 4. However, the choice of C remains arbitrary. In case \hat{AO} is large and in excess of C , one can adjust the observation y_t to obtain the AO-corrected y_t^* . In case of more than one AO, one can repeat this procedure until any \hat{AO} statistic becomes insignificant. In a final step one can re-estimate the parameters for all observations, where some of these have been corrected for AOs.

This can be done in five steps:

- Step 1: Estimate the parameters of (B-1)-(B-3) using the observed data (y_t) and construct $\hat{v}_t^* = |\hat{\varepsilon}_t^*|^\delta - \hat{\sigma}_t^{\delta*}$.
- Step 2: For each $t = \tau^*$, perform the regression (B-9) and calculate the statistics $\hat{AO}(\tau^*)$.
- Step 3: The observation on \hat{v}_t at $t = \tau^*$ with the largest value of the \hat{AO} statistic (which should exceed C), is replaced by \hat{v}_t^* using Eq. (B-8) and the estimated weight described in Eq. (B-10).
- Step 4: With \hat{v}_t^* and $\hat{\sigma}_t^{\delta*}$, construct $\varepsilon_t^{\delta*}$ via $\varepsilon_t^{\delta*} = \hat{v}_t^* + \hat{\sigma}_t^{\delta*}$ at time $t = \tau^*$. Furthermore, construct the AO-corrected residuals as $\varepsilon_t^* = \varepsilon_t$ for $t \neq \tau^*$ and $\varepsilon_t^* = (\varepsilon_t^{\delta*})^{1/\delta} \text{sign}(\varepsilon_t)$ for $t = \tau^*$ (to preserve the sign of the returns). The AO-corrected observations y_t^* are finally recovered by inverting Eq. (B-1).
- Step 5: Go back to Step 1 with the y_t^* series, and repeat all steps until no \hat{AO} test statistic values exceed C , i.e. until there appear to be no more AOs in the data.

Appendix C

Appendix Chapter 7

Forecasts

For the two-regime model, the variance forecast at time t of a single observation at time $t + j$ (denoted s_{t+j}^2) is computed as:¹

$$\begin{aligned} s_{t+j}^2 &\equiv \text{var} [y_{t+j} \mid \tilde{y}_t] \\ &= E_t [y_{t+j}^2] - E_t [y_{t+j}]^2 \\ &= p_{1t,t+j} (\sigma_1^2 + \mu_1^2) + (1 - p_{1t,t+j}) (\sigma_2^2 + \mu_2^2) \\ &\quad - [p_{1t,t+j} \mu_1 + (1 - p_{1t,t+j}) \mu_2]^2, \end{aligned}$$

where $\tilde{y}_t = \{y_t, y_{t-1}, \dots\}$ and $p_{1t,t+j} = \Pr [S_{t+j} = 1 \mid \tilde{y}_t]$ which is the first element in a two-element vector of regime probabilities for time $t + j$ given by

$$p_{t+j} = p_t' P^j.$$

The j -week variance forecast is then

$$j - \text{week} = \sum_{i=1}^j s_{t+i}^2.$$

For the four-regime forecasts, the variance forecasts are constructed in a similar way. For example, for a one-week forecast, we have:

$$\begin{aligned} E_t [\sigma_{t+1}^2 \mid \tilde{y}_t] &= E_t [y_{t+1}^2] - E_t [y_{t+1}]^2 \\ &= p_{1t,t+1} (\sigma_1^2 + \mu_1^2) + p_{2t,t+1} (\sigma_1^2 + \mu_2^2) \\ &\quad + p_{3t,t+1} (\sigma_2^2 + \mu_1^2) + p_{4t,t+1} (\sigma_2^2 + \mu_2^2) \\ &\quad - [(p_{1t,t+1} + p_{3t,t+1}) \mu_1 + (p_{2t,t+1} + p_{4t,t+1}) \mu_2]^2. \end{aligned}$$

¹See Bollen, Gray, and Whaley (2000).

Expected maximum likelihood procedure

This Markov-Switching model is estimated by the so-called Expected Maximum Likelihood (EML) procedure. As the name of this procedure suggests, it involves two main steps: the computation of the probabilities of being in each regime (the “expectation” part) and the maximization of the sum of log-likelihoods conditional on each regime (the “maximum likelihood” part). Basically, the estimation procedure may be described in three steps.

Step 1: Initialization of the filter at time $t = 0$.

At time t and for given values of P , one has to compute the steady-state probabilities (ergodic probabilities) $\pi_j = \text{Prob}(s_0 = j)$ in order to initialize the filter. For instance, in the two state case, this can be computed in a straightforward way:

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{22} - p_{11}} \quad (\text{C-3})$$

$$\pi_2 = \frac{1 - p_{11}}{2 - p_{22} - p_{11}}. \quad (\text{C-4})$$

Step 2: Computation of the log-likelihood for observation t .

