Lars Forsberg

On the Normal Inverse Gaussian Distributionin Modeling Volatility in the Financial Markets

UPPSALA
UNIVERSITET

ACTA UNIVERSITATIS UPSALIENSIS
Studia Statistica Upsaliensia
5

Lars Forsberg

On the Normal Inverse Gaussian Distribution in Modeling Volatility in the Financial Markets

Dissertation for the Doctoral Degree of Philosophy in Statistics presented at Uppsala University 2002

Abstract

Forsberg, Lars, On the Normal Inverse Gaussian Distribution in Modeling Volatility in the Financial Markets. Acta Univ. Ups., *Studia Statistica Upsaliensia* 5. Uppsala. ISBN 91-554-5298-1

We discuss the Normal inverse Gaussian (NIG) distribution in modeling volatility in the financial markets. Refining the work of Barndorff-Nielsen (1997) and Andersson (2001), we introduce a new parameterization of the NIG distribution to build the GARCH(p,q)-NIG model. This new parameterization allows the model to be a strong GARCH in the sense of Drost and Nijman (1993). It also allows us to standardize the observed returns to be i.i.d., so that we can use standard inference methods when we evaluate the fit of the model.

We use the realized volatility (RV), calculated from intraday data, to standardize the returns of the ECU/USD foreign exchange rate. We show that normality cannot be rejected for the RV-standardized returns, i.e., the Mixture-of-Distributions Hypothesis (MDH) of Clark (1973) holds. We build a link between the conditional RV and the conditional variance. This link allows us to use the conditional RV as a proxy for the conditional variance. We give an empirical justification of the GARCH-NIG model using this approximation.

In addition, we introduce a new General GARCH(p,q)-NIG model. This model has as special cases the Threshold-GARCH(p,q)-NIG model to model the leverage effect, the Absolute Value GARCH(p,q)-NIG model, to model conditional standard deviation, and the Threshold Absolute Value GARCH(p,q)-NIG model to model asymmetry in the conditional standard deviation. The properties of the maximum likelihood estimates of the parameters of the models are investigated in a simulation study.

Keywords: Volatility modeling, inverse Gaussian, normal inverse Gaussian, realized volatility, GARCH.

Lars Forsberg, Uppsala University, Department of Information Science, Division of Statistics, Box 513, SE-751 20 Uppsala, Sweden

©Lars Forsberg 2002

ISSN 1104-1560
ISBN 91-554-5298-1

Cover design: Jan-Olof Werkmäster
Printed in Sweden by Elanders Gotab, Stockholm 2002
Distributor: Uppsala University Library

DEDICATION

To Anna, my love,
and Lovisa, my pride and joy.

Strive for excellence in every step of the way.
George Tauchen

Contents

1 Review of volatility models **1**

1.1 Introduction . . . 1
1.2 Background to volatility modeling . . . 3
1.3 The stochastic volatility model . . . 4
1.4 The GARCH(p,q) model . . . 5
1.5 The NIGSV(p,q) model . . . 6
1.6 Continuous time models and volatility . . . 8
1.7 Outline of the thesis . . . 10

2 A new parameterization of the NIG **11**

2.1 Standardization of the NIGSV(1,1) . . . 11
2.2 A new scale invariant parameterization of NIG . . . 13
2.2.1 Scaling properties of the new parameterization . . . 14
2.3 The GARCH(p,q)-NIG model . . . 15

3 Test of the Mixing Distribution Hypothesis **17**

3.1 Introduction . . . 17
3.2 The Mixture-of-Distributions Hypothesis . . . 18
3.3 Data sources and realized volatility . . . 19
3.4 Raw Returns . . . 20
3.5 RV-standardized returns . . . 21
3.6 Conclusions . . . 22
3.7 Futher work . . . 22
3.8 Tables . . . 23
3.9 Figures . . . 28

4 Motivating the GARCH(p,q)-NIG **41**

4.1 The link between realized volatility and conditional variance . . 42
4.2 The inverse gamma and the inverse Gaussian distributions . . . 44
4.3 Unconditional distributions . . . 47
4.4 Conditional distributions . . . 48
4.4.1 Distributions conditional on lagged realized volatilities . 48

4.4.2 Distributions conditional on lagged squared returns . . . 51
4.4.3 Out-of-sample Euro predictions . . . 52
4.5 Conclusions . . . 53
4.6 Further work . . . 53
4.7 Tables . . . 54
4.8 Figures . . . 62

5 Temporal aggregation of RV and IG **73**
5.1 Fitting IG to realized volatility . . . 74
5.2 Fitting IG to standardized RV . . . 74
5.2.1 Fitting IG to standardized RV results . . . 76
5.2.2 Moment aggregation of the realized volatility . . . 76
5.2.3 Analytical aggregation of moments . . . 79
5.3 Conclusions . . . 79
5.4 Tables . . . 81
5.5 Figures . . . 87

6 The General GARCH-NIG model **92**
6.1 Asymmetry . . . 93
6.2 Modeling the conditional standard deviation . . . 94
6.3 A general GARCH-NIG(p,q) model . . . 94
6.3.1 T-GARCH(p,q)-NIG . . . 95
6.3.2 AV-GARCH(p,q)-NIG . . . 96
6.3.3 TAV-GARCH(p,q)-NIG . . . 96
6.4 Estimation of the models . . . 97
6.5 Concluding remarks . . . 98
6.6 Further work . . . 98

7 Simulation study **100**
7.1 Simulation setup . . . 100
7.2 Evaluation measures . . . 102
7.3 Results of the simulations . . . 103
7.3.1 Results for GARCH-NIG . . . 103
7.3.2 Results for T-GARCH-NIG . . . 104
7.3.3 Results for AV-GARCH-NIG . . . 104
7.3.4 Results for TAV-GARCH-NIG . . . 105
7.4 Concluding remarks . . . 106
7.5 Tables . . . 107

Remarks and further work **145**

References **148**

A Moments of the models **154**
A.1 Moment structure of the GARCH(1,1)-NIG 154
A.2 Moment structure of the T-GARCH(1,1)-NIG 155
A.3 Moment structure of the AV-GARCH(1,1)-NIG 157
A.4 Moment structure of the TAV-GARCH(1,1)-NIG 158

B Absolute moments of NIG **160**

C Gradients and Hessians of the models **162**
C.1 GARCH(p,q)-NIG 164
C.2 T-GARCH(p,q)-NIG 165
C.3 AV-GARCH(p,q)-NIG 166
C.4 TAV-GARCH(p,q)-NIG 167

List of Tables

3.1a Descriptives of ECU/USD returns, daily, weekly 23
3.1b Descriptives of ECU/USD returns, bi-weekly, monthly 24
3.2a Q-statistics of ECU/USD returns, daily, weekly 25
3.2b Q-statistics of ECU/USD returns, bi-weekly, monhtly 26
3.3 Descriptives of ECU/USD realized volatility 27
4.1 Distributions fitted to ECU/USD realized volatility 54
4.2 Distributions fitted to ECU/USD returns 55
4.3 RV-GARCH models fitted to ECU/USD RV (given F_{t-1}) 56
4.4 RV-GARCH models fitted to ECU/USD returns (given F_{t-1}) 57
4.5 GARCH models fitted to ECU/USD RV (given I_{t-1}) 58
4.6 GARCH models fitted to ECU/USD returns (given I_{t-1}) 59
4.7 Out-of-sample information criteria, unconditional distributions 60
4.8 Out-of-sample information criteria, GARCH models 61
5.1 ARMA(1,1)-IG estimates 81
5.2 Q-statistics of raw and standardized RV 82
5.3 Descriptives: RV standardized by ARMA(1,1)-IG 83
5.4 IG fitted to standardized RV 83
5.5a Temporal aggregation: mean 84
5.5b Temporal aggregation: variance 85
5.5c Temporal aggregation: skewness 85
5.5d Temporal aggregation: kurtosis 86
7.1a Parameters sets for the GARCH(1,1)-NIG 107
7.1b Parameters sets for the T-GARCH(1,1)-NIG 108
7.2a Parameters sets for the AV-GARCH(1,1)-NIG 109
7.2b Parameters sets for the TAV-GARCH(1,1)-NIG. 110
7.3a Bias and std err. of est. of GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 111
7.3b Bias and std err. of est. of GARCH(1,1)-NIG (ρ_1, π_1) 112
7.4a MAPE of est. of GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 113
7.4b MAPE of est. of GARCH(1,1)-NIG (ρ_1, π_1) 114
7.5a CI coverage of est. of GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 115
7.5b CI coverage of est. of GARCH(1,1)-NIG (ρ_1, π_1). 116
7.6a JB test of normality of est. of GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 117
7.6b JB test of normality of est. of GARCH(1,1)-NIG (ρ_1, π_1) 118
7.7a Bias and std err. of est. of T-GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 119
7.7b Bias and std err. of est. of T-GARCH(1,1)-NIG $(\rho_1, \omega_1, \pi_1)$ 120
7.8a MAPE of est. of T-GARCH(1,1)-NIG $(\overline{\alpha}, \rho_0)$ 121
7.8b MAPE of est. of T-GARCH(1,1)-NIG $(\rho_1, \omega_1, \pi_1)$ 122

7.9a CI coverage of est. of T-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 123
7.9b CIcoverage of est. of T-GARCH(1,1)-NIG (ρ_1, ω_1, π_1) 124
7.10a JB test of normality of est. of T-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 125
7.10b JB test of normality of est. of T-GARCH(1,1)-NIG (ρ_1, π_1) 126
7.10c JB test of normality of est. of T-GARCH(1,1)-NIG (π_1) 127
7.11a Bias and std err. of est. of AV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 128
7.11b Bias and std err. of est. of AV-GARCH(1,1)-NIG (ρ_1, π_1) 129
7.12a MAPE of est. of the AV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 130
7.12b MAPE of est. of the AV-GARCH(1,1)-NIG (ρ_1, π_1) 131
7.13a CI coverage of est. of AV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 132
7.13b CI coverage of est. of AV-GARCH(1,1)-NIG (ρ_1, π_1). 133
7.14a JB test of normality of est. of AV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 134
7.14b JB test of normality of est. of AV-GARCH(1,1)-NIG (ρ_1, π_1) 135
7.15a Bias and std err. of est. of the TAV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$).... 136
7.15b Bias and std err. of est. of TAV-GARCH(1,1)-NIG (ρ_1, ω_1, π_1)... 137
7.16a MAPE of est. of the TAV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$).............. 138
7.16b MAPE of est. of the TAV-GARCH(1,1)-NIG (ρ_1, ω_1, π_1) 139
7.17a CI coverage of est. of TAV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 140
7.17b CI coverage of est. of TAV-GARCH(1,1)-NIG (ρ_1, ω_1, π_1) 141
7.18a JB test of normality of est. of TAV-GARCH(1,1)-NIG ($\overline{\alpha}, \rho_0$) 142
7.18b JB test of normality of est. of TAV-GARCH(1,1)-NIG (ρ_1, ω_1) ... 143
7.18c JB test of normality of est. of TAV-GARCH(1,1)-NIG (π_1) 144

List of Figures

3.1a ECU/USD Returns, daily, weekly 28
3.1b ECU/USDReturns, bi-weekly, monthly 29
3.2a ECU/USD Relized volatility, daily, weekly 30
3.2b ECU/USD Relized volatility, bi-weekly, monthly 31
3.3a ECU/USD Returns and fitted normal, daily, weekly 32
3.3b ECU/USD Returns and fitted normal, bi-weekly, monthly 33
3.4a QQ-plot: ECU/USD Returns and normal, daily 34
3.4b QQ-plot: ECU/USD Returns and normal, weekly 35
3.4c QQ-plot: ECU/USD Returns and normal, bi-weekly 36
3.4d QQ-plot: ECU/USD Returns and normal, monthly 37
3.5 SACF of ECU/USD Returns 38
3.6a RV-standardized ECU/USD returns, daily, weekly 39
3.6b RV-standardized ECU/USD returns, bi-weekly, montly 40
4.1 QQ-plot: IG on RV 62
4.2 QQ-plot: IGamma on RV 62
4.3 QQ-plot: NIG on returns 63
4.4 QQ-plot: Student's t on returns 63
4.5 QQ-plot: normal on returns 64
4.6 QQ-plot: conditional IG on RV 64
4.7 QQ-plot: conditional IGamma on RV 65
4.8 QQ-plot: returns standardized by RV-GARCH(1,1)-NIG 65
4.9 QQ-plot: returns standardized by RV-GARCH(1,1)-N 66
4.10 QQ-plot: returns standardized by RV-GARCH(1,1)-t 67
4.11 QQ-plot: RV standardized by GARCH(1,1)-IG 68
4.12 QQ-plot: RV standardized by GARCH(1,1)-IGamma 69
4.13 QQ-plot: returns standardized by GARCH(1,1)-NIG 70
4.14 QQ-plot: returns standardized by GARCH(1,1)-n 71
4.15 QQ-plot: returns standardized by GARCH(1,1)-t 72
5.1 ACF of daily RV. 87
5.2 ACF of standardized RV 87
5.3a Standardized RV and fitted IG, daily, weekly 88
5.3b Standardized RV and fitted IG, bi-weekly, monthly 89
5.4a QQ-plot: RV standardized by ARMA(1,1)-IG against IG 90
5.4b QQ-plot: RV standardized by ARMA(1,1)-IG against IG, cont. 91

Acknowledgements

There are many people I would like to thank for helping me throughout my journey through the graduate school. First, would like to acknowledge my gratitude to my first supervisor, the late Reinhold Bergström. Many thanks to my second supervisor, Anders Ågren for guidance and interesting discussions. I would also like to thank Anders Christofferson for all the help throughout the years.

Furthermore, I would like to thank Johan Lyhagen for introducing me to the field of time series analysis back in 1998, for his unselfish help, for all the clarifying and stimulating discussions and for his support throughout my work on this thesis.

I have been highly influenced by Jonas Andersson, and I would like to say a big thanks to him for his comments and suggestions, and for many discussions from which I have learned much. Thanks to Tomas Petterson and Inger Persson, my roommates during the early years, for many interesting discussions. An additional thanks to Tomas for software support. I would like to thank Anders Eriksson for his inspiring manners, and for numerous constructive and creative discussions. Half a row of thanks to Rolf Larsson.

I would like to thank Professor George Tauchen for inviting me to the Department of Economics, Duke University, USA. During my time at Duke my work got a new boost that is still burning strong. Thanks to the Empirical Finance Lunch Group for many suggestions and comments on my work. I would also like to thank Professor Tim Bollerslev for inspiring discussions and many suggestions that have improved this thesis. A special thanks to Professor Eric Ghysels for many insightful, clarifying and stimulating discussions. His guidance through the literature has been of invaluable help.

I am thankful to Petrus Sundvall, who convinced me to go to the university in the first place, some ten years ago.

A special thanks goes to Robert Ekerlin for being a good role model for working hard, and for teaching me, solely by his way of being, that giving up is never an option.

Needless to say, I would like to thank my wife Anna, for her solid support and patience with me, and with my work on this thesis. It's not always easy to write a thesis, and I can imaging that it's not always easy to live with someone doing it. And to our beloved daughter Lovisa, for just being there, and making me realize that there is more to life than statistics...

Sist och mest, vill jag tacka mina föräldrar, Irène och Svante för att de alltid har funnits där med all sin kärlek, för att de i alla väder har stöttat mig, och i evinnerliga tider uppmuntrat mig att göra mina läxor.

Financial support from the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) and from Jan Wallanders and Tom Hedelius Foundation is gratefully acknowledged.

Chapter 1

Review of volatility models

1.1 Introduction

The modeling of variances of returns of financial assets is crucial for the financial practician. The uncertainty of returns measured as variances and covariances of the returns is important in derivative pricing, hedging and risk management.

Returns from financial markets are characterized by two stylized facts, non-normality and volatility clustering. Returns are not normally distributed, instead the empirical distribution of returns is leptokurtic, that is, it is more peaked and has fatter tails than the normal distribution. Volatility clustering means that small changes in price tend to be followed by small changes, and that large price changes tend to be followed by large price changes. Expressed differently, one could say that the squared returns are autocorrelated. This has been known for some time, see e.g. Mandelbrot (1963) and Fama (1965).

The seminal work by Engle (1982), where he introduced the Auto Regressive Conditional Heteroscedasticity (ARCH) model, and Bollerslev (1986), who introduced the Generalized Auto Regressive Conditional Heteroscedasticity (GARCH) model, triggered one of the most active and fruitful areas of research in econometrics over the past two decades. The success of the ARCH/GARCH class of models at capturing volatility clustering in financial markets is well documented (see, for example, Bollerslev, Chou, and Kroner, 1992). At the same time, the inability of the ARCH/GARCH models coupled with the auxiliary assumption of conditionally normally distributed errors to fully account for the mass in the tails of the distributions of, say, daily returns, is generally well recognized. Indeed, several alternative error distributions were proposed in the early ARCH literature to better account for the deviations from normality in the conditional distributions of the returns. For example, the t-distribution of Bollerslev (1987), the General Error Distribution (GED) of Nelson (1991), and more recently, the normal inverse Gaussian (NIG) distribution of Barndorff-Nielsen (1997), Andersson (2001) and Jensen and Lunde (2001). Meanwhile, the justification behind these alternative error distributions has been almost exclusively empirical and pragmatic in nature.

In this thesis, building on the Mixture-of-Distributions-Hypothesis (MDH) (Clark, 1973) along with the recent idea of so-called Realized Volatilities (RV) (Andersen, Bollerslev, Diebold and Labys 2001, 2002, and Barndorff-Nielsen and Shephard, 2001a,b, 2002a), we provide a sound empirical foundation for the distributional assumptions behind the GARCH-NIG model. Consistent with the absence of arbitrage and a time-changed Brownian motion (see, for example, Ane and Geman, 2000, and Andersen, Bollerslev, Diebold, 2002), the MDH postulates that the distribution of returns is normal, but with a stochastic (latent) variance. In the original formulation in Clark (1973) the variance is assumed to be i.i.d. lognormally distributed, resulting in a lognormal-normal mixture distributions for the returns. Numerous theoretical extensions and empirical investigations of these ideas involving various proxies for the mixing variable have been conducted in the literature (important early contributions include Epps and Epps, 1976; Taylor, 1982; Tauchen and Pitts, 1983). Importantly, to explicitly account for the volatility clustering effect, Taylor (1982, 1986) proposed an extension of the MDH setup by assuming that the (latent) logarithmic variances follow a Gaussian autoregression, resulting in the lognormal Stochastic Volatility (SV) model; see also Andersen (1996). Since the joint distribution of the returns in the SV model is not known in a closed form, estimation and inference for these types of models are considerably more complicated than for the ARCH/GARCH class of models (see, e.g., Shephard, 1996), which we will consider in the next section.

Barndorff-Nielsen (1997) and Andersson (2001) assume that the conditional variance is inverse Gaussian (IG). This assumption implies that the returns, conditional of an information set, are normal inverse Gaussian (NIG). That is, the joint distribution of the returns is known in a closed form, and maximum likelihood estimation is straightforward. Andersson (2001) denotes the model the "Normal inverse Gaussian Stochastic Volatility" (NIGSV) model. In this thesis, we give further empirical support for this model. We will use a slightly different parameterization of the NIG distribution, which enable one to consider the model to be a GARCH model, hence we refer to this model as the GARCH-NIG model.

1.2 Background to volatility modeling

Here we provide a background to the statistical modeling of financial data. We highlight the statistical properties of the data and discuss different explanations and ways to model these stylized facts.