The log-likelihood function is written as:

$$L_{\text{markov}} = \sum_{t=1}^T \ln \Delta(y_t | \Omega_{t-1}), \quad (\text{C-5})$$

where Δ is the Gaussian density. Starting with the conditional joint density of y_t , s_t and s_{t-1} , one can write:

$$\Delta(y_t, s_t, s_{t-1} | \Omega_{t-1}) = \Delta(y_t | s_t, s_{t-1}, \Omega_{t-1}) \text{Prob}(s_t, s_{t-1} | \Omega_{t-1}). \quad (\text{C-6})$$

The idea is to get $\Delta(y_t | \Omega_{t-1})$ by summing this joint density over all possible values of s_t and s_{t-1} . Again, in the two state case:

$$\Delta(y_t | \Omega_{t-1}) = \sum_{s_t=1}^2 \sum_{s_{t-1}=1}^2 \Delta(y_t | s_t, s_{t-1}, \Omega_{t-1}) \text{Prob}(s_t, s_{t-1} | \Omega_{t-1}). \quad (\text{C-7})$$

Using (C-5), the log-likelihood may be then computed for given values of $\text{Prob}(s_t, s_{t-1} | \Omega_{t-1})$.

Step 3: Filtering step; Calculate $\text{Prob}(s_t, s_{t-1}|\Omega_t)$.

For $t > 1$, update $\text{Prob}(s_t, s_{t-1}|\Omega_{t-1})$, using information up to time t :

$$\text{Prob}(s_t, s_{t-1}|\Omega_t) = \frac{\Delta(y_t, s_t, s_{t-1}|\Omega_{t-1})}{\Delta(y_t|\Omega_{t-1})}. \quad (\text{C-8})$$

This updated joint probability at time t is then used to apply again step 2 to the next observation, i.e. observation at time $t + 1$. From (C-8), it is also possible to compute the (2×1) vector of filtered probabilities:

$$\text{Prob}(s_t|\Omega_t) = \sum_{s_{t-1}=1}^2 \text{Prob}(s_t, s_{t-1}|\Omega_{t-1}) \quad (\text{C-9})$$

Bibliography

- ALMEKINDERS, G., AND S. EIJJFINGER (1993): "Accounting for Daily Bundesbank and Federal Reserve Intervention," Tilburg University, Mimeo.
- ANDERSEN, T., AND T. BOLLERSLEV (1997): "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, 4, 115–158.
- (1998): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.
- ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2000): "Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian," *Multinational Finance Journal*, 4, 159–179.
- (2001): "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.
- ANDERSEN, T., T. BOLLERSLEV, AND S. LANGE (1999): "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance*, 6, 457–477.
- ANDERSSON, M. (2000): "Do Long-Memory Models Have Long Memory?," *International Journal of Forecasting*, 16, 121–124.
- AREAL, N., AND S. TAYLOR (2000): "The Realized Volatility of FTSE-100 Futures Prices," Forthcoming in *Journal of Futures Markets*.
- AZZALINI, A., AND A. CAPITANIO (1996): "The Multivariate Skew-Normal Distribution," *Biometrika*, 83, 715–726.
- BAILLIE, R. (1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

- BAILLIE, R., AND T. BOLLERSLEV (1989): “The Message in Daily Exchange Rates: A Conditional-Variance Tale,” *Journal of Business and Economic Statistics*, 7, 297–305.
- BAILLIE, R., T. BOLLERSLEV, AND H. MIKKELSEN (1996): “Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 74, 3–30.
- BAILLIE, R., C. CHUNG, AND M. TIESLAU (1996): “Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model,” *Journal of Applied Econometrics*, 11, 23–40.
- BAILLIE, R., AND O. HUMPAGE (1992): “Post Louvre Intervention: Does Target Zones Stabilize the Dollar ?,” Federal Reserve Bank of Cleveland, Working Paper 9203.
- BAILLIE, R., AND W. OSTERBERG (1997a): “Central Bank Intervention and Risk in the Forward Market,” *Journal of International Economics*, 43, 483–497.
- (1997b): “Why do Central Banks Intervene ?,” *Journal of International Money and Finance*, 16, 909–919.
- BALL, C., AND A. ROMA (1993): “A Jump Diffusion Model for the European Monetary System,” *Journal of International Money and Finance*, 12, 475–492.
- BAUWENS, L., P. GIOT, J. GRAMMIG, AND D. VEREDAS (2000): “A Comparison of Financial Duration Models Via Density Forecasts,” CORE DP 2060.
- BAUWENS, L., AND S. LAURENT (2002): “A New Class of Multivariate Skew Densities, with Application to GARCH Models,” CORE DP 2002/20.
- BAUWENS, L., S. LAURENT, AND J. ROMBOUTS (2002): “A review of Multivariate GARCH Models with Applications to Financial Data,” unpublished manuscript.
- BEINE, M., A. BÉNASSY-QUÉRÉ, AND C. LECOURT (2002): “Central Bank Intervention and Foreign Exchange Rates: New Evidence from FIGARCH Estimations,” *Journal of International Money and Finance*, 21, 115–144.