From a statistical perspective, when one considers how the daily returns are constructed, the non-normality of the returns can be quite mysterious. The daily price changes are made up of many small intraday price changes. Let

$$x_i = \ln P_i - \ln P_{i-1},$$

where P_i is ith intraday price and x_i is the intraday log price change. The daily return can then be written as the sum of the intraday returns, that is

$$r_t = \Sigma_{i=1}^{m} x_i,$$

where m is the number of price changes within day t. According to financial theory all known information about the security is incorporated in the price. When new information arrives to the market place, this causes the market participants to re-evaluate the security and the price adjusts as trading takes place. In theory, every new piece of information triggers a trade, and therefore a price change. This means that the daily price is made up of, say, m trades (assuming that the information flow is constant over time, that is that we have the same number of trades each day).

Now, assuming that these intraday price changes are independent and identically distributed, the Central Limit Theorem (CLT) says that the daily return should be normally distributed. However, there is overwhelming evidence that the returns are NOT normally distributed. So, which assumption of the CLT is violated?

Mandelbrot (1963) argued that the failure of the CLT is due to the fact that the intraday changes are independent but they do not have a finite variance. Given this assumption, by utilizing a generalized CLT, one can show that the daily price changes follow a stable Paretian law.

Another explanation for the non-normality of the returns was presented by Clark (1973), who introduced the Mixture-of-Distributions Hypothesis (MDH). Consistent with the absence of arbitrage and a time-changed Brownian motion (see, for example, Ane and Geman, 2000, and Andersen, Bollerslev, and Diebold, 2002), the MDH postulates that the distribution of returns is normal but with a stochastic (latent) variance.

$$r_t \sim N\left(0, \sigma_t^2\right),$$

where σ_t^2 is a strictly positive random variable. In the original formulation in Clark (1973) the variance is assumed to be i.i.d. lognormally distributed, resulting in a lognormal-normal mixture distribution for the returns.

$$f(r_t) = \int_0^{\infty} f_{\text{Normal}}\left(r_t|\sigma_t^2\right) * g_{\text{Mixing}}\left(\sigma_t^2\right) d\sigma_t^2. \tag{1.1}$$

It can be shown that the resulting distribution has fatter tails than the normal distribution. To find the joint distribution of the returns, we need to integrate out the unobserved variance. For the lognormal assumption of the mixing variable, the integral in (1.1) is not known in a closed form. Numerous theoretical extensions and empirical investigations of these ideas involving various proxies for the mixing variable have been conducted in the literature (early contributions include, Epps and Epps, 1976; Taylor, 1982; Tauchen and Pitts, 1983)

1.3 The stochastic volatility model

The MDH of Clark (1973) might explain the non normality of the returns, but it does not explain the volatility clustering, or ARCH effects in the returns. To explicitly account for the volatility clustering effect Taylor (1982, 1986) proposed an extension of the MDH setup by making the (latent) logarithmic variances follow a Gaussian autoregression, resulting in the lognormal Stochastic Volatility (SV) model; see also Andersen (1996). For an excellent introduction to SV models, see Ghysels et al (1996). The SV model can be written

$$r_t|\sigma_t^2 \sim N(0, \sigma_t^2),$$

where

$$\begin{aligned} \sigma_t^2 &= \sigma^2 \exp(h_t), \\ h_t &= \gamma h_{t-1} + \eta_t, \\ \eta_t &\sim N(0, \sigma_\eta^2). \end{aligned}$$

We can rewrite the model to make it (more) apparent that the conditional variance is assumed to be lognormally distributed

$$\sigma_t^2|\Psi_{t-1} \sim LogN(\ln \sigma^2 + \gamma h_{t-1}, \sigma_\eta^2),$$

where Ψ_t denotes the information up to and including time t. The density of the returns is given by

$$r_t \sim \int_0^{\infty} f_{\text{Normal}}\left(r_t|\sigma_t^2\right) * g_{\text{Lognormal}}\left(\sigma_t^2\right) d\sigma_t^2. \tag{1.2}$$

Since the density of the returns in the SV model in (1.2) is not known in a closed form, estimation and inference are considerably more complicated for these types of models than for the ARCH/GARCH class of models (see, e.g., Shephard, 1996).

1.4 The GARCH(p,q) model

Another branch of volatility modeling is the ARCH/GARCH-literature, which started with Engle (1982) and Bollerslev (1986). In the GARCH(p,q) model the conditional variance is a deterministic function of lagged squared observations and lagged conditional variances. The GARCH(p,q) model can be written as

$$r_t = E\left(r_t | I_{t-1}\right) + \sigma_t \varepsilon_t,$$

where I_{t-1} denotes the information set containing all information up to time t and $E\left(r_t | I_{t-1}\right)$ is expected value of the return given the information set I_{t-1}, and

$$\varepsilon_t \sim i.i.d.\left(0,1\right),$$

and

$$\sigma_t^2 = \rho_0 + \Sigma_{i=1}^{q} \rho_i r_{t-i}^2 + \Sigma_{j=1}^{p} \pi_j \sigma_{t-j}^2, \tag{1.3}$$

where p is the number of lagged conditional variances, q is the number of lagged squared returns entering the variance equation, where $\rho_0 > 0$, $\rho_i \geq 0$ $i = 1, ..., q$ and $\pi_j \geq 0$ $j = 1, ..., p$.

Bollerslev (1986) assumes (conditional) normality of the returns, in which case the model can be written

$$r_t | I_{t-1} \sim N\left(\mu_t, \sigma_t^2\right),$$

where $\mu_t = E\left(r_t | I_{t-1}\right)$ and where σ_t^2 is defined in (1.3).

The GARCH(p,q) model explicitly models the volatility clustering, and one can show that the unconditional distribution of the returns has fatter tails than the normal distribution. Still, the normal distribution is not enough to fully account for the fat tails of the return distribution. For this reason fat-tailed distributions have been proposed in the literature, such as the t-distribution by Bollerslev (1987) and the General Error Distribution (GED) by Nelson (1991). A large number of GARCH-type models have been proposed in the literature, for a survey see Bollerslev et al (1992).

1.5 The NIGSV(p,q) model

When dealing with models where the conditional variance is random, such as the SVAR model of Taylor, we have the problem that the likelihood in (1.2) is not known in a closed form. Therefore, it makes sense to look for other distributions for the variance. This has been done by Barndorff-Nielsen (1997), and his model was generalized by Andersson (2001). They use the inverse Gaussian (IG) distribution as a mixing distribution.[1] The density of the IG distribution is given by

$$f(z;\delta,\alpha,\beta) = (2\pi)^{-1/2}\delta\exp(\delta\gamma)z^{-3/2}\exp(-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)),$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$. The first two moments are

$$E(z) = \frac{\delta}{\gamma},$$

and

$$V(z) = \frac{\delta}{\gamma^3}.$$

Note that the parameter δ is *proportional to* the mean of the distribution.[2]

If we have a normally distributed variable with the variance drawn from the IG distribution,

$$x|z \sim N(\mu, z),$$

where

$$z \sim IG(\delta, \sqrt{\alpha^2 - \beta^2}).$$

Then the distribution of the return is NIG

$$f\left(x\right) = \int f\left(x|z\right) g\left(z\right) dz \sim NIG\left(\alpha,\beta,\mu,\delta\right).$$

The $NIG\left(\alpha,\beta,\mu,\delta\right)$ density is given by

$$g(x;\alpha,\beta,\mu,\delta) = a\left(\alpha,\beta,\mu,\delta\right) q\left(\frac{x-\mu}{\delta}\right)^{-1} * K_1\left[\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right]\exp\left(\beta x\right), \tag{1.4}$$

[1] The name inverse Gaussian is due to the fact that the cumulant generating funtion of the IG density, is the inverse of the cumulant generating function of the Gaussian distribution.

[2] The inverse Gaussian distribution can be derived as the waiting time for a Brownian motion with drift α, to hit a barrier δ, see Seshadri (1993).

Alternatively, it can be derived as the number of links an internet "surfer" uses before finding the "right page", given that the surfing follows a Gaussian random walk, see Huberman, Pirolli, Pitkow and Lukose (1998).

where K_1 is the modified Bessel function of third order and index 1, that is $K_1(x) = \int_0^\infty \exp(-x\cosh(t))\cosh(t)dt$ and the functions $a(\alpha,\beta,\mu,\delta) = \pi^{-1}\alpha\exp\left[\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right]$ and $q(x) = \sqrt{1+x^2}$. Restrictions for the parameters are $0 \leq |\beta| \leq \alpha$, $\mu \in R$ and $\delta > 0$. The first four central moments are given by

$$\begin{aligned}
\mu_1 &= \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}, \\
\mu_2 &= \frac{\delta\alpha^2}{\left(\alpha^2-\beta^2\right)^{3/2}}, \\
\mu_3 &= \frac{3\delta\beta\alpha^2}{\left(\alpha^2-\beta^2\right)^{5/2}},
\end{aligned}$$

and

$$\mu_4 = 3\delta\alpha^2\frac{\alpha^2+4\beta^2+\delta\alpha^2\sqrt{\alpha^2-\beta^2}}{\left(\alpha^2-\beta^2\right)^{7/2}}.$$

The parameters can be interpreted as follows: α and β are shape parameters with β determining the skewness of the distribution and, with $\beta = 0$, α determining the degree of non-normality. The parameter δ is a scale parameter and μ is the location parameter, if $\beta = 0$, μ denotes the mean of the distribution.

Barndorff-Nielsen (1997) used the normal inverse Gaussian distribution to construct a volatility model of the mixing distribution type; his formulation was generalized by Andersson (2001). We present the model by Andersson (2001), where

$$\beta = \mu = 0,$$

which means that the resulting NIG distribution is symmetric about zero, however, it is straightforward to model the conditional first moment. It is also possible to let β be non zero, and include modeling of the skewness of the distribution. The observed variable r_t is, given the variance z_t, normally distributed

$$(r_t|z_t) \sim N(0, z_t),$$

where the variance z_t is inverse Gaussian given the information set $I_{t-1} = (\delta_{-p+1}, ..., \delta_0, r_{-q+1}, ..., r_{t-1})$:

$$(z_t|I_{t-1}) \sim IG(\delta_t, \alpha).$$

Conditional on the information set, the observed variable is now normal inverse Gaussian

$$(r_t|I_{t-1}) \sim NIG(\alpha, 0, 0, \delta_t),$$

and the conditional variance is given by

$$V\left(r_t | I_{t-1}\right) = \frac{\delta_t}{\alpha}.$$

Andersson (2001) makes the parameter δ_t time varying according to

$$\delta_t = \rho_0 + \Sigma_{i=1}^{q} \rho_i r_{t-i}^2 + \Sigma_{j=1}^{p} \pi_j \delta_{t-j}. \tag{1.5}$$

In the NIGSV(p,q) model we do not model the conditional variance directly, but the parameter δ_t, which is *proportional to* the conditional variance.

We can also write (1.5) using slightly different notation.[3] Let

$$B\left(L\right) = 1 - \Sigma_{j=1}^{p} \pi_j L^j, \tag{1.6}$$

and

$$A\left(L\right) = 1 - \Sigma_{i=1}^{q} \rho_i L^i, \tag{1.7}$$

where L^i denotes the lag operator, that is, $x_t L^i = x_{t-i}$ and $y_t = r_t / \sqrt{(\delta_t/\alpha)}$ is the sequence of errors. Then we can define the sequence of $\{\delta_t\}$ to be the stationary solution to

$$B\left(L\right)\delta_t = \rho_0 + \left(A\left(L\right) - 1\right) y_t^2. \tag{1.8}$$

We do not have a latent factor in the NIGSV(p,q) model as we had in the SVAR model. Instead one of the parameters is made time varying. In Andersson (2001), the parameter δ_t is made time varying.[4] In contrast to the SVAR of Taylor (1986), we know the joint distribution of the observed variable in closed form, which makes maximum likelihood estimation/ inference straightforward.

1.6 Continuous time models and volatility

In order to fully understand the ideas presented later in this thesis, we need some results from continuous time finance. Let us assume that the log price follows an univariate diffusion process with no mean dynamics

$$dp\left(t\right) = \sigma\left(t\right) dW\left(t\right), \tag{1.9}$$

[3] We will use this notation later, when dealing with the scaling properties of the NIGSV(p,q) model.

[4] Because the model does not have a latent factor, some authors claim that the model is not a stochastic volatility model. However, the model can be written as a product of two stochastic variables, i.e.,

$$x_t = \sqrt{\nu_t} \varepsilon_t,$$

where $\nu_t \sim IG\left(\delta_t, \alpha\right)$ and $\varepsilon_t \sim N\left(0, 1\right)$. This is in the same spirit as the (log normal) stochastic volatility model of Taylor (1986).

where $p(t)$ denotes the log price at time t, W is a standard Brownian motion, and $\sigma(t)$ is the instantaneous volatility or the spot volatility. The return at time t is defined as

$$r_t \equiv p(t) - p(t-1) = \int_{t-1}^{t} \sigma(s)\, dW(s), \tag{1.10}$$

where r_t is the continuously compounded return at time t. Andersen, Bollerslev, Diebold and Labys (2001) (ABDL 2001) used the quadratic variation (QV) of the process as a volatility measure. The quadratic variation is defined as

$$QV_t = \int_{t-1}^{t} \sigma^2(s)\, ds. \tag{1.11}$$

The expression $\int_{t-1}^{t} \sigma^2(s)\, ds$ also defines the so-called *integrated volatility* (IV_t). An interesting result in (ABDL 2001) is that the conditional expectation of the quadratic variation is the conditional variance of the returns. That is

$$E\left(QV_t|\psi_{t-1}\right) = V\left(r_t|\psi_{t-1}\right), \tag{1.12}$$

where ψ_t is all the information up to time t. We will make use of this result later in this thesis when linking the realized volatility to the conditional variance. The quadratic variation is a theoretical concept, of course unobservable in practice. To estimate the quadratic variation at time t, we use an estimate referred to as the *realized volatility* (RV_t). The RV_t is defined as

$$RV_t = \Sigma_{i=1}^{mh} r_{(m)}^2 (t - h + i/m), \tag{1.13}$$

where $r_{(m)}$ is the intraday return, sampled m times a day and h is the frequency, where $h = 1$ is daily, $h = 5$ is weekly and so on. This idea is has also been used in Schwert (1989), Hsieh (1991), and elsewhere. A formal justification of the realized volatility is given in (ABDL 2001). The realized volatility is a consistent estimate of the quadratic variation, that is

$$p \lim_{m\to\infty} RV_{t,h} = QV_{t,h}. \tag{1.14}$$

Given only data on a daily basis or with an even lower frequency, the standard way to estimate the conditional variance has been to use the square of the returns as an estimate of the volatility. Andersen and Bollerslev (1998) showed that the squared return is a very noisy estimate of the variance, so they preferred to use the idea of realized volatility (RV).

As can be seen from (1.13), by definition, the RV over a week is simply the sum of the RV for 5 days. That is, we aggregate the realized volatility in the same way as we aggregate compound returns.

When we have access to intraday data, we can model the RV directly, instead of resorting to modeling the squares of the, say, daily returns. Some steps have been taken in this direction. ABDL (2001) analyzed the RV of three FX series, and they developed a multivariate model for these series. The basic idea in their paper was to assume that the RV is lognormal, they take the natural log of the RV and model the serial dependence using an ARMA(p,q) model assuming the errors to be normal.

1.7 Outline of the thesis

The rest of the thesis is organized as follows. In Chapter 2 we introduce and motivate a new scale invariant parameterization of the normal inverse Gaussian distribution. In Chapter 3 we give further evidence that the MDH holds. Using an intraday dataset ECU/USD 1989 - 1998, we construct realized volatility and standardize the returns thereby showing that we cannot reject the null of normality.

In Chapter 4, using results from continuous time finance, we build a link between realized volatility and conditional variance. We use the realized volatility of the ECU/USD 1989 - 1998 dataset to show that the inverse Gaussian distribution gives a good fit to the conditional variance, giving empirical support to the GARCH(p,q)-NIG model.

In Chapter 5 we use the temporal aggregation properties of the realized volatility, and the convolution formulas of the inverse Gaussian to give further support to the hypothesis that the conditional realized volatility is well described by the inverse Gaussian distribution. Again, this gives direct support to the GARCH(p,q)-NIG model.

In Chapter 6 we introduce a new General GARCH(p,q)-NIG model. As special cases of this models we derive four GARCH-NIG models From this General GARCH(p,q) we derive the three GARCH-NIG models as special cases.[5] The special cases are: the Threshold-GARCH(p,q)-NIG, which is an asymmetric model for the conditional variance, the Absolute Value GARCH(p,q)-NIG model, which is a symmetric model for the conditional standard deviation, and the Threshold Absolute Value GARCH(p,q)-NIG model, which is an asymmetric model for the conditional standard deviation.

Chapter 7 conducts a maximum likelihood study for the four models in Chapter 6, we focus on a comparison of the small sample performance of the maximum likelihood (ML) estimator using numerical and analytical gradients.

[5] In a concurrent and independent work, Jensen and Lunde (2001) proposed a more general model "GARCH-NIG" model, which they refer to as the NIG-S&ARCH model, which is the A-PARCH model of Ding, Granger and Engle (1993) used with the NIG distribution.

Chapter 2

A new parameterization of the NIG

Here we propose and motivate a "scale invariant" parameterization of the normal inverse Gaussian distribution. Barndorff-Nielsen (1997) has also proposed a scale invariant parameterization of the normal inverse Gaussian distribution, however, our parameterization uses only one parameter for the variance, which is more intuitive in the context of conditional variance modeling. This will lead us to a new parameterization of the NIGSV(p,q) model of Andersson (2001); we refer to the new formulation of the model as the GARCH(p,q)-NIG model.

Using this parameterization, we can write the model not only as a SV model, but also as a (strong) GARCH model with a NIG error distribution.[1] We highlight some differences between the two parameterizations, and the implications for the modeling of the conditional variance.

2.1 Standardization of the NIGSV(1,1)

When we model time dependence of the conditional variance in real data, we might want to standardize the observed returns, i.e., divide the observed data by the (estimated) conditional standard deviation to get the standardized data. In doing so, we can use standard diagnostics to check whether the model gives a good description of the data. We might want to see if the standardized returns are normal, or if there are any serial correlations in the squared standardized returns. In a practical situation, if the model is correct we should have no dependencies left in the standardized data. Let r_t be the

[1] By strong GARCH, we mean a strong GARCH in the sense of Drost and Nijman (1993).

daily return at time t, and, as in the NIGSV(p,q) model, let δ_t/α be the conditional variance at time t, then we can standardize the return

$$r_t^* = \frac{r_t}{\sqrt{\frac{\delta_t}{\alpha}}},$$

where r_t^* is the standardized return at time t. To see how the standardized return r_t^* is distributed in the NIGSV(p,q) framework, we need to know the scaling properties of the NIG distribution. The scaling properties of the parameterization of the NIG distribution used in the NIGSV(p,q) model are as follows. Let $x \sim NIG(\alpha, 0, 0, \delta)$, then

$$cx \sim NIG(\frac{\alpha}{c}, 0, 0, c\delta). \tag{2.1}$$

By using (2.1), the standardized returns from the NIGSV(p,q) model are distributed according to

$$r_t^* \sim NIG(\sqrt{\alpha\delta_t}, 0, 0, \sqrt{\alpha\delta_t}).$$

We note that the parameters of the standardized returns are still time varying. The variance of the standardized return is

$$V(r_t^*|I_{t-1}) = \frac{\sqrt{\alpha\delta_t}}{\sqrt{\alpha\delta_t}} = 1,$$

so the conditional variance is constant, but higher moments are time varying. For instance, the kurtosis of the standardized return is

$$K_4\left(r_t^*\right)|I_{t-1} = 3 + \frac{3}{\alpha\delta_t},$$

that is, we have a time varying conditional kurtosis in the standardized returns. The reason for this is that when setting a time varying structure on the parameter δ_t in (1.5), we not only model the conditional *variance*, but since δ_t determines higher moments as well, we model the conditional *distribution.* This is basically what one wants to do when modeling financial data. Modeling the conditional variance as in the GARCH models, is just a convenient simplification of reality since one might suspect that higher order moments are time varying as well. The drawback of this parameterization is that we cannot standardize the returns to get i.i.d. variables, and then use the standard diagnostic tools.