- BEINE, M., AND S. LAURENT (1999): “Central Bank Interventions and Jumps in Double Long Memory Models of Daily Exchange Rates,” Mimeo, University of Liège.
- (2000): “La Persistance des Chocs de Volatilité sur le Marché des Changes s’est-elle Modifiée depuis le Début des Années Quatre-vingts ?,” *La Revue Economique*, 3, 703–711.
- (2001): “Structural Change and Long Memory in Volatility: New Evidence from Daily Exchange Rates,” in *Developments in Forecast Combination and Portfolio Choice*, ed. by C. Dunis, A. Timmermann, and J. Moody, chap. 6, pp. 145–157. J. Wiley & Sons.
- BEINE, M., S. LAURENT, AND C. LECOURT (2000): “Accounting for Conditional Leptokurtosis and Closing Days Effects in FIGARCH Models of Daily Exchange Rates,” Forthcoming in *Applied Financial Economics*.
- (2001): “Official Central Bank Interventions and Exchange Rate Volatility: Evidence from a Regime Switching Analysis,” mimeo, University of Liège.
- BERA, A., AND M. HIGGINS (1993): “ARCH Models: Properties, Estimation and Testing,” *Journal of Economic Surveys*.
- BLACK, F. (1976): “Studies of Stock Market Volatility Changes,” *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.
- BLAIR, B., S. POON, AND S. TAYLOR (2000): “Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns,” Forthcoming in *Journal of Applied Econometrics*.
- BOLLEN, P., S. GRAY, AND R. WHALEY (2000): “Regime-Switching in Foreign Exchange Rates: Evidence From Currency Option Prices,” *Journal of Econometrics*, 94, 239–276.
- BOLLERSLEV, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

- (1987): “A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return,” *Review of Economics and Statistics*, 69, 542–547.
- (1990): “Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model,” *Review of Economics and Statistics*, 72, 498–505.
- BOLLERSLEV, T., R. CHOU, AND K. KRONER (1992): “ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence,” *Journal of Econometrics*, 52, 5–59.
- BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1994): “ARCH Models,” in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, chap. 4, pp. 2959–3038. North Holland Press, Amsterdam.
- BOLLERSLEV, T., R. ENGLE, AND J. WOOLDRIDGE (1988): “A Capital Asset Pricing Model with Time Varying Covariances,” *Journal of Political Economy*, 96, 116–131.
- BOLLERSLEV, T., AND E. GHYSELS (1996): “Periodic Autoregressive Conditional Heteroskedasticity,” *Journal of Business and Economics Statistics*, 14, 139–152.
- BOLLERSLEV, T., AND H. O. MIKKELSEN (1996): “Modeling and Pricing Long-Memory in Stock Market Volatility,” *Journal of Econometrics*, 73, 151–184.
- BOLLERSLEV, T., AND J. WOOLDRIDGE (1992): “Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances,” *Econometric Reviews*, 11, 143–172.
- BOND, S. (2000): “A Review of Asymmetric Conditional Density Functions in Autoregressive Conditional Heteroscedasticity Models,” mimeo, Duke University, Durham.
- BONSER-NEAL, C., AND G. TANNER (1996): “Central Bank Intervention and the Volatility of Foreign Exchange Rates: Evidence from the Options Market,” *Journal of International Money and Finance*, 15, 853–878.

- BOX, G., AND G. JENKINS (1970): *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.
- BRANCO, M., AND D. DEY (2000): "A class of Multivariate Skew-Elliptical Distributions," Forthcoming in *Journal of Multivariate Analysis*.
- BRENNAN, M. (1979): "The Pricing of Contingent Claims in Discrete Time Models," *Journal of Finance*, 34, 53–68.
- BROOKS, C., S. BURKE, AND G. PERSAUD (2001): "Benchmarks and the Accuracy of GARCH Model Estimation," *International Journal of Forecasting*, 17, 45–56.
- CATTE, P., G. GALLI, AND S. REBECCHINI (1992): "Exchange Markets Can Be Managed!," Report on the G-7, International Economic Insights.
- CHANG, I., G. C. TIAO, AND C. CHEN (1988): "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193204.
- CHEN, C., AND L. LIU (1993): "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, 284296.
- CHRISTOFFERSEN, P., AND F. DIEBOLD (2000): "How Relevant is Volatility Forecasting for Financial Risk Management?," *Review of Economics and Statistics*, 82, 1–11.
- CHUNG, C.-F. (1999): "Estimating the Fractionally Integrated GARCH Model," National Taiwan University working paper.
- CHUNG, C.-F., AND R. BAILLIE (1993): "Sample Bias in Conditional Sum-of-Squares Estimator of Fractionally Integrated ARMA Models," *Empirical Economics*, 18, 791–806.
- CONNOLLY, R., AND W. TAYLOR (1994): "Volume and Intervention Effects on the Yen/Dollar Volatility, 1977-1979," in *Advanced Financial Planning and Forecasting*, chap. 5, pp. 181–200. JAI Press, Greenwich, Connecticut.
- CRIBARI-NETO, F., AND S. ZARKOS (2001): "Econometric and Statistical Computing Using Ox," Forthcoming in *Computational Economics*.

- DANIELSSON, J., AND C. DE VRIES (2000): “Value-at-Risk and Extreme Returns,” *Annales d’Economie et Statistique*, 3, 73–85.
- DAVIDSON, J. (2001): “Moment and Memory Properties of Linear Conditional Heteroscedasticity Models,” Manuscript, Cardiff University.
- DE CEUSTER, M., AND D. TRAPPERS (1992): “Diagnostic Checking of Estimation with a Student-t Error Density,” Working paper UFSIA, Centrum voor Bedrijfeconomie en Bedrijfeconometrie, 173.
- DIEBOLD, F., J.-H. LEE, AND G. WEINBACH (1994): “Regime Switching with Time-Varying Transition Probabilities,” in *Nonstationary Time Series Analysis and Cointegration*, ed. by C. Hargreaves, pp. 283–302. Oxford University Press, Oxford.
- DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (1998): “Evaluating Density Forecasts, with Applications to Financial Risk Management,” *International Economic Review*, 39, 863–883.
- DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): “A Long Memory Property of Stock Market Returns and a New Model,” *Journal of Empirical Finance*, 1, 83–106.
- DOMINGUEZ, K. (1998): “Central Bank Intervention and Exchange Rate Volatility,” *Journal of International Money and Finance*, 17, 161–190.
- (1999): “The Market Microstructure of Central Bank Intervention,” NBER Working Paper 7337.
- DOMINGUEZ, K., AND J. FRANKEL (1993): “Does Foreign Exchange Intervention Work ?,” Institute for International Economics, Washington, DC.
- DOORNIK, J. A. (1999): *An Object Oriented Matrix Programming Language*. Timberlake Consultant Ltd., London.
- DOORNIK, J. A., AND M. OOMS (1999): “A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima package 1.0 for Ox,” Discussion paper, Econometric Institute, Erasmus University Rotterdam.

- DRÈZE, J. (1978): “Bayesian Regression Analysis using poly-t Densities,” *Journal of Econometrics*, 6, 329–354.
- EBENS, H. (1999): “Realized Stock Index Volatility,” Working Paper No. 420, Department of Economics, John Hopkins University, Baltimore.
- EL BABSIRI, M., AND J.-M. ZAKOIAN (1999): “Contemporaneous Asymmetry in GARCH Processes,” Forthcoming in *Journal of Econometrics*.
- ENGLE, R. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987–1007.
- (1995): *ARCH Selected Readings*. Oxford University Press, Oxford.
- (2001): “Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models,” Mimeo, UCSD.
- ENGLE, R., AND T. BOLLERSLEV (1986): “Modeling the Persistence of Conditional Variances,” *Econometric Reviews*, 5, 1–50.
- ENGLE, R., AND G. GONZÁLEZ-RIVERA (1991): “Semiparametric ARCH Model,” *Journal of Business and Economic Statistics*, 9, 345–360.
- ENGLE, R., AND F. KRONER (1995): “Multivariate Simultaneous Generalized ARCH,” *Econometric Theory*, 11, 122–150.
- ENGLE, R., AND G. LEE (1999): “A Permanent and Transitory Component Model of Stock Return Volatility,” in *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, ed. by R. Engle, and H. White, pp. 475–497. Oxford University Press, Oxford.
- ENGLE, R., AND V. NG (1993): “Measuring and Testing the Impact of News on Volatility,” *Journal of Finance*, 48, 1749–1778.
- ENGLE, R., AND A. PATTON (1999): “What Good is a Volatility Model?,” Mimeo, San Diego, Department of Economics.
- FERNÁNDEZ, C., AND M. STEEL (1998): “On Bayesian Modelling of Fat Tails and Skewness,” *Journal of the American Statistical Association*, 93, 359–371.

- FILARDO, A. (1994): “Business Cycles Phases and their Transitions,” *Journal of Business and Economics Statistics*, 12, 299–308.
- (1998): “Choosing Information Variables for Transition Probabilities in a Time-Varying Transition Probability Markov Switching Model,” Federal Reserve Bank of Kansas City RWP 98-09.
- FIorentini, G., G. CALZOLARI, AND L. PANATTONI (1996): “Analytic Derivatives and the Computation of GARCH Estimates,” *Journal of Applied Econometrics*, 11, 399–417.
- FIorentini, G., E. SENTANA, AND G. CALZOLARI (2000): “The Score of Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations and an LM Test for Multivariate Normality,” Working paper N 0007, CEMFI.
- FRANSES, P., AND H. GHIJSELS (1999): “Additive Outliers, GARCH and Forecasting Volatility,” *International Journal of Forecasting*, 15, 1–9.
- FRANSES, P., AND D. VAN DIJK (2000): *Non-Linear Series Models in Empirical Finance*. Cambridge University Press, Cambridge.
- FRENCH, K., G. SCHWERT, AND R. STAMBAUGH (1987): “Expected Stock Returns and Volatility,” *Journal of Financial Economics*, 19, 3–29.
- GABLE, J., S. VAN NORDEN, AND R. VIGFUSSON (1997): “Analytical Derivatives for Markov Switching Models,” *Computational Economics*, 10, 187–194.
- GALLANT, R., D. HSIEH, AND G. TAUCHEN (1991): “On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974-1983,” in *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen, pp. 199–240. Cambridge University Press, Cambridge.
- GALLANT, R., AND G. TAUCHEN (1989): “Seminonparametric Estimation of Conditionally Constrained Heterogenous Processes: Asset Pricing Applications,” *Econometrica*, 57, 1091–1120.
- GARCIA, R. (1998): “Asymptotic Null Distribution of the Likelihood Ratio Test in Markov-Switching Models,” *International Economic Review*, 39, 763–788.