We can view this in another way: We cannot write the model like a GARCH model in the sense of Bougerol and Picard (1992), that is, split up r_t into a time constant distribution and an I_{t-1} measurable one

$$r_t = \delta_t y_t, \tag{2.2}$$

where, δ_t is I_{t-1} measurable, so, given the information up to time $t-1$, δ_t is a constant, and where y_t is i.i.d. To write the NIGSV(p,q) model as a product, as in (2.2), we have to choose

$$y_t \sim NIG\left(\alpha/\delta_t, 0, 0, 1\right),$$

where the $NIG\left(\alpha/\delta_t, 0, 0, 1\right)$ distribution clearly is not time constant. On the other hand, if we start out with $y_t \sim NIG\left(\alpha, 0, 0, 1\right)$, it is impossible to find a scale factor c^*, such that $r_t \sim NIG\left(\alpha, 0, 0, \delta_t\right)$, owing to the scaling properties in (2.1).

One might also say that the NIGSV(p,q) of Andersson is not a strong GARCH in the sense of Drost and Nijman (1993). Let us first define the idea of strong (and semi-strong) GARCH. Let $\{y_t\}$ be a sequence of stationary errors with finite fourth moments. Let $A\left(L\right)$ and $B\left(L\right)$ be as defined in (1.7) and (1.6) respectively, and let the sequence $\left\{\sigma_t^2\right\}$ be defined as the stationary solution of

$$B\left(L\right)\sigma_t^2 = \rho_0 + \left(A\left(L\right) - 1\right) y_t^2.$$

Then, the sequence $\{r_t\}$ is defined to be generated by a *strong* GARCH(p,q) process if ρ_0, $A\left(L\right)$, and $B\left(L\right)$ can be chosen such that

$$y_t = \frac{r_t}{\sigma_t} \sim i.i.d.\left(0, 1\right). \tag{2.3}$$

Similarly, the sequence $\{r_t\}$ is defined to be generated by a *semi-strong* GARCH(p,q) process if ρ_0, $A\left(L\right)$, and $B\left(L\right)$ can be chosen such that

$$E\left(r_t | r_{t-1}, r_{t-2}, ...\right) = 0,$$

and

$$E\left(r_t^2 | r_{t-1}, r_{t-2}, ...\right) = \delta_t.$$

It is clear from the above that the sequence of δ_t in (1.8) does not fulfill the condition for strong GARCH. Instead, the NIGSV(p,q) model of Andersson (2001) is a semi-strong GARCH.

2.2 A new scale invariant parameterization of NIG

We would like to find a parameterization of the NIG distribution that is a strong GARCH and where we can write the model as a product of a time-constant distribution and an I_{t-1} measurable one, i.e., a GARCH model in the sense of Bougerol and Picard (1992), or a strong GARCH in the sense of Drost and Nijman (1993). Furthermore, as we are dealing with conditional variances, it would be more intuitive to find a parameterization of the NIG distribution that has only one parameter defining the variance.[2] This is possible if we start out from the scale invariant parameterization in (1.4), and we let

$$\begin{aligned} \beta &= 0, \\ \overline{\alpha} &= \alpha\delta, \end{aligned}$$

and

$$\sigma^2 = \frac{\delta}{\alpha}.$$

The density of the resulting parameterization, which we shall denote $NIG_{\sigma^2}\left(\overline{\alpha}, 0, \mu, \sigma^2\right)$ can be written

$$g\left(\overline{\alpha}, \sigma^2\right) = \frac{\sqrt{\overline{\alpha}}}{\pi\sqrt{\sigma^2}} \exp\left(\overline{\alpha}\right) q\left(\frac{z-\mu}{\sqrt{\sigma^2\overline{\alpha}}}\right)^{-1} K_1\left(\overline{\alpha} q\left(\frac{z-\mu}{\sqrt{\sigma^2\overline{\alpha}}}\right)\right), \tag{2.4}$$

where $q(x) = \sqrt{1+x^2}$, and $K_1(z)$ is the modified Bessel function of third order and index one. Restrictions for the parameters are $0 \leq \overline{\alpha}$, $\mu \in R$ and $0 \leq \sigma^2$.

The first four central moments are

$$\begin{aligned} \mu_1 &= \mu, \\ \mu_2 &= \sigma^2, \\ \mu_3 &= 0, \end{aligned}$$

and

$$\mu_4 = 3\sigma^4 + \frac{3\sigma^4}{\overline{\alpha}}.$$

Note that the variance is represented by one parameter, σ^2, which might be more intuitive in the context of volatility modeling. The kurtosis is given by

$$K = 3 + \frac{3}{\overline{\alpha}}.$$

2.2.1 Scaling properties of the new parameterization

The scaling properties of the $NIG_{\sigma^2}\left(\overline{\alpha}, 0, \mu, \sigma^2\right)$ parameterization are given by the following. Let $Z_1 \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, \mu, \sigma^2\right)$, then

$$cZ_1 \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, \mu, (c\sigma)^2\right),$$

[2] Jensen and Lunde (2001) use the scale invariant parameterization of Barndorff-Nielsen (1997) to build their model.

i.e., $\overline{\alpha}$ does not change under scaling. This means that if we use this parameterization in a conditional variance modeling framework, we can fit the model, standardize the observed returns using the conditional standard deviation, and the parameters of the distribution for the standardized returns will be constant. To see this, let

$$r_t \sim NIG\left(\overline{\alpha}, 0, \mu_t, \sigma_t^2\right),$$

be the daily returns, where $\mu_t = E\left(r_t | I_{t-1}\right)$ is the conditional mean of the returns and where the conditional variance σ_t^2 is modelled using a GARCH-specification.[3] Now, we standardize the observed returns

$$z_t = \frac{r_t - \mu_t}{\sigma_t},$$

and the standardized returns are distributed according to

$$NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, 1\right),$$

$$V\left(z_t\right) = 1.$$

2.3 The GARCH(p,q)-NIG model

We can derive the GARCH(p,q)-NIG model in two ways. We can view the model as a mixture-of-distributions model and start with the normal distribution, take the inverse Gaussian as the mixing density and then derive the model. Alternatively, we can view the model as a GARCH model with a NIG distribution instead of the normal or Student's t distribution.

In this thesis, we will derive the model using both methods, starting with the MDH derivation. Later we will use only the GARCH formulation, which tends to be easier to understand.

To derive the GARCH-NIG model we assume that the return r_t conditional on its variance z_t is normally distributed

$$(r_t | z_t) \sim N(\mu_t, z_t),$$

where $\mu_t = E\left(r_t | I_{t-1}\right)$ is the conditional mean. The variance z_t is inverse Gaussian given the information set up to, and including time $t-1$,

$$(z_t | I_{t-1}) \sim IG_{\sigma^2}(\sigma_t^2, \overline{\alpha}),$$

[3] For instance, we can model the conditional mean of the returns, μ_t by an ARMA(p,q) model.

where $I_{t-1} = \left(\sigma^2_{-p+1}, ..., \sigma^2_{t-1}, r_{-q+1}, ..., r_{t-1}\right)$. Note that the $E\left(z_t|I_{t-1}\right) = \sigma_t^2$, that is, the parameter σ_t^2 denotes the conditional mean of the variance. Now, the returns conditionally on I_{t-1} are normal inverse Gaussian

$$(r_t|I_{t-1}) \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, \mu_t, \sigma_t^2\right).$$

The conditional variance of the returns is given by

$$V\left(r_t|I_{t-1}\right) = \sigma_t^2,$$

which we model as

$$\sigma_t^2 = \rho_0 + \Sigma_{i=1}^q \rho_i r_{t-i}^2 + \Sigma_{j=1}^p \pi_j \sigma_{t-j}^2. \tag{2.5}$$

The conditional mean of the variance and the conditional variance of the returns are the same. That is, when we model the returns we implicitly model the mean of the (latent) variance. One would be justified in discussing whether it would be more appropriate to call this model a GARCH or a stochastic volatility model. For simplicity, we refer to this parameterization of the model as the GARCH-NIG model.

It is clear from the above that the GARCH-NIG is a strong GARCH(p,q) in the sense of Drost and Nijman (1993). Furthermore, we can write the model as a GARCH model in the sense of Bougerol and Picard (1992), with a standardized NIG error distribution. Whereby we split up the r_t into a factor with a time-constant distribution, and an I_{t-1} measurable one, i.e.,

$$r_t = y_t \sigma_t,$$

where $y_t \sim NIG\left(\overline{\alpha}, 0, 0, 1\right)$ and σ_t follows (2.5).

Chapter 3

Test of the Mixing Distribution Hypothesis

The Mixture-of-Distributions Hypothesis of Clark (1973) predicts that returns standardized by their conditional variance should be normally distributed, which we refer to henceforth as 'normal'. In this chapter we use a high frequency data set ECU/USD 1989 - 1998, sampled every five minutes, to construct realized volatility. We use these realized volatilities to standardize the returns and investigate whether the RV-standardized returns are normal, i.e., if the mixing distribution of Clark (1973) holds.

3.1 Introduction

Consistent with the absence of arbitrage and a time-changed Brownian motion (see, for example, Ane and Geman, 2000, and Andersen, Bollerslev and Diebold, 2002), the MDH postulates that the distribution of returns is normal, but with a stochastic (latent) variance. In the original formulation in Clark (1973) the variance is assumed to be i.i.d. lognormally distributed, resulting in a lognormal-normal mixture distribution for the returns. Numerous theoretical extensions and empirical investigations of these ideas, involving various proxies for the mixing variable have can be found in the literature (important early contributions include, Epps and Epps, 1976, Taylor, 1982, Tauchen and Pitts, 1983). Importantly, to account explicitly for the volatility clustering effect Taylor (1986) proposed an extension of the MDH setup by having the (latent) logarithmic variances follow a Gaussian autoregression, resulting in the lognormal Stochastic Volatility (SV) model; see also Andersen (1996). Since the joint distribution of the returns in the SV model is not known in a closed form, estimation and inference are considerably more complicated for these types of models than for the ARCH/GARCH class of models (see, e.g., Shephard, 1996).

In contrast to the existing SV literature where the mixing variable is treated as latent, here we proceed to show that by measuring the daily variance by the corresponding realized volatility, constructed from the sum of intraday high-frequency returns, the daily return standardized by the realized volatility is approximately normally distributed. Therefore, even though the realized volatilities are subject to measurement error vis-à-vis the true daily latent volatilities (see for instance Andreou and Ghysels, 2002, and Barndorff-Nielsen and Shephard, 2001b, 2002b), the (approximate) normality of the standardized returns is consistent with the basic tenets of the MDH and the reliance on the realized volatility as the underlying mixing variable. The empirical analysis is based on a ten-year sample of high-frequency five-minute returns for the ECU basket of currencies versus the U.S. Dollar spanning the period from January 3, 1989 through December 30, 1998.

Our results build directly on recent empirical findings and related theoretical developments in the literature. First, however, it should be noted that the idea of explicitly modeling realized volatility proxies has a long history in empirical finance (see for example, Schwert, 1989, and Hsieh, 1991, and more recently Andersen, Bollerslev, Diebold and Labys, 2002, and Maheu and McCurdy, 2002). Second, empirical results in Andersen, Bollerslev, Diebold and Labys (2000) and Andersen, Bollerslev, Diebold, and Ebens (2001) have previously demonstrated the approximate normality of the returns when standardized by the realized volatility for other asset classes and time periods.

3.2 The Mixture-of-Distributions Hypothesis

The Mixture-of-Distributions Hypothesis (MDH), starts from the premise that the distribution of discretely sampled returns, conditional on some latent information arrival process, is Gaussian. This assumption is justified theoretically if the underlying price process follows a continuous sample path diffusion as outlined in the introduction (Equation (1.9)), (see also the discussion in Andersen, Bollerslev, and Diebold, 2002, and Barndorff-Nielsen and Shephard, 2001b). In this setting, Barndorff-Nielsen and Shephard (1998) show that the returns conditional on the quadratic variation are normally distributed, that is

$$r_t | QV_t \sim N\left(0, QV_t\right), \tag{3.1}$$

where r_t is given by (1.10) and QV_t is the quadratic variation defined in (1.11), also called the integrated volatility of the process. However, the integrated volatility process, which serves as the mixture variable in this situation, is not directly observable. As noted above, this has spurred numerous empirical investigations into alternative volatility proxies and/or mixture variables. Meanwhile, as outlined in Chapter 1 and discussed further below, by using increasingly finer sampled returns, the integrated volatility in a diffusion process may in theory be estimated arbitrarily well by the so-called realized volatility, constructed by summing the of the squared high-frequency returns. This suggests the following empirically testable starting point for the MDH,

$$f(r_t|RV_t) \sim N(0, RV_t), \tag{3.2}$$

where r_t refers to the one-period returns sampled discretely from time $t-1$ to t, and RV_t denotes the corresponding realized volatility proxy measured over the same time interval. Recall that the realized volatility used in (3.2) is a consistent esimate of the quadratic variation used in (3.1) . Consistent with earlier related empirical results in ABDL (2000), the results for the high-frequency foreign exchange rates discussed in the next section are generally supportive of this hypothesis.

3.3 Data sources and realized volatility

Our primary data set consists of daily returns and realized volatilities for the ECU/US Dollar exchange rate from January 3, 1989 through December 30, 1998.[1] [2]Following standard practice in the literature, the daily realized volatilities are constructed from the summation of squared five-minute high-frequency returns. Formally, for $t = h, 2h, ..., T$

$$Var_{t,h} = \Sigma_{i=1}^{288h} r_{(288)}^2 \left(t - h + (i/288)\right), \tag{3.3}$$

where $r_{(288)}^2 \left(t + (i/288)\right)$ denotes the continuously compounded return for day t over the ith five-minute interval calculated on the basis of the linearly interpolated logarithmic midpoint of the bid-ask prices, and where h is the frequency $h = 1, 5, 10$ or 20 days, that is, for the daily, weekly, bi-weekly and monthly frequencies. We omit non-trading days and weekend periods as described in ABDL (2001). All in all, this leaves us with a total of 2,428 days.[3] Time series plots of the relevant returns and realized volatilities are given in Figures 3.1 and 3.2.

[1] All of the raw data were obtained from Olsen and Associates in Zürich, Switzerland.

[2] For simplicity, we refer to this dataset as the ECU/USD 1989 - 1998 dataset.

[3] We also excluded nine days in January and February 1989 on which the realized volatility was less than 0.005. These days are directly associated with problems in the data-feed early on in the sample. None of the results are sensitive to these additional exclusions.

The median of the data before the exclusion was 0.344 and the minimum was $2.3*10^{-5}$. The median of the data after the exclusion was 0.348 and the minimum was 0.016.

As can be seen from (3.3), by definition, the RV of a week is simply the sum of the RV over 5 days. That is, we aggregate the realized volatility in the same way as we aggregate compound returns.

3.4 Raw Returns

We start with a description of the raw returns. Figure 3.1 shows time series plots of the returns for the daily, weekly, bi-weekly and monthly frequencies. The volatility clustering-effects are obvious, at least for the daily, weekly and bi-weekly frequencies. The volatility clustering, or ARCH effects, are also seen in Figure 3.2 a,b, which shows the time series plots of the realized volatilities for the different frequencies. Recall, that the dependence in the *conditional variance* for the returns, translates into a dependence in the *conditional mean* for the realized volatility.

In Figure 3.3 a,b we see the unconditional distribution of the raw returns together with a fitted normal distribution. The empirical distribution of the raw returns is peaked and have a fatter tail than the normal distribution, at least at the daily and weekly frequencies. QQ-plots of the probability integral transform of the raw returns assuming them to be normal against the quantiles of the $U(0,1)$ distribution, are given in Figure 3.4.[4] For the daily frequency (Figure 3.4 a), we see the typical s-shaped QQ-plot, meaning that the daily raw returns have a fatter tail than the normal distribution. This pattern is less obvious for the weekly, bi-weekly and monthly frequencies. Descriptive statistics for the raw returns are presented in the left columns of Table 3.1a and b. We note, that except for the daily raw returns, the raw returns are skewed with a coefficient of skewness of about -0.5 for the weekly, bi-weekly and monthly frequencies. The kurtosis of the daily returns is 5.425, and 5.076 for the weekly returns, while for the bi-weekly and monthly frequencies it is 3.495 and 4.01, respectively.[5] The Jarque-Bera test (JB-test) for normality ((Jarque and Bera, 1987)) is rejected for all the frequencies. Taken together, this is strong evidence for non-normality of the raw returns.

Ljung-Box Q-statistics for serial dependence in the returns and the squared returns are reported in Table 3.2a and b. As noted frequently in the literature there seems to be no dependence in the first moment, but the daily squared raw returns display significant serial dependence, both at lag 1 and lag 10. The lower frequencies do not show serial dependence for the squared raw returns,

[4] The PIT is defined as $z_t^{PIT} = \int_{-\infty}^{x_t} f(u)\,du$. If $f(u)$ is the correct distribution, then $z_t^{PIT} \sim U(0,1)$.

[5] This kurtosis measure is $K_4 = \frac{E(x-\mu_x)^4}{\left(E(x-\mu_x)^2\right)^2}$, so the normal distribution has $K_4 = 3$.

suggesting that the volatility clustering vanishes with aggregation. The left panel of Figure 3.5 displays the sample autocorrelation function for the daily raw returns. The Sample Autocorrelation Function (SACF) of the absolute returns starts at 0.12 and decays slowly. The SACF of the squared raw returns also starts at 0.12 and decays slowly, but faster than for the absolute returns.

3.5 RV-standardized returns

Here we report the results for the RV-standardized returns, that is

$$r_t^* = \frac{r_{t,raw}}{\sqrt{RV_t}},$$

where $r_{t,raw}$ is the observed daily return at time t, RV_t is the realized volatility, and r_t^* is the RV-standardized return.

Figure 3.6 shows the RV-standardized returns and a fitted normal distribution. The empirical distributions of the RV-standardized returns is less peaked than the distribution for the raw returns, and the fit of the normal distribution is better for the RV-standardized returns than that of the raw returns in Figure 3.3, which is supported in the QQ-plots in Figure 3.4. The QQ-plot of the daily RV-standardized returns against the normal distribution are almost a straight line, and the visual impression is the same for the weekly, bi-weekly and monthly frequency, indicating that the normal distribution gives a good fit to the RV-standardized returns.

The normality of the RV-standardized returns is confirmed by the statistics in Table 3.1a and b. Compared to the statistics for the raw returns, the skewness is lower for all the frequencies, and the kurtosis is closer to three for all the frequencies for the RV-standardized returns than for the raw returns. Formally, using the JB-test, we cannot reject normality for any of the frequencies of the RV-standardized returns.

The serial dependence of the squared RV-standardized daily returns is lower than for the squared daily raw returns, as seen in the right panel of Figure 3.3, and Table3.2a and b. The p-value of the Ljung-Box statistic for the squares for one lag is 0.271 for the RV-standardized returns, in contrast to 0.000 for the raw squared return, indicating that standardizing the returns by the realized volatility takes out the volatility clustering effect as predicted by the MDH.

Both the normality of the RV-standardized returns, and the lack of serial dependence in the squares of the RV-standardized returns provide support for the Mixture-of-Distributions Hypothesis.

3.6 Conclusions

Using a high frequency data set consisting of five minute returns from the ECU/USD 1989 - 1998 exchange rate, for which we calculate the realized volatility, we have shown we cannot reject normality of the RV-standardized returns. That is, we found that the Mixture-of-Distributions Hypothesis of Clark (1973) cannot be rejected for this dataset.

3.7 Futher work

In this chapter, we study only one realized volatility dataset. It would be interesting to see whether the same result holds true for other datasets. It would also be interesting to see if the result would change in any direction if we filter the realized volatilty using the filters proposed in Andreou and Ghysels (2002). One could also try to incorporate the results concerning the asymptotic distribution of the (sampling) error in the realized volatility, i.e., the results of Barndorff-Nielsen and Shephard, 2001b, 2002b.

3.8 Tables

Table 3.1a:
Descriptives of unconditional returns of ECU/USD 1989 - 1998.

	Daily n=2428		Weekly n=445	
	Raw	RV-Stand.	Raw	RV-Stand
Mean	0.002	0.008	0.013	0.020
Median	0.000	-0.001	0.016	0.018
Maximum	3.141	3.497	4.374	2.564
Minimum	-3.257	-3.212	-7.583	-3.242
Std	0.638	0.950	1.417	0.934
Skewness	-0.079	0.027	-0.533	-0.220
Kurtosis	5.426	3.199	5.076	3.057
JB test stat	598.1	4.318	110.1	3.976
	(0.000)	(0.115)	(0.000)	(0.137)

Notes: RV-stand. means that returns are standardized by using the realized volatility, $r^*_{t,h} = r^{raw}_{t,h}/\sqrt{RV_{t,h}}$, where $r^*_{t.h}$ is the RV-standardized returns the raw return at time t for frequency h and $RV_{t,h}$ is the realized at time t and h is the frequency, $r^{raw}_{t,h}$ is volatility at time t for frequency h. The daily returns are standardized by using the daily RV and the weekly RV-standardized RV standardized returns are standardized using the weekly RV. JB stands for the Jarque-Bera test for normality,the p-values are in parenthesis.

Table 3.1b:
Descriptives of unconditional returns of ECU/USD 1989 - 1998.

	Bi-weekly n=242		Monthly n=121	
	Raw	RV-Stand.	Raw	RV-Stand
Mean	0.031	0.037	0.062	0.049
Median	0.090	0.049	0.175	0.089
Maximum	4.126	1.987	7.112	2.235
Minimum	-7.065	-2.520	-9.945	-2.785
Std	1.998	0.939	2.783	0.916
Skewness	-0.541	-0.308	-0.584	-0.334
Kurtosis	3.495	2.673	4.010	3.169
JB test stat	14.26	4.904	12.018	2.390
	(0.001)	(0.086)	(0.003)	(0.303)

Notes: RV-stand. means that returns are standardized by using the realized volatility, $r^*_{t,h}= r^{raw}_{t,h}/\sqrt{RV_{t,h}}$, where $r^*_{t.h}$ is the RV-standardized returns the raw return at time t for frequency h and $RV_{t,h}$ is the realized at time t and h is the frequency, $r^{raw}_{t,h}$ is volatility at time t for frequency h. The daily returns are standardized by using the daily RV and the weekly RV-standardized RV standardized returns are standardized using the weekly RV. JB stands for the Jarque-Bera test for normality,the p-values are in parenthesis.

Table 3.2a:
Descriptives of time series properties of returns of ECU/USD
1989 - 1998.

	Daily ($h=1$) n=2428		Weekly ($h=5$) n=445	
	Raw	RV-Stand.	Raw	RV-Stand
$Q\,(1)$	0.041	1.034	0.003	0.051
	(0.840)	(0.309)	(0.954)	(0.822)
$Q\,(10)$	20.269	21.468	10.772	6.099
	(0.027)	(0.018)	(0.376)	(0.807)
$Q^2\,(1)$	34.386	1.208	0.990	0.200
	(0.000)	(0.271)	(0.319)	(0.654)
$Q^2\,(10)$	208.4	30.948	14.673	23.540
	(0.000)	(0.001)	(0.144)	(0.009)

Notes: RV-stand. means that returns are standardized by using the realized volatility, $r^*_{t,h}= r^{raw}_{t,h}/\sqrt{RV_{t,h}}$, where $r^*_{t.h}$ is the RV-standardized returns at time and h is the frequency, $r^{raw}_{t,h}$ is the raw return at time t for frequency h and $RV_{t,h}$ is the realized volatility at time t for frequency h. The daily RV-standardized returns are standardized by using the daily RV and the weekly RV standardized returns are standardized using the weekly RV. JB stands for the Jarque-Bera test for normality, the p-values are given in parenthesis.
$Q\,(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized returns forup to lag (n). $Q^2\,(n)$ denotes the Ljung-Box statistic for the squared standardized residuals. The p-values are given in parenthesis.

Table 3.2b: Returns
Descriptives of time series properties of returns of ECU/USD
1989 - 1998.

	Bi-weekly ($h = 10$) n=242		Monthly ($h = 10$) n=121	
	Raw	RV-Stand.	Raw	RV-Stand
$Q\,(1)$	0.976	0.765	2.545	1.558
	(0.323)	(0.382)	(0.111)	(0.212)
$Q\,(10)$	8.516	5.575	10.176	8.104
	(0.579)	(0.850)	(0.425)	(0.619)
$Q^2\,(1)$	1.933	1.639	1.424	0.219
	(0.164)	(0.200)	(0.233)	(0.640)
$Q^2\,(10)$	17.254	13.782	7.249	11.379
	(0.069)	(0.183)	(0.702)	(0.329)

Notes: RV-stand. means that returns are standardized by using the realized volatility, $r^*_{t,h}= r^{raw}_{t,h}/\sqrt{RV_{t,h}}$, where $r^*_{t.h}$ is the RV-standardized returns at time and h is the frequency, $r^{raw}_{t,h}$ is the raw return at time t for frequency h and $RV_{t,h}$ is the realized volatility at time t for frequency h. The daily RV-standardized returns are standardized by using the daily RV and the weekly RV standardized returns are standardized using the weekly RV. JB stands for the Jarque-Bera test for normality, the p-values are given in parenthesis.
$Q\,(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized returns forup to lag (n). $Q^2\,(n)$ denotes the Ljung-Box statistic for the squared standardized residuals. The p-values are given in parenthesis.

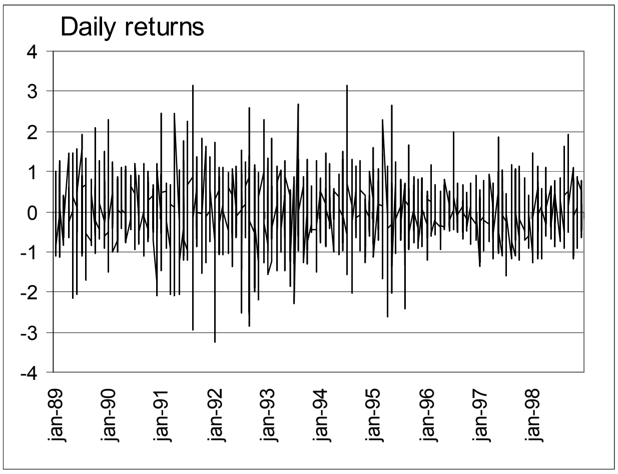
Table 3.3:
Descriptives of unconditional raw RV of ECU/USD
1989 - 1998.

	Frequency			
	Daily	Weekly	Bi-weekly	Monthly
Number of obs.	2428	485	242	121
Mean	0.427	2.135	4.271	8.542
Median	0.346	1.902	3.761	7.894
Maximum	3.870	9.477	16.865	28.481
Minimum	0.017	0.353	1.265	3.009
Std	0.311	1.177	2.096	3.703
Skewness	2.813	2.008	1.832	1.617
Kurtosis	17.984	9.231	8.786	8.594
$Q\,(1)$	667.9	184.8	79.123	29.352
	(0.000)	(0.000)	(0.000)	(0.000)
$Q\,(10)$	3147.0	426.1	135.0	43.242
	(0.000)	(0.000)	(0.000)	(0.000)

$Q\,(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized returns forup to lag (n). $Q^2\,(n)$ denotes the Ljung-Box statistic for the squared standardized residuals. The p-values are given in parenthesis.

3.9 Figures

Figure 3.1a:
Returns of the ECU/USD 1989 - 1998.



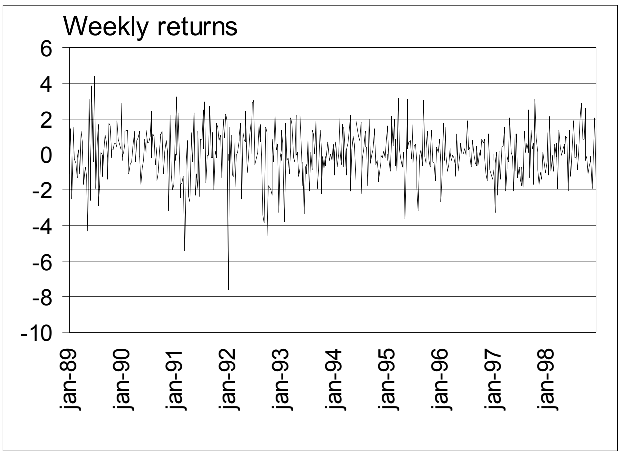


Figure 3.1b:
Returns of the ECU/USD 1989 - 1998.

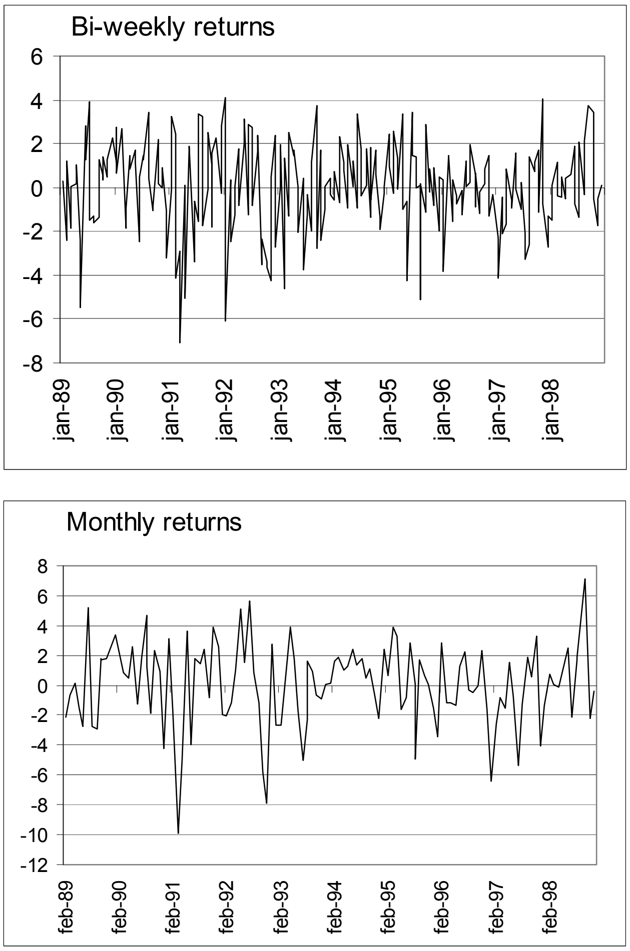


Figure 3.2a:
Time series plots of realized volatility of ECU/USD 1989 - 1998.

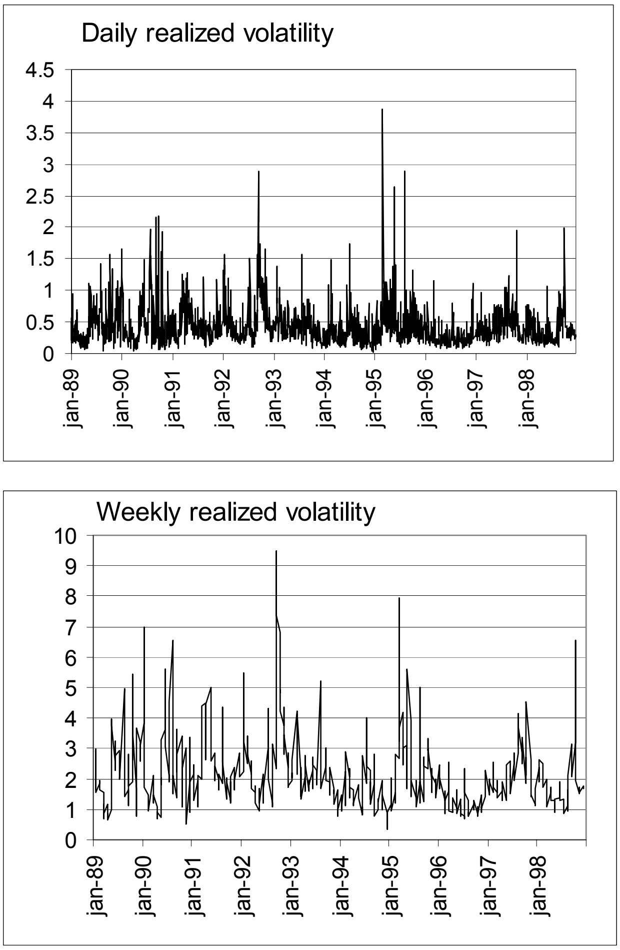
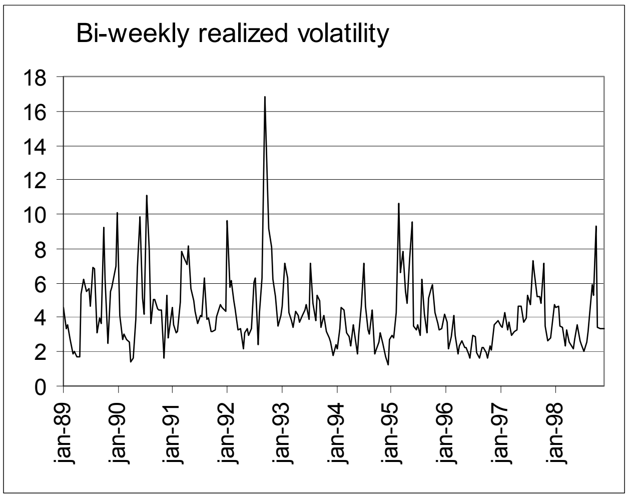


Figure 3.2b:
Time series plots of realized volatility of ECU/USD 1989 - 1998.



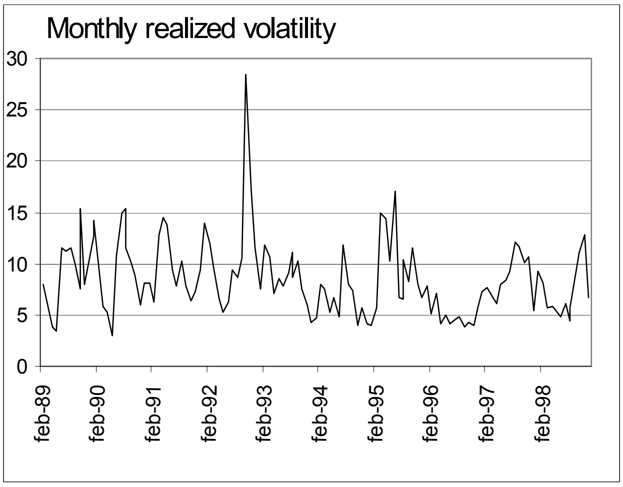


Figure 3.3a:
Raw returns of ECU/USD 1989 - 1998,
and fitted normal distribution.

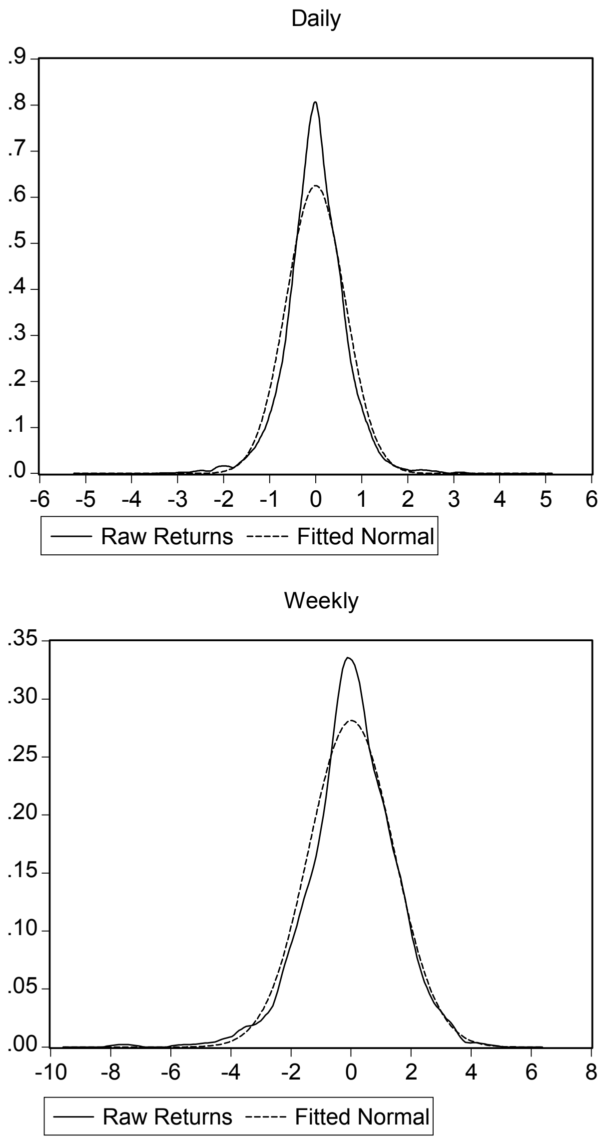
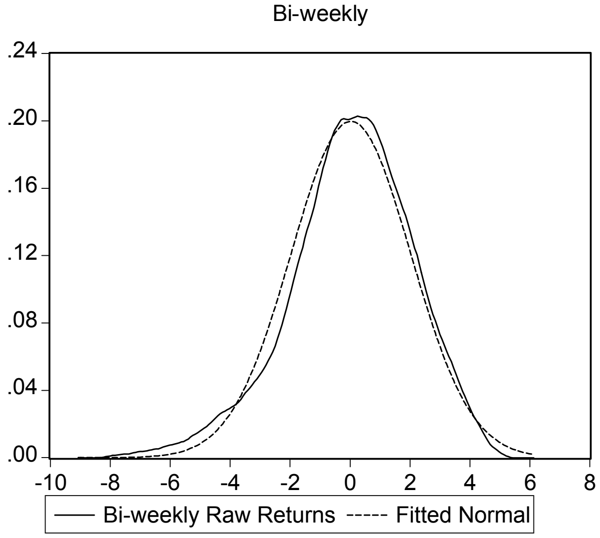


Figure 3.3b:
Raw returns of ECU/USD 1989 - 1998,
and fitted normal distribution.