- GEWEKE, J. (1986): “Modeling the Persistence of Conditional Variances: A Comment,” *Econometric Reviews*, 5, 57–61.
- (1989): “Bayesian Inference in Econometric Models Using Monte Carlo Integration,” *Econometrica*, 57, 1317–1339.
- GEWEKE, J., AND G. AMISANO (2001): “Compound Markov Mixture Models with Application in Finance,” Mimeo, University of Iowa.
- GEWEKE, J., AND S. PORTER-HUDAK (1983): “The Estimation and Application of Long Memory Time Series Models,” *Journal of Time Series Analysis*, 4, 221–238.
- GIOT, P. (2000): “Intraday Value-at-Risk,” CORE DP 2045 and Maastricht University METEOR RM/00/030.
- GIOT, P., AND S. LAURENT (2001a): “Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models,” Maastricht University METEOR RM/01/026.
- (2001b): “Quantifying Market Risk for Long and Short Traders,” Forthcoming in *European Investment Review*.
- (2001c): “Value-at-Risk for Long and Short Positions,” CORE DP 2001-22 and Maastricht University METEOR RM/01/005.
- GLOSTEN, L., R. JAGANNATHAN, AND D. RUNKLE (1993): “On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks,” *Journal of Finance*, 48, 1779–1801.
- GOURIEROUX, C. (1997): *ARCH Models and Financial Applications*. Springer Verlag, Berlin.
- GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): “Pseudo Maximum Likelihood Methods: Theory,” *Econometrica*, 52, 681–700.
- GRANGER, C. (1980): “Long Memory Relationships and the Aggregation of Dynamic Models,” *Journal of Econometrics*, 14, 227–238.

- GRANGER, C., AND N. HYUNG (1999): “Occasional Structural Breaks and Long Memory,” UCSD Discussion Paper 99-14, June 1999.
- GRANGER, C., AND R. JOYEUX (1980): “An Introduction to Long-Memory Time Series Models and Fractional Differencing,” *Journal of Time Series Analysis*, 1, 15–29.
- GRAY, S. (1996): “Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process,” *Journal of Financial Economics*, 42, 27–62.
- HAFNER, C. (2001): “Fourth Moment of Multivariate GARCH Processes,” CORE DP 2001-39.
- HAMILTON, J. (1994): *Time Series Analysis*. Princeton University Press, Princeton.
- HANSEN, B. (1992): “The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP,” *Journal of Applied Econometrics*, 7, S61–S82.
- (1994): “Autoregressive Conditional Density Estimation,” *International Economic Review*, 35, 705–730.
- HARDLE, W., AND C. HAFNER (2000): “Discrete Time Option Pricing with Flexible Volatility Estimation,” *Finance and Stochastics*, 4, 189–207.
- HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): “Multivariate Stochastic Variance Models,” *Review of Economic Studies*, 61, 247–264.
- HARVEY, C., AND A. SIDDIQUE (1999): “Autoregressive Conditional Skewness,” *Journal of Financial and Quantitative Analysis*, 34, 465–487.
- HE, C., AND T. TERÄSVIRTA (1999a): “Higher-order Dependence in the General Power ARCH Process and a Special Case,” Stockholm School of Economics, Working Paper Series in Economics and Finance, No. 315.
- (1999b): “Statistical Properties of the Asymmetric Power ARCH Process,” in *Cointegration, causality, and forecasting. Festschrift in honour of Clive W.J. Granger*, ed. by R. Engle, and H. White, chap. 19, pp. 462–474. Oxford University Press, Oxford.

- HE, H., AND H. LELAND (1993): "On Equilibrium Asset Price Processes," *Review of Financial Studies*, 6, 593–617.
- HENDRICKS, D. (1996): "Evaluation of Value-at-Risk Models using Historical Data," *Federal Reserve Bank of New York Economic Policy Review*, April 1996.
- HIGGINS, M., AND A. BERA (1992): "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.
- HONG, C. (1988): "Options, Volatilities and the Hedge Strategy," Unpublished Ph.D. diss., University of San Diego, Dept. of Economics.
- HOSKING, J. (1981): "Fractional Differencing," *Biometrika*, 68, 165–176.
- HOTTA, L., AND R. TSAY (1998): "Outliers in GARCH Process," Unpublished Manuscript.
- HSIEH, D. (1989): "Modeling Heteroskedasticity in Daily Foreign Exchange Rates," *Journal of Business and Economic Statistics*, 7, 307–317.
- HURVICH, C. M., R. DEO, AND J. BRODSKY (1998): "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Long-Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 19, 19–46.
- JARQUE, C., AND A. BERA (1987): "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.
- JEANTHEAU, T. (1998): "Strong Consistency of Estimators for Multivariate ARCH models," *Econometric Theory*, 14, 70–86.
- JOHNSON, M. (1987): *Multivariate Statistical Simulation*. J. Wiley & Sons, New York.
- JONDEAU, E., AND M. ROCKINGER (2000): "Conditional Volatility, Skewness, and Kurtosis: Existence and Persistence," Mimeo, HEC-School of Management, Jouyen-Josas, France.
- JONES, M. (2000): "Multivariate T and Beta Distributions Associated with the Multivariate F Distribution," Forthcoming in *Metrika*.