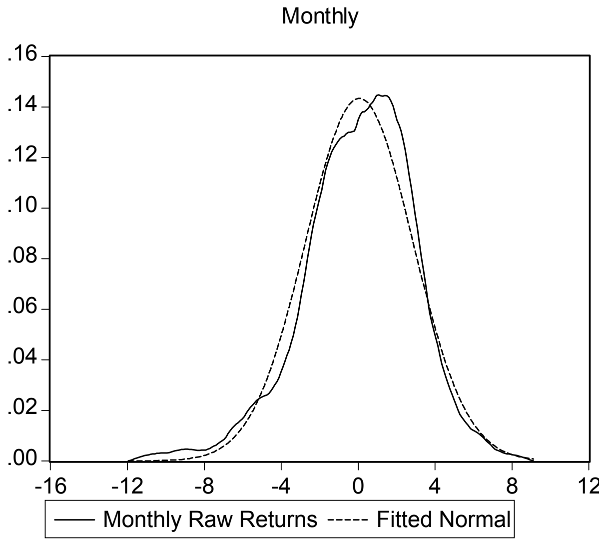
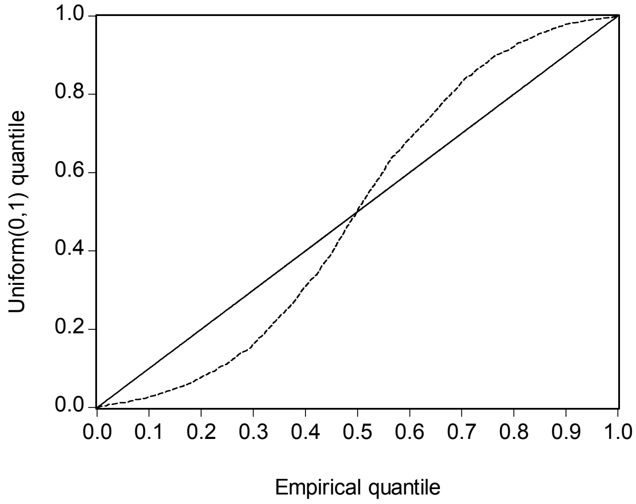


Figure 3.4a:

QQ plot of PIT of Daily ECU/USD 1989-1998 assuming them to be N(0,1).



QQ plot of PIT of Daily ECU/USD 1989-1998 standardized using RV assuming them to be N(0,1).

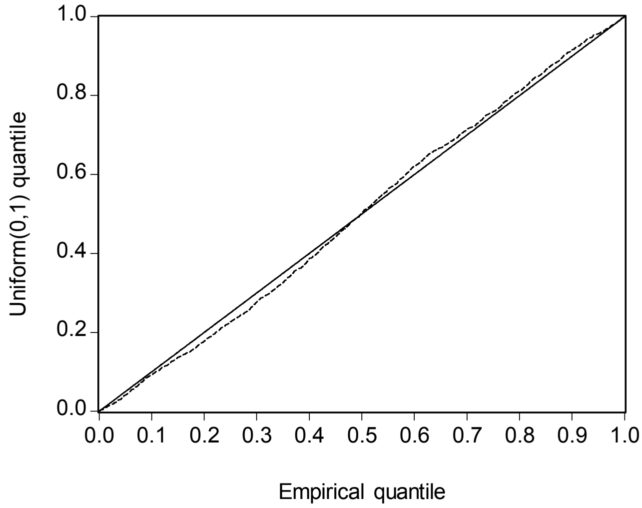
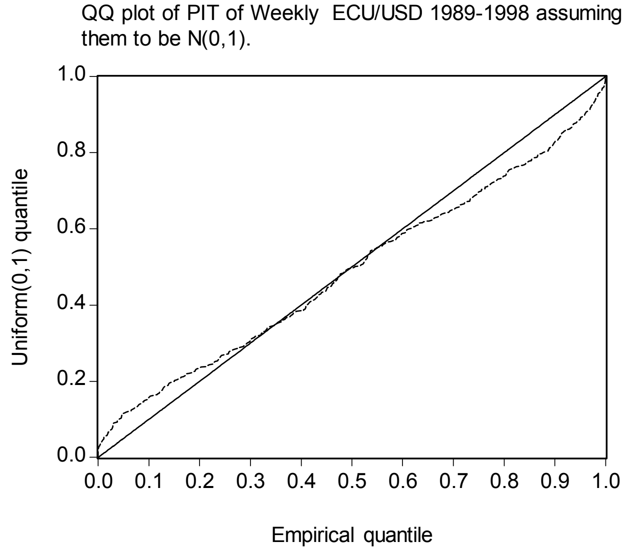


Figure 3.4b:


QQ plot of PIT of Weekly ECU/USD 1989-1998 assuming them to be N(0,1).
Uniform(0,1) quantile
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Empirical quantile

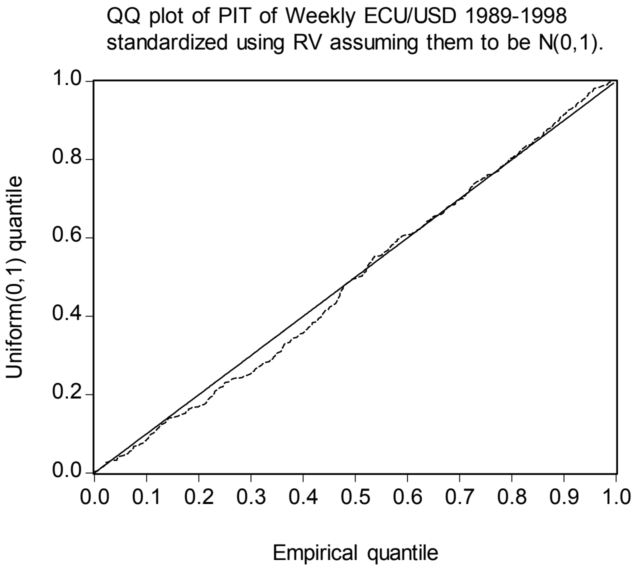

QQ plot of PIT of Weekly ECU/USD 1989-1998 standardized using RV assuming them to be N(0,1).
Uniform(0,1) quantile
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Empirical quantile

Figure 3.4c:

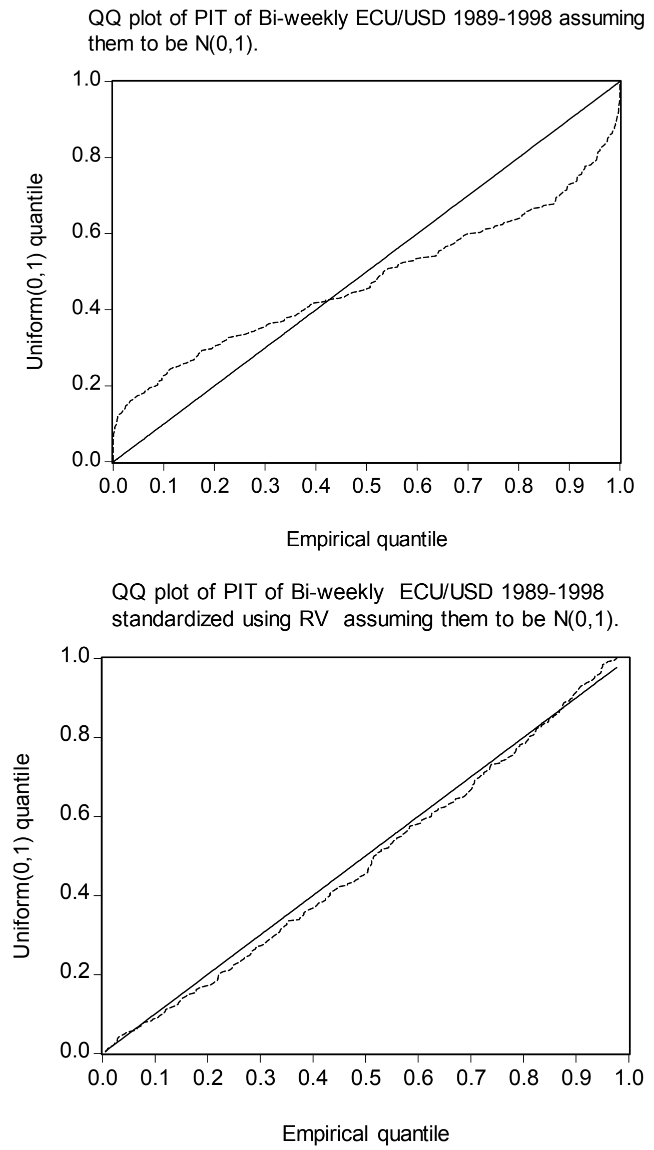
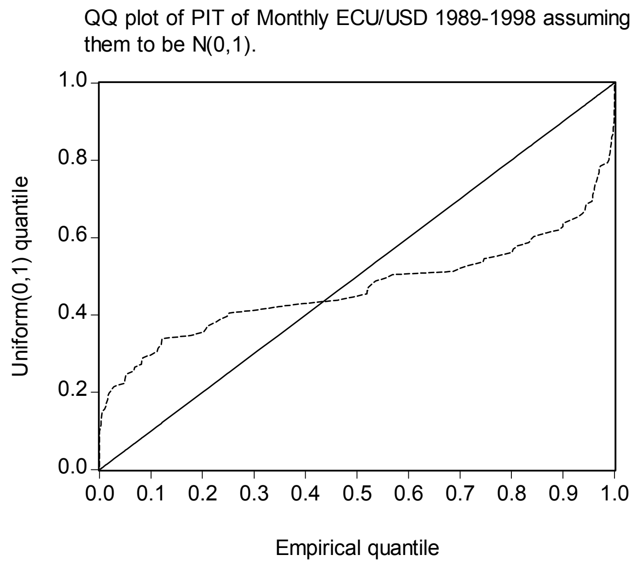

QQ plot of PIT of Bi-weekly ECU/USD 1989-1998 assuming them to be N(0,1).
Uniform(0,1) quantile
Empirical quantile
QQ plot of PIT of Bi-weekly ECU/USD 1989-1998 standardized using RV assuming them to be N(0,1).
Uniform(0,1) quantile
Empirical quantile

Figure 3.4d:


QQ plot of PIT of Monthly ECU/USD 1989-1998 assuming them to be N(0,1).
Uniform(0,1) quantile
Empirical quantile
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

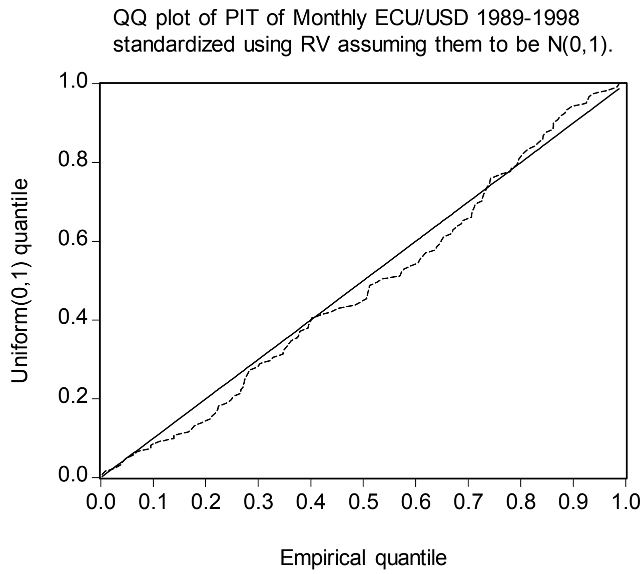
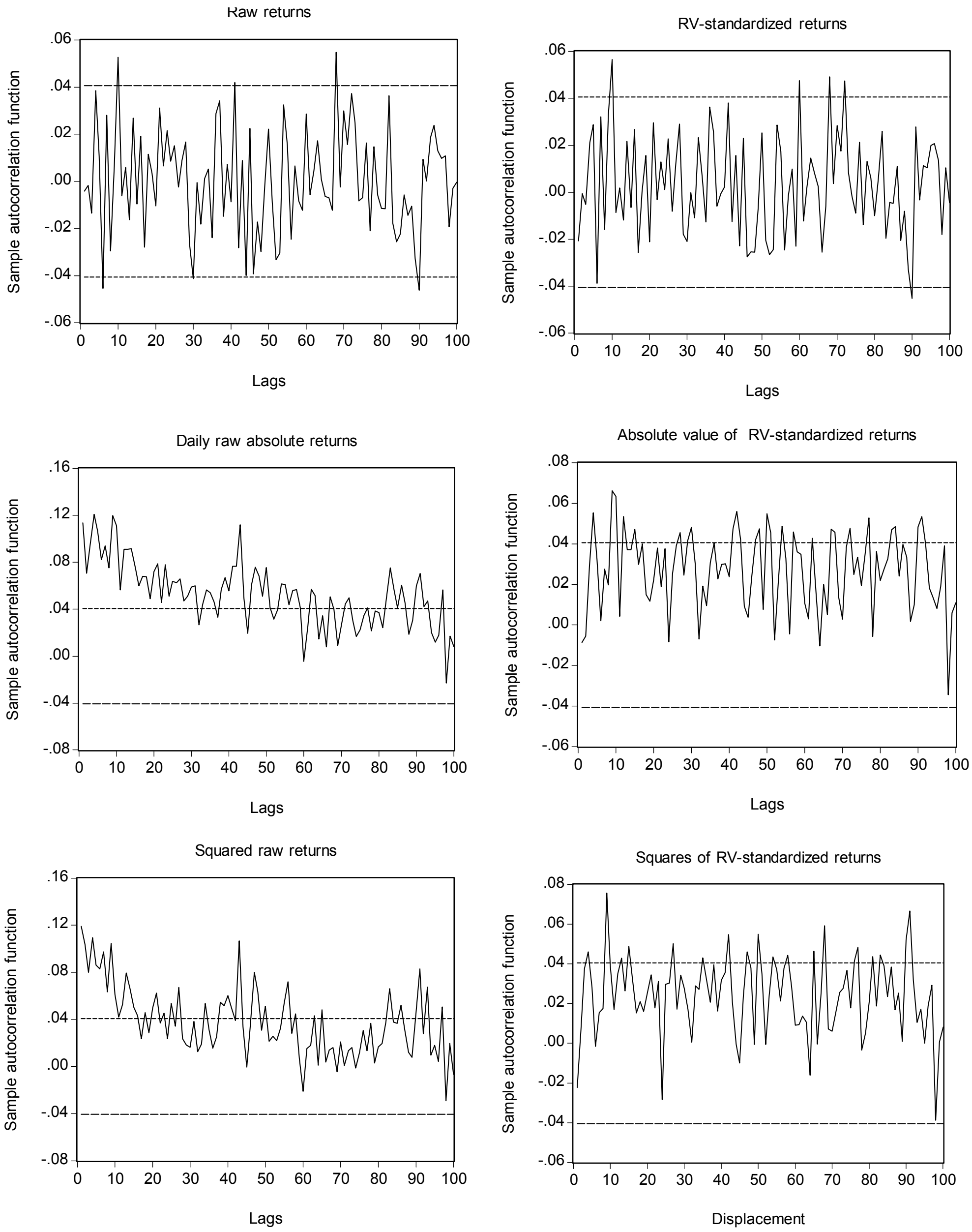
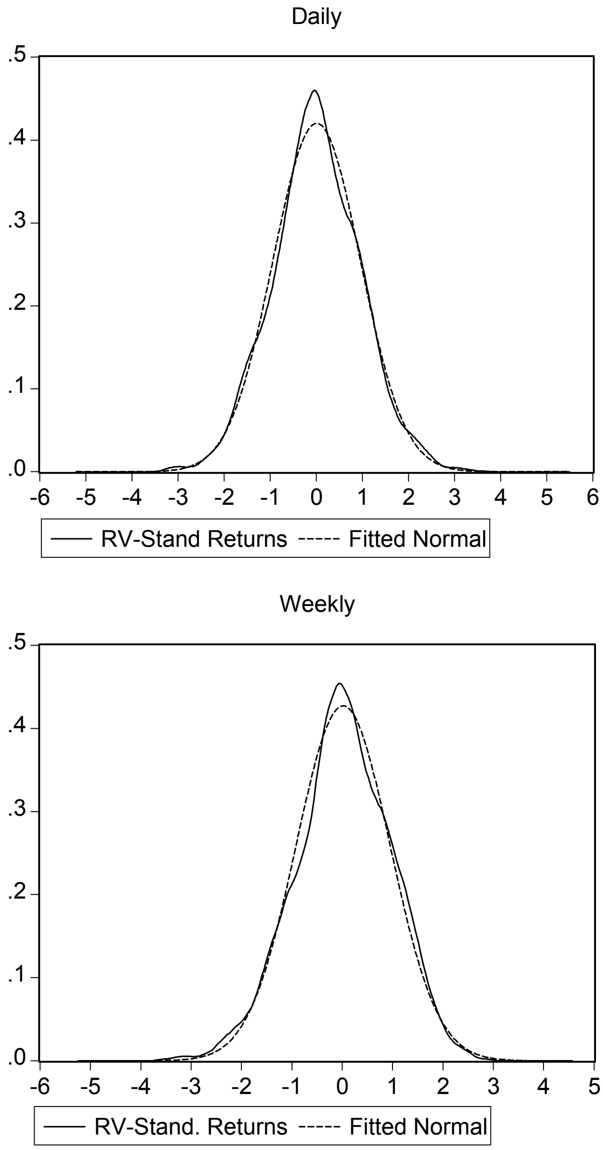

QQ plot of PIT of Monthly ECU/USD 1989-1998 standardized using RV assuming them to be N(0,1).
Uniform(0,1) quantile
Empirical quantile
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Figure 3.5:
Sample autocorrelation functions of daily exchange rate returns of the ECU/USD 1989 - 1998.



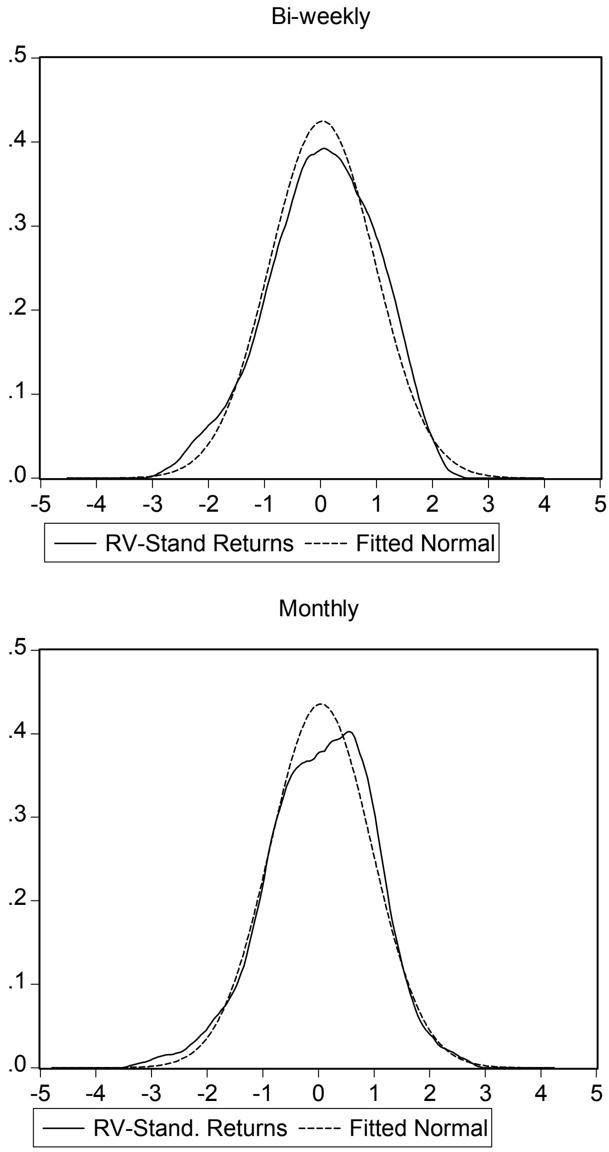
Notes: The dotted lines are 95% confidence bands.

Figure 3.6a:
RV-standardized returns of ECU/USD
1989 - 1998, and fitted normal distribution.



Notes: The RV-standardized returns are standardized by using the realized volatility, $r_t^* = r_t^{raw}/\sigma_t$, where σ_t^2 is the realized volatility. The daily RV-standardized returns are standardized using the daily RV, the weekly RV-standardized returns are standardized using the weekly RV, and so on.

Figure 3.6b:
RV-standardized returns of ECU/USD 1989 - 1998,
and fitted normal distribution.



Notes: The RV-standardized returns are standardized by using the realized volatility, $r_t^* = r_t^{raw}/\sigma_t$, where σ_t^2 is the realized volatility. The daily RV-standardized returns are standardized using the daily RV, the weekly RV-standardized returns are standardized using the weekly RV, and so on.

Chapter 4

Motivating the GARCH(p,q)-NIG

In Chapter 3, we found that when using realized volatility the standardized returns are normal, i.e., that the MDH holds. We need to make a distributional assumption for the volatility if we are to build a model for it. In this chapter, building on the Mixture-of-Distributions Hypothesis (MDH) (Clark, 1973) along with the recent idea of so-called Realized Volatilities (RV) (e.g., Andersen, Bollerslev, Diebold and Labys 2001, 2002, and Barndorff-Nielsen and Shephard, 2001a, b, 2002a), we provide a sound empirical foundation for the distributional assumptions behind the GARCH-NIG model.

We use high frequency data from the ECU/USD 1989 - 1998 FX rate to motivate the GARCH-NIG model.[1] By calculating realized volatilities for that period, we show that the inverse Gaussian distribution gives a good fit to the realized volatility, both unconditionally and conditionally on two different information sets.

In addition to the daily ECU returns and realized volatilities, we perform an out-of-sample predictive analysis based on 780 daily returns for the Euro from January 5, 1999 through December 31, 2001.[2] We do not have access to the high-frequency data underlying the construction of the realized volatilities over this more recent time period. Hence, our empirical verification of the various distributional hypotheses involving the RV_t variable outlined below, will be based exclusively on the ECU data.

The chapter is organized as follows. We start by establishing the link be-

[1] This is the same dataset that was used in Chapter 3, i.e., ECU/USD 10 Jan. 1989 - 30 Dec. 1998.

[2] We refer to this dataset as the Euro/USD 1999 - 2001 dataset, or just the Euro/USD dataset.

tween the realized volatility and the conditional variance by linking together the idea of realized volatility, quadratic variation and conditional variance. This will enable us to use the (conditional) realized volatility as a proxy for the conditional variance to evaluate the assumption of an inverse Gaussian distribution for the conditional variance. We introduce the inverse Gaussian and the inverse gamma distributions for the variance, and we outline how to use these distributional assumptions for the variance to derive the distributions for the returns, starting with the unconditional returns. Then we derive the distribution of the returns, conditional on lagged realized volatilities, and the distribution of the returns conditionally on the lagged returns, which will give us the GARCH(p,q)-NIG model.

4.1 The link between realized volatility and conditional variance

We use the realized volatility to motivate the GARCH-NIG model. Recall that the GARCH-NIG model assumes that the *conditional variance* is inverse Gaussian. We will show that the IG distribution gives a good fit to the *conditional realized volatility.* Therefore, we need to establish the link between the conditional realized volatility and the conditional variance. For this, we need to use continuous time arguments. As outlined in the introduction, and repeated here for the reader's convenience, we assume that the log price follows a univariate diffusion process with no mean dynamics

$$dp\left(t\right) = \sigma\left(t\right) dW\left(t\right), \tag{4.1}$$

where $p\left(t\right)$ denotes the log price at time t, W is a standard Brownian motion, and $\sigma\left(t\right)$ is the instantaneous volatility or the spot volatility. The return at time t is then defined as

$$r_t \equiv p\left(t\right) - p\left(t-1\right) = \int_{t-1}^{t} \sigma\left(s\right) dW\left(s\right), \tag{4.2}$$

where r_t is the continuously compounded return at time t. ABDL (2001) use the quadratic variation (QV) of the process as a volatility measure, that is

$$QV_t = \int_{t-1}^{t} \sigma^2\left(s\right) ds. \tag{4.3}$$

The expression $\int_{t-1}^{t} \sigma^2\left(s\right) ds$ also defines the so-called *integrated volatility* (IV_t). ABDL (2001) show that the conditional expectation of the quadratic variation is the conditional variance of the returns, that is

$$E\left(QV_t|\psi_{t-1}\right) = V\left(r_t|\psi_{t-1}\right), \tag{4.4}$$

where ψ_t is the σ field that reflects the information at time t. Equation (4.4) is of course silent about the distribution of $QV_t|\psi_{t-1}$, but if we make an assumption concerning the distribution of $QV_t|\psi_{t-1}$, this will imply a distribution for the conditional return.

However, the quadratic variation is a theoretical concept so it is unobservable in practice. The empirical justification of a model requires the use of observables. Here, we will use the realized volatility as an estimate of the quadratic variation. The realized volatility is a consistent estimate of the quadratic variation,

$$p \lim_{m\to\infty} RV_{t,h} = QV_{t,h}. \tag{4.5}$$

This means that we will use the approximation

$$E\left(RV_t|F_{t-1}\right) \simeq E\left(QV_t|\psi_{t-1}\right), \tag{4.6}$$

where F_{t-1} is the information set consisting of lagged realized volatilities. That is,

$$F_{t-1} = \sigma\left\{RV_{t-1}, RV_{t-2}, ...\right\}. \tag{4.7}$$

The information set F_t is, of course, only a subset of ψ_t. Using the result in (4.4), the approximation in (4.6) implies that we also make the following approximation

$$E\left(RV_t|F_{t-1}\right) \simeq V\left(r_t|\psi_{t-1}\right). \tag{4.8}$$

Since we can observe RV_t, we can investigate different distributional assumptions for $RV_t|F_{t-1}$. This means that, given the approximation in (4.8), we can use the realized volatility to investigate the validity of the distributional assumptions for the conditional variance. We will show that the distribution for $RV_t|F_{t-1}$ is well described by the inverse Gaussian distribution, that is

$$RV_t|F_{t-1} \sim IG_{\sigma^2}\left(\sigma_t^2, \overline{\alpha}\right),$$

this implies

$$E\left(RV_t|F_{t-1}\right) = \sigma_t^2, \tag{4.9}$$

that is, σ_t^2 is the conditional mean of the inverse Gaussian distribution. We can now write the (conditional) distribution of the returns as

$$f\left(r_t|F_{t-1}\right) = \int_0^\infty f\left(r_t|RV_t\right) f\left(RV_t|F_{t-1}\right) d\left(RV_t|F_{t-1}\right) \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, \sigma_t^2\right).$$

Since the returns now are NIG, this implies that the conditional variance of the returns is given by

$$V\left(r_t|F_{t-1}\right) = \sigma_t^2.$$

Again, recall that the conditional mean of the inverse Gaussian distribution is the conditional variance of the normal inverse Gaussian distribution.

However, when we only have access to daily data, we cannot use the information set F_{t-1}. In this situation, we define an information set generated by the historical squared returns

$$I_{t-1} = \sigma \left\{ r_{t-1}^2, r_{t-2}^2 ... \right\}. \tag{4.10}$$

Now we proceed to investigate distributional assumptions of $RV_t | I_{t-1}$. Assuming that

$$RV_t | I_{t-1} \sim IG_{\sigma^2} \left(\sigma_t^2, \overline{\alpha} \right),$$

yields that the returns, conditional on the information set I_{t-1} is NIG distributed, that is

$$f \left(r_t | I_{t-1} \right) = \int_0^\infty f \left(r_t | RV_t \right) f \left(RV_t | I_{t-1} \right) d \left(RV_t | I_{t-1} \right) \sim NIG_{\sigma^2} \left(\overline{\alpha}, 0, 0, \sigma_t^2 \right).$$

We let the conditional variance parameter σ_t^2 be time varying according to a GARCH specification, which gives us the GARCH(p,q)-NIG model. Showing that $RV_t | I_{t-1}$ is well described by the IG distribution will give us the empirical motivation for the GARCH(p,q)-NIG model.

To summarize, our link between the realized volatility and the conditional variance is that the conditional expectation of the quadratic variation is the conditional variance of the returns, i.e., Equation (1.12), and the fact that realized volatility is a consistent estimate of the quadratic variation.

Given this link between the realized volatility and the conditional variance that we established using continuous time arguments, we will go back to discrete time in the remainder of this chapter.

We will compare the assumption that the conditional variance is IG distributed, with the assumption that it is inverse gamma distributed. These two distributions for the conditional variance, will imply the GARCH-NIG and the GARCH-t models for the conditional returns, respectively.

4.2 The inverse gamma and the inverse Gaussian distributions

The distribution in (3.2) and the unconditional and the conditional distribution of RV_t jointly determine the distribution of the returns. In the original MDH formulation advocated by Clark (1973), the latent mixing variable is assumed to be i.i.d. lognormally distributed, resulting in an unconditional lognormal-normal mixture distribution. This same lognormal distribution has previously been found to approximate the unconditional distribution of realized exchange rate volatility closely (Andersen, Bollerslev, Diebold and Labys, 2001). Meanwhile, the lognormal distribution is not closed under temporal aggregation, and the density function for the resulting lognormal-normal mixture distribution is only available in integral form. Partly in response to these concerns, Barndorff-Nielsen and Shephard (2001b, 2002a) have recently demonstrated that the unconditional distribution of the realized volatility may be equally well approximated by the inverse Gaussian (IG) distribution.

$$f\left(RV_t\right) \sim IG_{\sigma^2}\left(\sigma^2, \overline{\alpha}\right), \tag{4.11}$$

where the density function for the IG distribution may be expressed in standardized form as,[3]

$$IG_{\sigma^2}\left(z; \sigma^2, \overline{\alpha}\right) = \frac{\left(\frac{1}{\overline{\alpha}\sigma^2}\right)^{-1/2} z^{-3/2}}{\left(2\pi\right)^{1/2}} \exp\left\{\overline{\alpha} - \frac{1}{2}\left(\frac{\overline{\alpha}\sigma^2}{z} + \frac{\overline{\alpha}z}{\sigma^2}\right)\right\}. \tag{4.12}$$

The mean, variance, skewness and kurtosis of this IG distribution are, respectively,

$$\begin{aligned} E\left(z\right) &= \sigma^2, \\ V\left(z\right) &= \frac{\sigma^4}{\overline{\alpha}}, \\ S\left(z\right) &= \frac{3}{\sqrt{\overline{\alpha}}}, \end{aligned} \tag{4.13}$$

and

$$K\left(z\right) = 3 + \frac{15}{\overline{\alpha}}.$$

In contrast to the lognormal distribution, the IG distribution is formally closed under temporal aggregation. Hence, if, say, the daily realized volatility is IG distributed, the weekly realized volatility, defined by the summation of the daily realized volatilities within the week, will also be IG distributed. The empirical evidence for the realized exchange rate volatility series in the next section strongly supports the IG distribution in Equation (4.11).

Now, combining the distributional assumptions in (3.2) and (4.11), the implied *unconditional* distribution for the returns should be a normal inverse Gaussian (NIG),

$$f\left(r_t\right) = \int_0^\infty f\left(r_t|RV_t\right) f\left(RV_t\right) dRV_t \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, \sigma^2\right), \tag{4.14}$$

[3] We use the notation $z \sim IG_{\sigma^2}\left(\sigma^2, \overline{\alpha}\right)$ for this parameterization. Note that $cz \sim IG_{\sigma^2}\left(c^2\sigma^2, \overline{\alpha}\right)$ i.e. the parameter $\overline{\alpha}$ does not change during scaling.

where the density of $NIG_{\sigma^2}\left(\overline{\alpha},0,0,\sigma^2\right)$ is given in (2.4).

The NIG distribution may be viewed as a special case of the Generalized Hyperbolic Distribution in Barndorff-Nielsen (1978). It was first used for modeling speculative returns in Barndorff-Nielsen (1997).

Another distribution, that is used in order to reproduce the fat tails of the return distributions is the Student's t distribution. When using the t-distribution for the returns, we implicitly assume, that the latent variance is inverse gamma (IGamma$\left(\sigma^2,\nu\right)$) distributed. In deriving the GARCH-t model, Bollerslev (1987) assumed the latent mixing variable to be inverse gamma. As an alternative hypothesis, we will assume that the realized volatility is inverse gamma distributed, that is

$$f\left(RV\right)_t \sim IGamma\left(\sigma^2,\nu\right), \tag{4.15}$$

where the density can be expressed as

$$IGamma\left(\sigma^2,\nu\right) = \Gamma\left(\frac{\nu}{2}\right)^{-1}\left(\left(\frac{\nu}{2}-1\right)\sigma^2\right)^{\nu/2}(y)^{-1-\frac{\nu}{2}}\exp\left(\left(1-\frac{\nu}{2}\right)\sigma^2 y^{-1}\right).$$

By combining (3.2) and (4.15) , we find that the unconditional distribution for the returns will be Student's t distributed

$$f\left(r_t\right) = \int_0^\infty f\left(r_t|RV_t\right) f_{IGamma}\left(RV_t\right) dRV_t \sim t\left(\sigma^2,\nu\right),$$

with the following density function

$$g_t = \Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}\left(\left(\nu-2\right)\pi\sigma^2\right)^{-1/2}\left(1+x^2\sigma^{-2}\left(\nu-2\right)^{-1}\right)^{-(\nu+1)/2}. \tag{4.16}$$

The Student's t distribution have fatter tails than the normal distribution, which is a as desireed when we are looking for distributions with which to model returns.

By explicitly treating the RV_t mixing variable in Equations (3.2), (4.11) and (4.15) as an observable, the present empirical approach allows for direct verification of the (implicitly) underlying distributional assumptions of the distribution of the volatility.

Although the NIG distribution in Equation (4.14) or the Student's t distribution in (4.16) may adequately capture the fat tailed unconditional return distributions, it does not account for the well-documented volatility clustering or ARCH effects. In order to incorporate conditional heteroscedasticity in the returns process within the MDH framework, we use the two information sets, I_{t-1} and F_{t-1} as defined above. The standard ARCH class of models then essentially entails the joint specification of a conditional density for r_t given

I_{t-1} along with a parametric expression for the conditional variance. In the conditional variance equation, these models use the lagged squared returns as proxies for the lagged variance. That is

$$\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2, \tag{4.17}$$

where σ_t^2 denotes the conditional variance for time t, and r_{t-1}^2 is the lagged squared return.[4] Andersen and Bollerslev (1998) showed that r_t^2 is a very noisy estimator of the true latent variance. They also showed that the realized volatility is a better estimator of the true latent volatility than r_t^2. So, when available, one should use the realized volatility as a proxy for the volatility, i.e., we should use the information set F_{t-1} as the conditioning set. When we have access to intraday data, and therefore can calculate the realized volatility, we should use the lagged realized volatilities in the variance equation. Recently, Barndorff-Nielsen and Shephard (2001b) showed, that the realized volatility has a measurement error. However, using the arguments outlined above and guided by the results in Chapter 3, we will use the realized volatility as an empirical proxy for the true (latent) volatility.[5]

Next, we will examine the fit of the inverse Gaussian and the inverse gamma distribution to the realized volatility. Building on the results in Chapter 3, we will explicitly treat the realized volatility as a proxy for the true (latent) variance. We will examine the fit of these two alternative distributions for the realized volatility unconditionally and conditionally upon the two information sets F_{t-1} and I_{t-1}. We will use the information in the two information sets to standardize the realized volatility and thereby examine the fit of the distributions once we have taken the serial dependence in the realized volatility into account.

4.3 Unconditional distributions

Results from the estimation of the inverse gamma and inverse Gaussian in (4.12) and the inverse gamma in (4.15) on the realized volatility are given in Table 4.1. The σ^2 parameter represents the mean of both the distributions, and the estimates are 0.427 and 0.457 respectively. The log likelihood values indicate that the inverse Gaussian distribution gives a much better fit than the inverse gamma distribution. Figure 4.1 shows a QQ-plot of the probability integral transform (PIT) of the data, assuming them to be IG$(0.427, 2.028)$

[4] For simplicity, we use the (1,1) case here. The more general (p,q) parameterization is given in (1.3).

[5] Again, one can say that we use the realized volatility as an estimate of the quadratic variation and we assume that the error is negligible.

and with a reference line of 45 degrees.[6] Similarly, Figure 4.2 shows the PIT of the data, assuming them to be IGamma$(0.457, 5.344)$.

By the MDH argument, the distributions for the unconditional returns in Equation (3.2) and the realized volatilities in Equation (4.19) imply that the returns should be unconditionally NIG distributed as in Equation (4.14). The QQ-plot in Figure 4.3, does indeed indicate that the NIG distribution gives a very good fit to the daily raw returns. The corresponding maximum likelihood estimates for the NIG parameters are reported in the second column in Table 4.2. Meanwhile, the distribution for the unconditional returns in (3.2) and the assumption that the realized volatilities follows an inverse gamma distribution, as in (4.15) implies that the returns should be unconditionally t-distributed. Figure 4.4 shows a plot of the PIT of the fitted t-distribution against quantiles of the $U(0,1)$ distribution. This plot indicates a good fit of the t-distribution to the returns.

Using only visual inspection of Figures 3 and 4, it might be hard to evaluate the fit of the two distributions. However, the results in Table 4.2 indicate that the inverse Gaussian distribution gives a better fit to the realized volatility and that for the returns, the corresponding NIG distribution, gives the better fit.

These results further emphasise the validity of the NIG in characterizing the unconditional distribution of speculative returns. However, the unconditional NIG distribution obviously does not account for the well-documented volatility clustering phenomena. In order to do so, we now turn to a discussion of the conditional return and volatility distributions.

4.4 Conditional distributions

The significant volatility clustering effects in the squared returns and the lack of any serial correlation in the squared standardized returns, is further manifested in the strong temporal dependencies in the realized volatility series. For instance, the Ljung-Box Q(1) statistic for test of first order autocorrelation in RV_t is equal to 667.9.

4.4.1 Distributions conditional on lagged realized volatilities

To take the serial correlation of the RV_t into account when fitting the inverse Gaussian and the inverse gamma distribution, we parameterize the mean of the two distributions according to a GARCH structure, where we use lagged

[6] As noted in Chapter 3, the PIT is defined as $z_t^{PIT} = \int_{-\infty}^{x_t} f(u)\,du$. If $f(u)$ is the correct distribution, then $z_t^{PIT} \sim U(0,1)$.

RV_t as the explanatory variable in the conditional mean equation. Thereby, we obtain the distribution of the standardized RV_t, that is

$$RV_t^* = \frac{RV_t}{\widehat{\sigma}_t^2}, \tag{4.18}$$

where RV_t^* is the standardized realized volatility, and $\widehat{\sigma}_t^2$ is the estimated conditional mean of the RV_t, given the two assumptions for the distributions.

To make the discussion more precise, when assuming $RV_t|F_{t-1}$ to follow the inverse Gaussian distribution, we estimated the model

$$RV_t|F_{t-1} \sim IG_{\sigma^2}\left(\sigma_t^2, \overline{\alpha}\right), \tag{4.19}$$

where the conditional mean of the IG is time varying following a GARCH-like structure

$$\sigma_t^2 = \rho_0 + \Sigma_{i=1}^q \rho_i RV_{t-i} + \Sigma_{j=1}^p \pi_j \sigma_{t-j}^2. \tag{4.20}$$

The standardized realized volatility in (4.18) should then be $IG_{\sigma^2}(1, \overline{\alpha})$. Let us denote the model we get from combining (4.19) and (4.20) the RV-GARCH(p,q)-IG model.

Similarly, when assuming that the realized volatility follows and inverse gamma distribution, we fit the model

$$RV_t|F_{t-1} \sim IGamma\left(\sigma_t^2, \nu\right), \tag{4.21}$$

where σ_t^2 follows (4.20). Let us denote the model that is the result of combining (4.20) and (4.21) the RV-GARCH(p,q)-inverse gamma.

Using the two models, the standardized realized volatilities should be inverse gamma$(1, \nu)$ and inverse Gaussian$(1, \overline{\alpha})$, respectively.

Estimates of the parameters of the two models are reported in Table 4.3.[7] We see that the log likelihood for the RV-GARCH(1,1)-IG model is 761.7, which is higher than the value of 710.20 obtained for the RV-GARCH(1,1)-inverse gamma, indicating that when we take the serial dependence of the realized volatility into account, using lagged realized volatilities, the inverse Gaussian distribution gives a better fit. The Q-statistics for the standardized realized volatilities (the residuals) in the lower part of Table 4.3, indicates that the RV-GARCH(1,1)-IG also does a good job of removing the serial dependence in the realized volatility.