- JONES, M., AND M. FADDY (2000): “A Skew Extension of the t Distribution, with Applications,” mimeo, Department of Statistics, Open University, Walton Hall, UK.
- JORION, P. (1988): “On Jump Processes in the Foreign Exchange and Stock Markets,” *The Review of Financial Studies*, 68, 165–176.
- (1996): “Risk and Turnover in the Foreign Exchange Market,” in *The Microstructure of Foreign Exchange Markets*, ed. by J. Frankel, G. Galli, and A. Giovanni. The University of Chicago Press, Chicago.
- (2000): *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York.
- KIM, C. (1994): “Dynamic linear Models with Markov Switching,” *Journal of Econometrics*, 60, 1–22.
- KIM, C.-J., AND C. NELSON (1999): *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press, Cambridge.
- KIM, D., AND S. KON (1999): “Structural Change and Time Dependence in Models of Stock Returns,” *Journal of Econometrics*, 6, 283–308.
- KNIGHT, J., S. SATCHELL, AND K. TRAN (1995): “Statistical Modelling of Asymmetric Risk in Asset Returns,” *Applied Mathematical Finance*, 2, 155–172.
- KON, S. (1982): “Models of Stock Returns, a Comparison,” *Journal of Finance*, 39, 147–165.
- KOOPMAN, S., N. SHEPARD, AND J. DOORNIK (1998): “Statistical Algorithms for Models in State Space using SsfPack 2.2,” *Econometrics Journal*, 1, 1–55.
- KRAFT, D., AND R. ENGLE (1982): “Autoregressive Conditional Heteroskedasticity in Multiple Time Series,” unpublished manuscript, Department of Economics, UCSD.
- KRAUS, A., AND R. LITZENBERGER (1976): “Skewness Preference and the Valuation of Risky Assets,” *Journal of Finance*, 31, 1085–1100.

- KRONER, F., AND V. NG (1998): “Modelling Asymmetric Comovements of Asset Returns,” *The Review of Financial Studies*, 11, 817–844.
- KUPIEC, P. (1995): “Techniques for Verifying the Accuracy of Risk Measurement Models,” *Journal of Derivatives*, 2, 173–84.
- LAMBERT, P., AND S. LAURENT (2000): “Modelling Skewness Dynamics in Series of Financial Data,” Discussion Paper, Institut de Statistique, Louvain-la-Neuve.
- (2001): “Modelling Financial Time Series Using GARCH-Type Models and a Skewed Student Density,” Mimeo, Université de Liège.
- LAURENT, S. (2001): “Analytical Derivates of the APARCH Model under a Skewed Student Assumption for the Innovation Process,” Mimeo, Université de Liège.
- LAURENT, S., AND J.-P. PETERS (2001): “G@RCH 2.2 : An Ox Package for Estimating and Forecasting Various ARCH Models,” Forthcoming in *Journal of Economic Surveys*.
- LECOURT, C. (2000): “Dépendance de Court et Long Terme des Rendements de Taux de Change,” *Economie et Prévision*, 5, 127–137.
- LEE, S., AND B. HANSEN (1994): “Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models,” *Econometric Theory*, 10, 29–52.
- LEE, Y., AND T. TSE (1991): “Term Structure of Interest Rates in the Singapore Asian Dollar Market,” *Journal of Applied Econometrics*, 6, 143–152.
- LING, S., AND M. MCALEER (2002): “Stationarity and the Existence of Moments of a Family of GARCH Processes,” *Journal of Econometrics*, 106, 109–117.
- LIU, S.-M., AND B. BRORSEN (1995): “Maximum Likelihood Estimation of a GARCH - STABLE Model,” *Journal of Applied Econometrics*, 2, 273–285.
- LOMBARDI, M., AND G. GALLO (2001): “Analytic Hessian Matrices and the Computation of FIGARCH Estimates,” Manuscript, Università degli studi di Firenze.

- LUMSDAINE, R. (1996): “Asymptotic Properties of the Quasi Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models,” *Econometrica*, 64, 575–596.
- MACDONALD, R. (2000): “Expectations Formation and Risk in Three Financial Markets: Surveying What the Surveys Say,” *Journal of Economic Surveys*, 14, 69–100.
- MAULEÓN, I., AND J. PEROTE (1999): “Estimation of Multivariate Densities with Financial Data: the Performance of the Multivariate Edgeworth-Sargan Density,” Proceedings of the 12th Australian Finance and Banking Conference, Sidney.
- MCCULLOCH, J. (1985): “Interest-Risk Sensitive Deposit Insurance Premia: Stable ACH Estimates,” *Journal of Banking and Finance*, 9, 137–156.
- MCCULLOUGH, B., AND H. VINOD (1999): “The Numerical Reliability of Econometric Software,” *Journal of Economic Literature*, 37, 633–665.
- MCDONALD, J. (1984): “Some Generalized Functions for the Size Distribution of Income,” *Econometrica*, 53, 647–663.
- (1991): “Parametric Models for Partially Adaptive Estimation with Skewed and Leptokurtic Residuals,” *Economics Letters*, 37, 272–278.
- MCLEOD, A., AND W. LI (1983): “Diagnostic Checking ARMA Time Series Models Using Squared Residuals Autocorrelations,” *Journal of Time Series Analysis*, 4, 269–273.
- MERTON, R. (1980): “On Estimating the Expected Return on the Market. An Exploratory Investigation,” *Journal of Financial Economics*, 8, 323–361.
- MITTNIK, S., AND M. PAOLELLA (2000): “Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates,” *Journal of Forecasting*, 19, 313–333.
- MITTNIK, S., M. PAOLELLA, AND S. RACHEV (1998): “The Prediction of Downside Market Risk with GARCH-stable Models,” Working paper of the Institute for Statistics and Econometrics at Kiel University, n 118.