The standardized volatilities from the two models should follow an inverse gamma$(1, \nu)$ and $IG(1, \overline{\alpha})$ distribution, respectively. Figure 4.6 displays a QQ-plot of the PIT of the RV-GARCH(1,1)-IG standardized RV. Figure 4.7

[7] Throughout this chapter, we use the (1,1) specification of the models. For all the models estimated, the parameters for higher orders of the lags were not significant.

presents a QQ-plot of the PIT of the RV-GARCH(1,1)-inverse gamma standardized RV. Both the plots indicate a good fit.

We have also estimated the corresponding models for the returns. Recall that, by the MDH argument, when we assume that the conditional variance follows an inverse Gaussian distribution, the returns follows a NIG distribution. And when we assume that the conditional variance follows an inverse gamma distribution, the returns follow a Student's t distriution. Therefore, by combining (3.2), with the different assumptions for the conditional distribution for RV_t as in (4.19) and (4.21) with the conditional variance equation in (4.17), we obtain two models for the conditional returns. These models use lagged realized volatilities in the conditional variance equation.

The model for the returns, that we get from (3.2), (4.19) and (4.20) can be written

$$(r_t - \mu)\,|F_{t-1} \sim NIG\left(\overline{\alpha}, 0, 0, \sigma_t^2\right), \tag{4.22}$$

where σ_t^2 follows (4.20).[8] We will refer to this model as the RV-GARCH-NIG model.

Similarly to the above, we will get the RV-GARCH-t model by combining (3.2), (4.21) and (4.20)

$$(r_t - \mu)\,|F_{t-1} \sim t\left(\sigma_t^2, \nu\right), \tag{4.23}$$

where σ_t^2 follows (4.20). We also estimate a GARCH model with a normal distribution, that is

$$(r_t - \mu)\,|F_{t-1} \sim N\left(0, \sigma_t^2\right), \tag{4.24}$$

where σ_t^2 follows (4.20).

These models were fitted to the daily returns of the ECU/USD 1989 - 1998 dataset and the results are reported in Table 4.4. As seen in Table 4.5, AIC and BIC criteria are minimized for the RV-GARCH(1,1)-NIG model.[9] Figure 4.8 shows the QQ-plot of the PIT of the standardized residuals from the RV-GARCH(1,1)-NIG model. Figures 4.9 and 4.10 show the corresponding plots using the RV-GARCH(1,1)-N and RV-GARCH(1,1)-t models respectively. According to Figure 4.9 and 4.11, the NIG and the t-distribution give an equally good fit to the standardized residuals.

[8] For the ECU/USD 1989 - 1998 dataset, there are no serial dependencies in the conditional mean. Therefore, for all models throughout this chapter, we use a constant for the conditional mean and we demean the data and fit the models on the demeaned data. The estimation of the constant and the equation for the conditional variance are done simultaneously.

[9] We will use the log likelihood values, together with the information criteria AIC and BIC to evaluate the models, as outlined in Bollerslev and Mikkelsen (1999).

4.4.2 Distributions conditional on lagged squared returns

Above we used the lagged realized volatilities to model the conditional mean of the inverse gamma and the inverse Gaussian distribution, and to model the conditional variance of the NIG, t and normal distribution. Although intraday data is becoming increasingly available, still most data concerning assets, and undoubtedly most historical data are comprised of daily data. Therefore, we repeat the above analysis using the information set I_{t-1}, i.e., the lagged squared returns.

For the inverse Gaussian assumption we have

$$(RV_t|I_{t-1}) \sim IG\left(\sigma_t^2, \overline{\alpha}\right).$$

where σ_t^2 follows (4.17). We refer to this model as the GARCH(1,1)-IG model.

When we assume the realized volatility, conditional on lagged squared returns, to be inverse gamma, we have

$$f\left(RV_t|I_{t-1}\right) \sim IGamma\left(\sigma_t^2, \nu\right), \tag{4.25}$$

where σ_t^2 follows (4.17). We refer to this model as the GARCH(1,1)-inverse gamma model.

We use these models to standardize the realized volatilities, while taking into account the serial dependence in the realized volatility as in (4.18).

Again, we also estimate the corresponding models for the returns, assuming the realized volatilities conditional on the lagged squared returns to be inverse Gaussian, which leads to

$$\left(r_t - \mu\right)|I_{t-1} \sim NIG\left(\overline{\alpha}, 0, 0, \sigma_t^2\right),$$

i.e., the GARCH-NIG model of Barndorff-Nielsen (1997), Andersson (2001) and Jensen and Lunde (2001).

For the assumption that realized volatility, given lagged squared returns, follows an inverse gamma distribution, the conditional distribution of the returns will be

$$\left(r_t - \mu\right)|I_{t-1} \sim t\left(\sigma_t^2, \nu\right),$$

which is the GARCH-t model of Bollerslev (1987). Additionally, we estimate the standard GARCH-normal model, that is

$$\left(r_t - \mu\right)|I_{t-1} \sim N\left(0, \sigma_t^2\right),$$

where σ_t^2 follows (4.17).

First, we report the estimates of the GARCH-IG and the GARCH-inverse gamma models, presented in Table 4.5. The log likelihood values for the two models show that the GARCH-IG model gives a better fit than the GARCH-inverse gamma model. The log likelihood values in Table 4.5 indicate a slightly better fit of the inverse Gaussian distribution, however, this is hard to see in the QQ-plots. Figure 4.11 shows the PIT of the realized volatilities standardized by the RV-GARCH(1,1)-IG model against the quantiles of the $U(0.1)$ distribution. Similarly, Figure 4.12 shows the PIT of the realized volatilities standardized by the RV-GARCH(1,1)-inverse gamma model against the quantiles of the $U(0.1)$.

In Table 4.6, the results of the estimation of the corresponding models for the returns, i.e., the "usual" GARCH models, are reported . The information criteria AIC and BIC are minimized for the GARCH(1,1)-NIG model, but the values for the GARCH(1,1)-t model are very close. This, again, indicates that the inverse Gaussian gives a slightly better fit to the realized volatility.

4.4.3 Out-of-sample Euro predictions

So far, we have fitted the different models and distributions using the ECU/USD 1989 - 1998 data. Up until 1998, the ECU was a basket of currencies. In the beginning of 1999 it started to trade as a currency. An interesting question is, whether we can use the information in the ECU/USD 1989 - 1998 data to choose the most appropriate model for the Euro/USD 1999 - 2001 data. We will investigate this by using the estimates of the ECU period to make predictions for the Euro period.

To start out, we use the unconditional distributions estimated on the daily ECU/USD 1989 - 1998 data, that is, the normal, NIG and Student's t to investigate the fit of these distributions on the 1999 - 2001 Euro/USD data. Table 4.7 shows the log likelihood, AIC and BIC values for the unconditional normal, NIG and Student's t distribution for the Euro/USD 1999 - 2001 data, given the estimates obtained using the ECU/USD 1989 - 1998 data. That is, we fit the normal, NIG and Student's t to the ECU/USD 1989 - 1998 data, and given these estimates (from Table 4. 2), we use the data from the Euro/USD 1999 - 2001 period to calculate the log likelihood, AIC and BIC. The results in Table 4.7 show that the NIG distribution gives the best fit, even out-of-sample.

We do not have the realized volatilities for the Euro/USD 1999 - 2001 data, so we cannot evaluate the fit of the RV-GARCH models for the out-of-sample data. Still, we can evaluate the GARCH model that uses only the lagged squared returns in the conditional variance equation. We use the estimates of the GARCH models (with the NIG, normal and Student's t distribution) obtained on the ECU/USD 1989 - 1998 data (results in Table 4.6) to evaluate the log likelihood, AIC and BIC for the out-of-sample Euro/USD 1999 - 2001 data. These results are reported in Table 4.8. The last four rows of the table report the Ljung-Box statistics for test of serial correlation in the Euro-return residuals based on the model estimates for the ECU.

4.5 Conclusions

In this chapter, we have established a link between the realized volatility and the conditional variance. Using this link, we have shown that the inverse Gaussian distribution gives a good fit to the realized volatility. The inverse Gaussian gives a good fit to the realized volatility both unconditionally, and conditional on two different information sets. This gives direct empirical support to the GARCH-NIG model, which uses the assumption that the conditional variance follows the inverse Gaussian distribution. For comparison, we also tried the inverse gamma distribution for the realized volatility. When the conditional variance is assumed to be inverse gamma distributed, the returns are Student's t distributed. When fitting these distributions to the returns, the NIG distribution gave a better fit than the Student's t distribution both unconditionally, and conditional on the two information sets.

4.6 Further work

It would be of interest to conduct a more comprehensive study of which distribution gives the best fit to the conditional variance using the realized volatility. This study would include more distributions, such as the lognormal, generalized inverse Gaussian (GIG), and all the distributions that are nested in, or special cases of the GIG distribution, such as the reciprocal inverse Gaussian and the positive hyperbolic distributions. This fishing expedition would also include the use of more high frequency datasets, i.e., more datasets where we can calculate realized volatility. It would also be interesting to see if there are any differences between different types of data, such as foreign exchange data, stock data and other equities.

In this thesis we do not deal with the sampling error in the realized volatilty, as highlighted by Barndorff-Nielsen and Shephard, 2001b, 2002b. A natural extension of the work in this chapter is to try assume a (small sample) distribution for the error in the realized volatility, and to incorporate that in the parametric modeling of the realized volatility.

4.7 Tables

Table 4.1:
Distributions fitted to the daily unconditional realized volatility of the ECU/USD 1989 - 1998 dataset.

	Distributions	
	IG($\sigma^2, \overline{\alpha}$)	Inverse gamma(σ^2, ν)
σ^2	0.427	0.457
	(0.006)	(0.010)
$\overline{\alpha}, \nu$	2.028	5.344
	(0.051)	(0.121)
Log lik.	**217.32**	156.54
AIC	**-430.64**	-309.09
BIC	**-419.05**	-297.50

Notes: For both the inverse Gaussian and the inverse Gamma distribution, the parameter σ^2 denotes the mean. The estimation is done by maximum likelihood. Standard errors of the estimates are given in parenthesis.

Table 4.2:
Distributions fitted to the unconditional returns of daily ECU/USD 1989 - 1998.

		Distributions		
	Data	Normal	NIG	Student's t
μ	-	0.002 (0.013)	0.005 (0.011)	0.006 (0.012)
$\overline{\alpha}, \nu$	-	-	1.138 (0.167)	4.630 (0.489)
σ^2	-	0.407 (0.008)	0.407 (0.017)	0.420 (0.024)
Log lik.	-	-2353.957	**-2251.076**	-2252.489
AIC	-	4711.915	**4508.152**	4510.979
BIC	-	4723.504	**4525.536**	4528.363
$Q^2(1)$	34.4 (0.000)	-	-	-
$Q^2(5)$	123.0 (0.000)	-	-	-
$Q^2(10)$	208.5 (0.000)	-	-	-

Notes: The table shows the parameter estimates of the three distributions fitted to the ECU/USD 1989 - 1998 returns. Estimation is done by maximum likelihood. Standard errors of the estimates are given in parenthesis. $Q^2(n)$ denotes the Ljung-Box statistic for serial dependency in the squared standardized residuals up to lag (n). The p-values are given in parenthesis.

Table 4.3:
Models for the daily realized volatility ECU/USD 1989 - 1998 using lagged realized volatilities in the conditional mean equation, that is, models for $RV_t|F_{t-1}$.

		Models	
	Data	RV-GARCH-IG$^{(1)}$	RV-GARCH-IGamma$^{(2)}$
$\overline{\alpha}, \nu$	-	3.4178	8.081
		(0.071)	(0.166)
ρ_0	-	0.025	0.024
		(0.003)	(0.002)
ρ_1	-	0.270	0.273
		(0.014)	(0.012)
π_1	-	0.671	0.681
		(0.016)	(0.012)
Log lik.	-	**761.7**	710.20
AIC	-	**-1515.5**	-1412.4
BIC	-	**-1492.4**	-1389.2
$Q(1)$	667.9	1.857	2.201
	(0.000)	(0.172)	(0.137)
$Q(5)$	2133.9	10.970	11.539
	(0.000)	(0.052)	(0.0417)
$Q(10)$	3147.0	21.165	21.683
	(0.000)	(0.020)	(0.016)

Notes: For both the inverse Gaussian and the inverse Gamma distribution, the parameter σ_t^2 denotes the conditional mean of the Standard errors of the estimates are given in parenthesis. distributions. F_{t-1} denotes the information set consisting of lagged realized volatilities, $F_{t-1}= \sigma\left\{RV_{t-1}, RV_{t-2,}...\right\}$.

(1) RV-GARCH-IG is the model $RV_t|F_{t-1} \sim IG\left(\sigma_t^2, \overline{\alpha}\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$.

(2) RV-GARCH-Inverse Gamma is the model $RV_t|F_{t-1} \sim \text{IGamma}\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$.

The lower part of the table shows the results from the realized volatilities standardized using the different models. $Q(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized residuals for up to lag (n). The p-values are given in parenthesis.

Table 4.4:
RV-GARCH models fitted to the daily ECU/USD 1989 - 1998 using lagged RV in the variance equation, that is, models for $(r_t - \mu)\,|F_{t-1}$.

		Models: RV-GARCH		
	Data	N(1)	NIG(2)	t(3)
μ	-	0.006 (0.012)	0.006 (0.011)	0.006 (0.011)
$\overline{\alpha}, \nu$	-	-	1.500 (0.244)	5.476 (0.654)
ρ_0	-	0.013 (0.005)	0.001 (0.003)	0.000 (-)
ρ_1	-	0.181 (0.023)	0.090 (0.019)	0.0822 (0.0166)
π_1	-	0.778 (0.029)	0.903 (0.022)	0.915 (0.017)
Log lik.	-	-2265.19	**-2198.32**	-2199.13
AIC	-	4538.39	**4406.64**	4408.26
BIC	-	4561.57	**4435.62**	4437.24
$Q^2\,(1)$	34.35 (0.000)	7.758 (0.005)	1.126 (0.288)	0.698 (0.403)
$Q^2\,(5)$	123.0 (0.000)	17.139 (0.004)	3.975 (0.553)	2.531 (0.772)
$Q^2\,(10)$	208.5 (0.000)	24.139 (0.007)	9.416 (0.493)	6.747 (0.749)

Notes: F_{t-1} denotes the information set consisting of lagged realized volatilities, $F_{t-1}= \sigma\left\{RV_{t-1}, RV_{t-2,} ...\right\}$. For all the models, the parameter σ_t^2 denotes the conditional variance. Standard errors of the estimates are given in parenthesis.

(1) RV-GARCH-N is the model $(r_t - \mu)\,|F_{t-1} \sim N\left(0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$.

(2) RV-GARCH-NIG is the model $(r_t - \mu)\,|F_{t-1} \sim NIG\left(\overline{\alpha}, 0, 0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$.

(3) RV-GARCH-N is the model $(r_t - \mu)\,|F_{t-1} \sim t\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$.

$Q\,(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized residuals for up to lag (n). The p-values are given in parenthesis.

Table 4.5:
Models for realized volatility using lagged squared returns in the equation for the conditional mean of the RV_t, that is models for $RV_t|I_{t-1}$.

		Models	
	Data	GARCH-IG$^{(1)}$	GARCH-IGamma$^{(2)}$
$\overline{\alpha}, \nu$	-	2.732	6.856
		(0.062)	(0.145)
ρ_0	-	0.037	0.033
		(0.003)	(0.002)
ρ_1	-	0.077	0.077
		(0.006)	(0.005)
π_1	-	0.841	0.856
		(0.012)	(0.009)
Log lik.	-	**527.5**	492.31
AIC	-	**-1047.01**	-976.62
BIC	-	**-1023.83**	-953.44
$Q(1)$	667.9	333.84	341.21
	(0.000)	(0.000)	(0.000)
$Q(5)$	2133.9	953.06	974.73
	(0.000)	(0.000)	(0.000)
$Q(10)$	3147.0	1355.85	1381.62
	(0.000)	(0.000)	(0.000)

Notes: For both the inverse Gaussian and the inverse Gamma distribution, the parameter σ_t^2 denotes the conditional mean of the distributions. I_{t-1} denotes the information set consisting of lagged squared returns, $I_{t-1} = \sigma\left\{r_{t-1}^2, r_{t-2}^2, ...\right\}$. Standard errors of the estimates are given in parenthesis.
$^{(1)}$ RV-GARCH-IG is the model $RV_t|I_{t-1} \sim IG\left(\sigma_t^2, \overline{\alpha}\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$.
$^{(2)}$ RV-GARCH-Inverse Gamma is the model $RV_t|I_{t-1} \sim IGamma\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$.
The lower part of the table shows the results from the realized volatilities standardized using the different models.
$Q(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized residuals for up to lag (n). The p-values are given in parenthesis.

Table 4.6:
GARCH models fitted to the daily ECU/USD 1989 - 1998 using lagged squared returns in the variance equation, that is, models for $(r_t - \mu)\,|I_{t-1}$.

		Models		
	Data	GARCH-N$^{(1)}$	GARCH-NIG$^{(2)}$	GARCH-t$^{(3)}$
μ	-	0.007 (0.012)	0.004 (0.010)	0.005 (0.011)
$\overline{\alpha}, \nu$	-	-	1.705 (0.287)	5.992 (0.768)
ρ_0	-	0.006 (0.001)	0.004 (0.002)	0.004 (0.002)
ρ_1	-	0.049 (0.006)	0.041 (0.008)	0.041 (0.008)
π_1	-	0.937 (0.007)	0.950 (0.010)	0.951 (0.009)
Log lik	-	-2237.96	**-2177.93**	-2178.71
AIC	-	4483.92	**4365.87**	4367.42
BIC	-	4507.10	**4394.85**	4396.40
$Q^2(1)$	34.4 (0.000)	0.044 (0.832)	0.324 (0.569)	0.332 (0.564)
$Q^2(5)$	123.0 (0.000)	3.164 (0.674)	2.350 (0.798)	2.400 (0.791)
$Q^2(10)$	208.5 (0.000)	15.78 (0.106)	4.964 (0.894)	4.993 (0.892)

Notes: I_{t-1} denotes the information set consisting of lagged squared returns, $I_{t-1} = \sigma\left\{r_{t-1}^2, r_{t-1}^2, ...\right\}$. For all the models the parameter σ_t^2 denotes the conditional variance. Standard errors of the estimates are given in parenthesis.

(1) GARCH-N is the model $(r_t - \mu)\,|I_{t-1} \sim N\left(0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$.

(2) GARCH-NIG is the model $(r_t - \mu)\,|I_{t-1} \sim NIG\left(\overline{\alpha}, 0, 0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$.

(3) GARCH-t is the model $(r_t - \mu)\,|I_{t-1} \sim t\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$.

$Q(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized residuals for up to lag (n). The p-values are given in parenthesis.

Table 4.7:
Information criteria for the out-of-sample data i.e., the Euro/USD 1999 - 2001 data, given the estimates obtained using the ECU/USD 1989 - 1998 data.

	Distributions		
	Normal	NIG	Student's t
Log lik.	-693.80	**-648.42**	-653.58
AIC	1391.60	**1302.85**	1313.17
BIC	1406.00	**1324.45**	1334.77

Notes: The distributions were fitted to the ECU/USD 1989 - 1998 data. We assume that the same distributions hold for the out-of-sample data, that is, for the Euro/USD 1999 - 2001 data. The table shows the information criteria of the Euro/USD 1999 - 2001 data, using the estimates from the ECU/USD data from 1989 - 1998 period.