- MUNDACA, B. (2001): “Central Bank Interventions and Exchange Rate Bands,” *Journal of International Money and Finance*, 20, 677–700.
- MURRAY, J., M. ZELMER, AND D. MCMANUS (1996): “The Effect of Intervention on Canadian Dollar Volatility,” Bank of Canada, Mimeo.
- NEELY, C. (1999): “Target Zones and Conditional Volatility: the Role of Realignments,” *Journal of Empirical Finance*, 6, 177–192.
- (2000): “The Practice of Central Bank Intervention: Looking Under the Hood,” Federal Reserve Bank Working Paper no. 2000-028.
- NELSON, D. (1991): “Conditional Heteroskedasticity in Asset Returns: a New Approach,” *Econometrica*, 59, 349–370.
- NELSON, D., AND C. CAO (1992): “Inequality Constraints in the Univariate GARCH Model,” *Journal of Business and Economic Statistics*, 10, 229–235.
- NEWKEY, W., AND D. STEIGERWALD (1997): “Asymptotic Bias for Quasi Maximum Likelihood Estimators in Conditional Heteroskedasticity Models,” *Econometrica*, 3, 587–599.
- NG, V., R. ENGLE, AND M. ROTHSCHILD (1992): “A Multi-Dynamic Factor Model for Stock Returns,” *Journal of Econometrics*, 52, 245–265.
- NYBLOM, J. (1989): “Testing for the Constancy of Parameters Over Time,” *Journal of the American Statistical Association*, 84, 223–230.
- OOMEN, R. (2001): “Using High Frequency Stock Market Index Data to Calculate, Model and Forecast Realized Volatility,” Manuscript, European University Institute, Department of Economics.
- OOMS, M., AND J. A. DOORNIK (1998): “Estimation, Simulation and Forecasting for Fractional Autoregressive Integrated Moving Average Models,” Discussion paper, Econometric Institute, Erasmus University Rotterdam, presented at the fourth annual meeting of the Society for Computational Economics, June 30, 1998, Cambridge, UK.
- PAGAN, A. (1996): “The Econometrics of Financial Markets,” *Journal of Empirical Finance*, 3, 15–102.

- PAGAN, A., AND Y. HONG (1990): “Non-Parametric Estimation and the Risk Premium,” in *Semiparametric and Nonparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge University Press, Cambridge.
- PAGAN, A., AND G. SCHWERT (1990): “Alternative Models for Conditional Stock Volatility,” *Journal of Econometrics*, 45, 267–290.
- PALM, F. (1996): “GARCH Models of Volatility,” in *Handbook of Statistics*, ed. by G. Maddala, and C. Rao, pp. 475–497. Elsevier Science, Amsterdam.
- PALM, F., AND P. VLAAR (1997): “Simple Diagnostics Procedures for Modelling Financial Time Series,” *Allgemeines Statistisches Archiv*, 81, 85–101.
- PAOLELLA, M. S. (1997): “Using Flexible GARCH Models with Asymmetric Distributions,” Working paper, Institute of Statistics and Econometrics Christian Albrechts University at Kiel.
- PEIRÓ, A. (1999): “Skewness in Financial Returns,” *Journal of Banking and Finance*, 23, 847–862.
- PENTULA, S. (1986): “Modeling the Persistence of Conditional Variances: A Comment,” *Econometric Reviews*, 5, 71–74.
- RAVN, M., AND M. SOLA (1995): “Stylized Facts and Regime Changes: Are Prices Procyclical?,” *Journal of Monetary Economics*, 36, 497–526.
- RICHARD, J.-F., AND H. TOMPA (1980): “On the Evaluation of poly-t Density Functions,” *Journal of Econometrics*, 12, 335–351.
- SAHU, S., D. DEY, AND D. BRANCO (2001): “A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models,” mimeo, Department of Statistics, Univerisity of Sao Paulo.
- SCAILLET, O. (2000): “Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall,” Mimeo, Université Catholique de Louvain, IRES.
- SCHWERT, W. (1990): “Stock Volatility and the Crash of ’87,” *Review of Financial Studies*, 3, 77–102.

- SHARPE, W., G. ALEXANDER, AND J. BAILEY (1999): *Investments*. Prentice-Hall, New York.
- SHEPHARD, N. (1996): “Statistical Aspects of ARCH and Stochastic Volatility,” in *Time Series Models: In Econometrics, Finance and Other Fields*, ed. by D. Cox, D. Hinkley, and O. Barndorff-Nielsen, chap. 1, pp. 1–67. Chapman & Hall, London.
- SIMKOWITZ, M., AND W. BEEDLES (1980): “Asymmetric Stable Distributed Security Returns,” *Journal of the American Statistical Association*, 75, 306–312.
- SO, J. (1987): “The Distribution of Foreign Exchange Price Changes: Trading Day Effects and Risk measurement - A Comment,” *Journal of Finance*, 42, 181–188.
- SOWEL, F. (1992): “Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models,” *Journal of Econometrics*, 53, 165–188.
- TAYLOR, S. (1986): *Modelling Financial Time Series*. J. Wiley & Sons, New York.
- (1994): “Modeling Stochastic Volatility: A Review and Comparative Study,” *Mathematical Finance*, 4, 183–204.
- TAYLOR, S., AND X. XU (1997): “The Incremental Volatility Information in One Million Foreign Exchange Quotations,” *Journal of Empirical Finance*, 4, 317–340.
- TEYSSIÈRE, G. (1997): “Double Long-Memory Financial Time Series,” Paper presented at the ESEM, Toulouse.
- TIMMERMANN, A. (2000): “Density Forecasting in Economics and Finance,” *Journal of Forecasting*, 19, 120–123.
- TSCHERNIG, R. (1995): “Long Memory in Foreign Exchange Rates Revisited,” *Journal of International Financial Markets, Institutions and Money*, 5, 53–78.
- TSE, Y. (1998): “The Conditional Heteroscedasticity of the Yen-Dollar Exchange Rate,” *Journal of Applied Econometrics*, 13, 49–55.

- TSE, Y., AND A. TSUI (1998): "A Multivariate GARCH Model with Time-Varying Correlations," Mimeo, Department of Economics, National University of Singapore.
- VAN DEN GOORBERGH, R., AND P. VLAAR (1999): "Value-at-Risk Analysis of Stock Returns. Historical Simulation, Variance Techniques or Tail Index Estimation?," De Nederlandse Bank-Staff Report, 40.
- VLAAR, P., AND F. PALM (1993): "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps," *Journal of Business and Economic Statistics*, 11, 351–360.
- VON ROHR, P., AND I. HOESCHELE (1999): "Robust Bayesian Analysis Using Skewed Student-t Distributions," 50th Annual Meeting of EAAP, August 22-26, Zurich, Switzerland.
- WANG, K.-L., C. FAWSON, C. BARRETT, AND J. McDONALD (2001): "A Flexible Parametric GARCH Model with an Application to Exchange Rates," *Journal of Applied Econometrics*, 16, 521–536.
- WEBER, A. (1996): "Intervention, International Policy Coordination and the Future of EMU: A German Perspective," in *The New Transatlantic Economy*, ed. by M. Canzoneri, W. Ethier, and V. Grilli, pp. S. 54–113. Cambridge University Press, Cambridge.
- WEISS, A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131.
- WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.
- ZAKOIAN, J.-M. (1994): "Threshold Heteroskedasticity Models," *Journal of Economic Dynamics and Control*, 15, 931–955.

Nederlandse Vertaling

Het is algemeen gekend dat aandelen returns, waargenomen op een intra-dagelijkse, dagelijkse en wekelijkse basis, in het algemeen serieel gecorreleerd en dikwijls heteroskedastisch zijn. Dikwijls hebben de verdelingen van deze returns dikke staarten en zijn ze scheef.

Het doel van deze thesis was een model te vinden dat de hierboven genoemde eigenschappen van de returns kan reproduceren, en om de belangrijkheid aan te tonen van dichtheidsfuncties die verschillen van de normale dichtheidsfunctie in financiële toepassingen. Een AR-APARCH specificatie werd in staat bevonden de eerste twee conditionele momenten te beschrijven van de onderzochte series. Om rekening te houden met de mogelijke asymmetrie and de hoge kurtosis die we empirisch waarnemen, hebben we voorgesteld om gebruik te maken van de scheef verdeelde student dichtheid van Fernández and Steel (1998). We breiden deze dichtheid uit in twee richtingen.

Vooreerst hebben we deze dichtheid geherformuleerd opdat deze een nul-gemiddelde en unitaire variantie heeft. We geven ook een analytische uitdrukking van de gradient vector voor dit model. We hebben ook de gebruiksvriendelijkheid aangetoond van deze dichtheid (gemakkelijke implementatie, een duidelijke interpretatie van de parameters en de empirische relevantie) ten aanzien van haar concurrenten. Vanuit een empirisch perspectief werd aangetoond dat een AR-APARCH gecombineerd met een scheve student verdeling accurate VaR voorspellingen gaf.

Het meest uitdagende deel van deze thesis was het ontwikkelen van een software pakket voor het schatten en voorspellen van de meest populaire univariate modellen van het ARCH type. Dit pakket, genaamd G@RCH, werd ontwikkeld met de OX 3.0 matrix georiënteerde programmatietaal van Doornik (1999). Het biedt een gebruiksvriendelijke venster georiënteerde interface aan te vergelijken met

de welgekende PcGive software.

De laatste contributie van deze thesis is nog steeds gerelateerd met het gebruik van niet normale verdelingen. Het idee was om het effect te bestuderen van centrale bank interventies op de wekelijkse returns en de volatiliteit van de DEM/USD en de YEN/USD wisselkoersen. Hier, in tegenstelling tot de vorige hoofdstukken, laten we regime-afhankelijke specificaties toe (dit is een extensie van normale mixtuur voorgesteld in het derde hoofdstuk) en we onderzoeken of officiele interventies de geobserveerde regime veranderingen in de volatiliteit kunnen verklaren. De schattingsresultaten werpen een interessant licht op de conclusies gegeven in de literatuur. Het werd bevonden dat afhankelijk van het gegeven volatiliteits niveau, gecordineerde centrale bank interventies een stabiliserend of een destabiliserend effect kunnen hebben. Onze resultaten zijn in opspraak met de algemene visie dat deze interventies enkel stijgingen in de volatiliteit impliceren.

Curriculum Vitae

Sébastien Laurent was born on June 4th, 1974 in Huy, Belgium. From 1992 until 1996 he studied Economics (BA) at the University of Liège and from 1996 until 1998 he did a Masters degree (DEA) in Economics at the University of Liège and at Maastricht University.

From December 1998 until June 2002, he was a research assistant in Econometrics at the Department of Economics at the University of Liège. His research, in financial econometrics, was supervised by Prof. dr. Franz Palm and dr. Jean-Pierre Urbain at Maastricht University.

Since January 2001, he has been visiting the Department of Quantitative Economics Maastricht University and since September 2001, he is visiting the CORE in Louvain-la-Neuve.