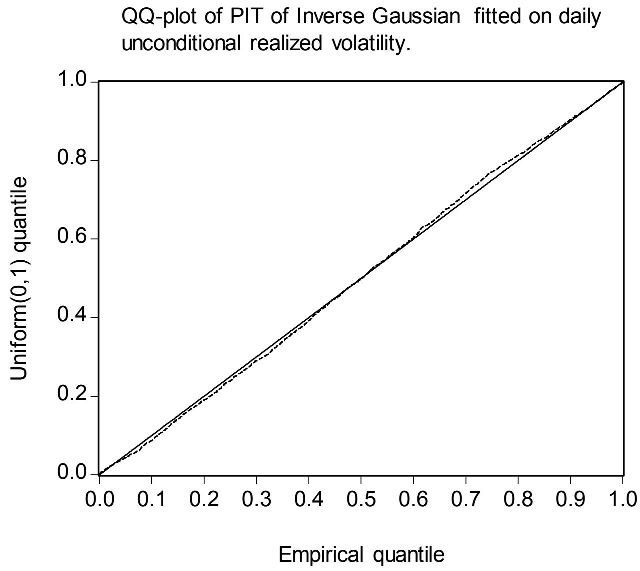
Table 4.8:
Information criteria from the out-of-sample period (Euro/USD 1999 - 2001), given the estimates from the in-sample period (ECU/USD 1989 - 1998).

		Models		
	Data	GARCH-N$^{(1)}$	GARCH-NIG$^{(2)}$	GARCH-t$^{(3)}$
Log lik.	-	-679.21	**-640.45**	-643.65
AIC	-	1366.42	**1290.91**	1297.31
BIC	-	1395.22	**1326.92**	1333.32
$Q^2\,(1)$	8.109	0.465	0.659	0.668
	(0.004)	(0.495)	(0.416)	(0.413)
$Q^2\,(5)$	10.584	0.955	11.467	11.478
	(0.060)	(0.044)	(0.042)	(0.042)
$Q^2\,(10)$	35.466	17.884	17.990	17.982
	(0.000)	(0.057)	(0.055)	(0.055)

Notes: The models were fitted using the ECU/USD 1989 - 1998 data. We assume that the same models hold for the out-of-sample data, that is, for the Euro/USD 1999 - 2001 data. The table shows the information criteria of the Euro/USD 1999 - 2001 data, using the estimates from the ECU/USD 1989 - 1998 period.
$Q^2\,(n)$ denotes the Ljung-Box statistic for serial correlation in the squared standardized residuals up to lag (n). The p-values are given in parenthesis

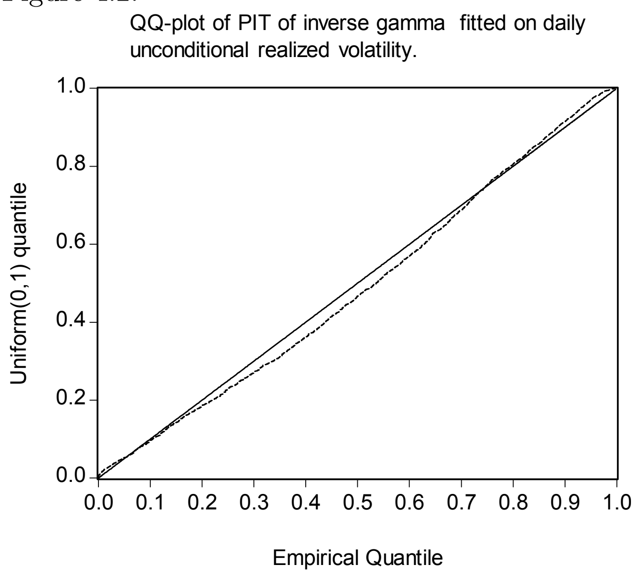
4.8 Figures

Figure 4.1:



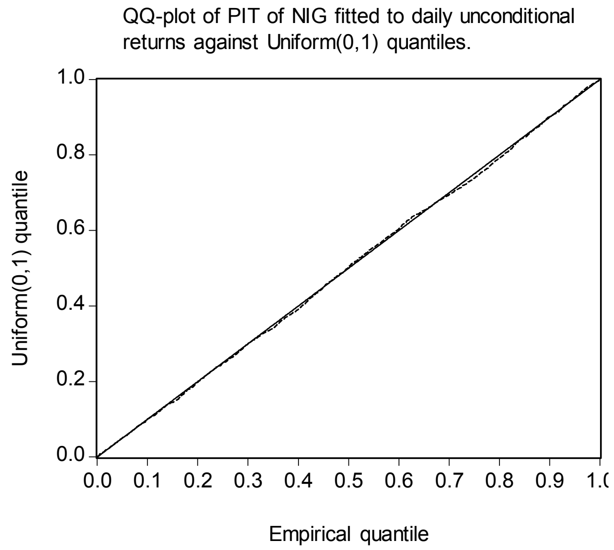
Notes: The figure shows a QQ-plot of the PIT for the unconditional daily realized volatility for the ECU/USD 1989 - 1998 assumed to be $IG\,(0.427, 2.028)$ against the quantiles of the U(0,1) distribution.

Figure 4.2:



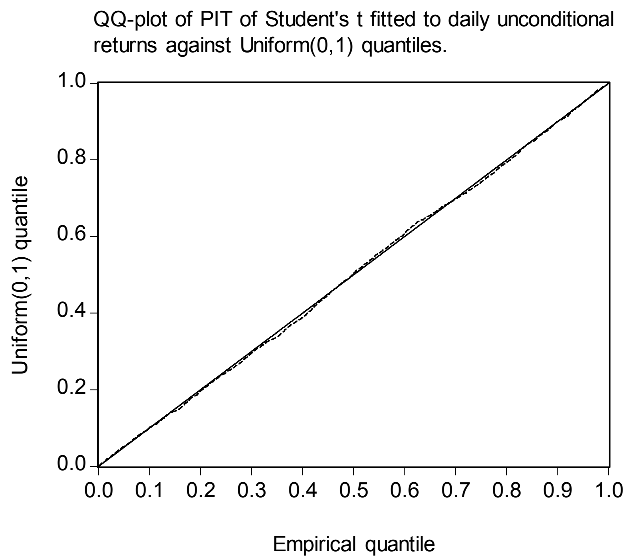
Notes: The figure shows a QQ-plot of the PIT for the unconditional daily realized volatility for the ECU/USD 1989 - 1998 assumed to be $IGamma\,(0.457, 5.344)$ against the quantiles of the U(0,1) distribution.

Figure 4.3:



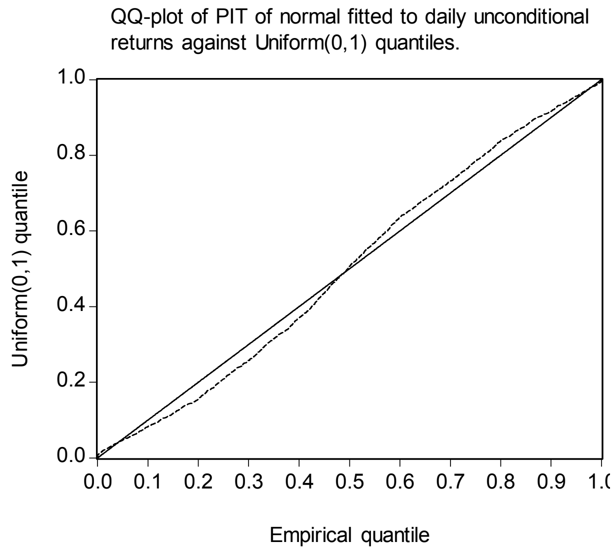
Notes: The figure shows a QQ-plot of the PIT of the daily demeaned returns for the ECU/USD 1989 - 1998 data assumed to be NIG$(1.138, 0, 0, 0.407)$ against the quantiles of the U(0,1) distribution.

Figure 4.4:



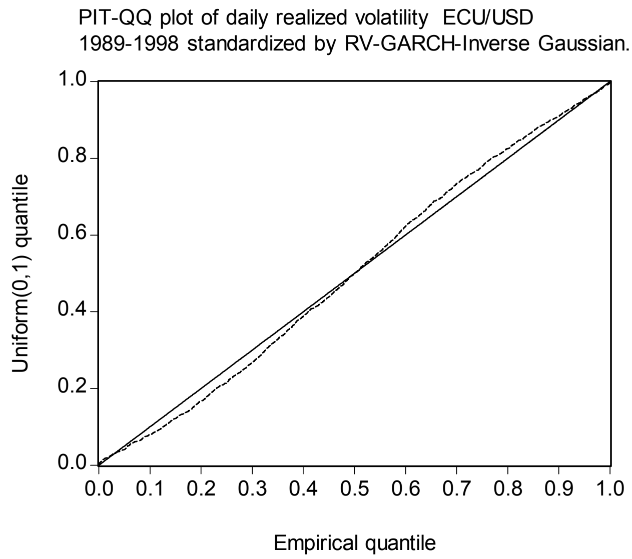
Notes: The figure shows a QQ-plot of the PIT of the daily demeaned returns for the ECU/USD 1989 - 1998 data assumed to be $t\,(0.420, 3.630)$ against the quantiles of the U(0,1) distribution.

Figure 4.5:



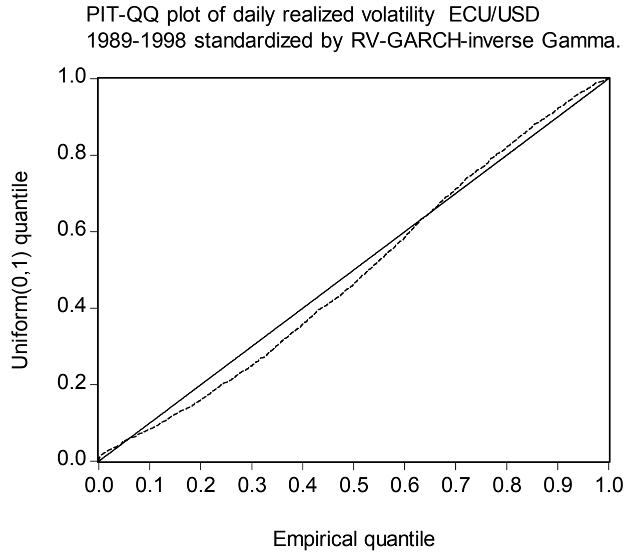
Notes: The figure shows a QQ-plot of the PIT of the daily demeaned returns for the ECU/USD 1989 - 1998 data assumed to be $\mathrm{N}(0, 0.407)$ against the quantiles of the U(0,1) distribution.

Figure 4.6:



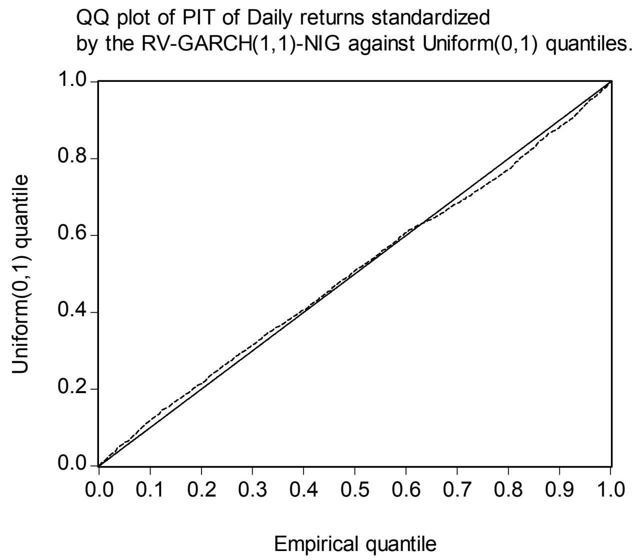
Notes: The figure shows a QQ-plot of PIT of Daily RV of ECU/USD standardized by the RV-GARCH(1,1)-IG, that is $RV_t|F_{t-1} \sim IG_{\sigma^2}\left(\sigma_t^2, \overline{\alpha}\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U(0,1)$ distribution. The standardized RV_t is assumed to be $IG_{\sigma^2}(1, 3.418)$.

Figure 4.7:



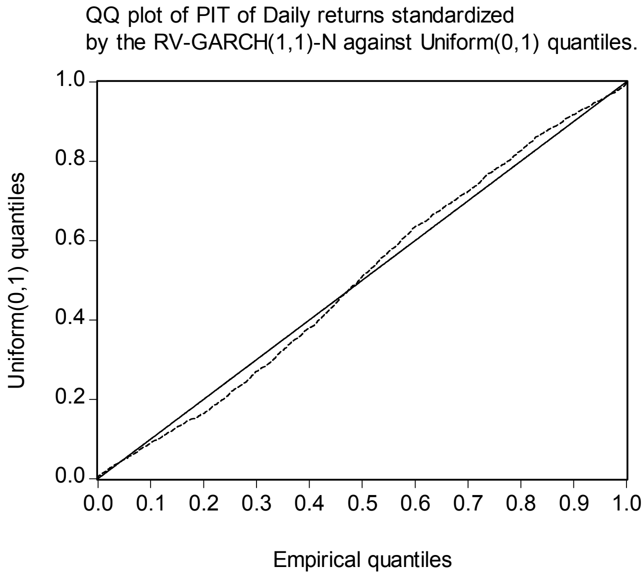
Notes: The figure shows a QQ-plot of PIT of Daily RV of ECU/USD standardized by the RV-GARCH(1,1)-IGamma, that is $RV_t|F_{t-1} \sim IGamma\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\left(0,1\right)$ distribution. The standardized RV_t is assumed to be IGamma$(1, 8.081)$.

Figure 4.8:



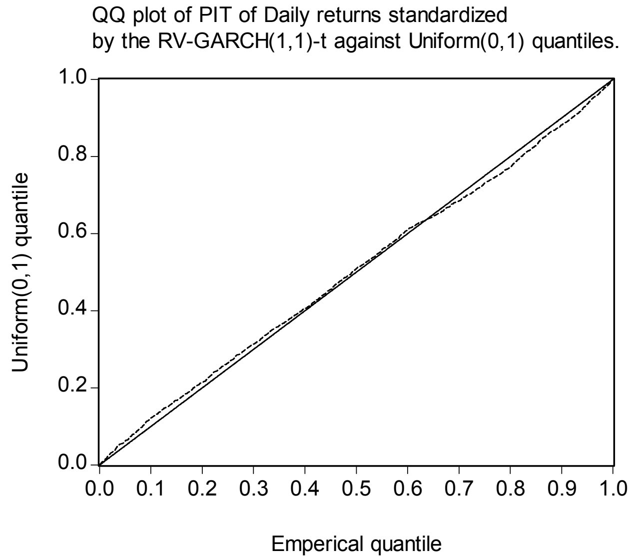
Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the RV-GARCH(1,1)-NIG model, that is $\left(r_t - \mu\right)|F_{t-1} \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, \sigma_t^2\right)$ where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\left(0,1\right)$ distribution. The standardized residuals are assumed to be $NIG_{\sigma^2}\left(1.5, 0, 0, 1\right)$.

Figure 4.9:



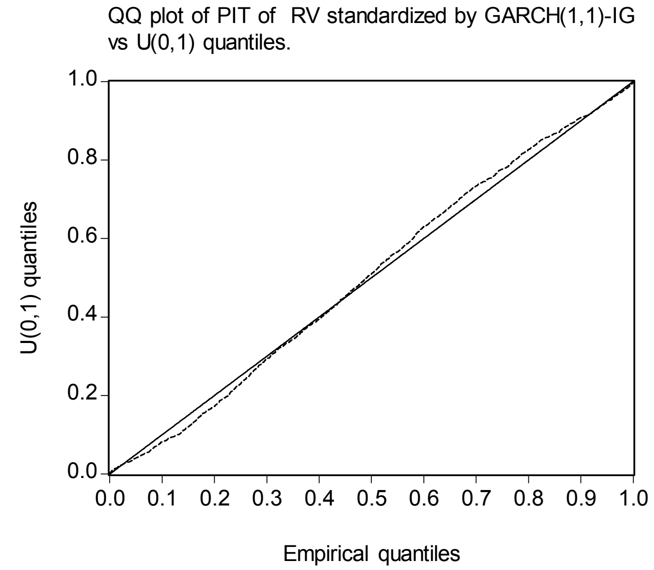
Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the RV-GARCH(1,1)-N model, that is, $(r_t - \mu)\,|F_{t-1} \sim N\left(0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\,(0,1)$ distribution. The standardized residuals are assumed to be $N\,(0,1)$.

Figure 4.10:



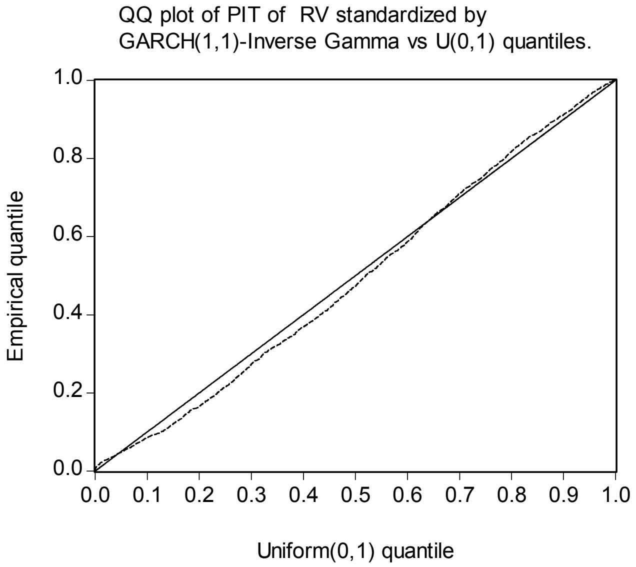
Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the RV-GARCH(1,1)-t model, that is, $(r_t - \mu)\,|F_{t-1} \sim \mathrm{t}\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 RV_{t-1} + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\,(0,1)$ distribution. The standardized residuals are assumed to be $t\,(1, 5.476)$.

Figure 4.11:



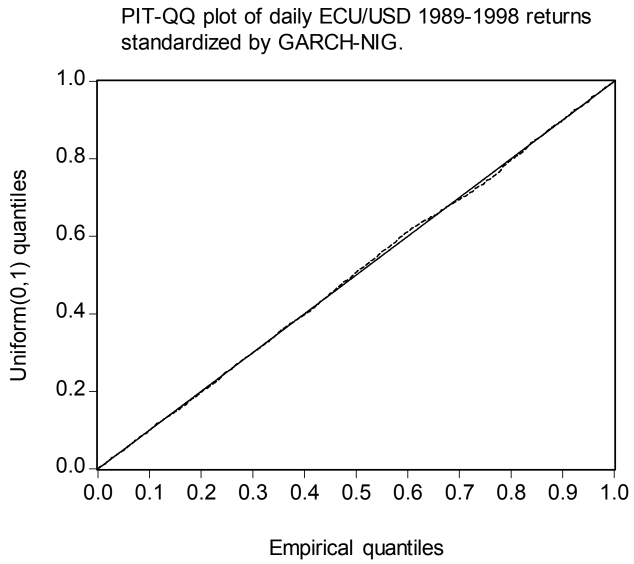
Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 realized volatility standardized by the GARCH(1,1)-IG model, that is, $RV_t|I_{t-1} \sim IG\left(\sigma_t^2, \overline{\alpha}\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\left(0,1\right)$ distribution. The standardized residuals are assumed to be $IG_{\sigma^2}\left(1, 2.732\right)$.

Figure 4.12:



Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 realized volatility standardized by the GARCH(1,1)-IG model, that is, $RV_t|I_{t-1} \sim IGamma\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\left(0,1\right)$ distribution. The standardized residuals are assumed to be $IGamma\left(1, 6.856\right)$.

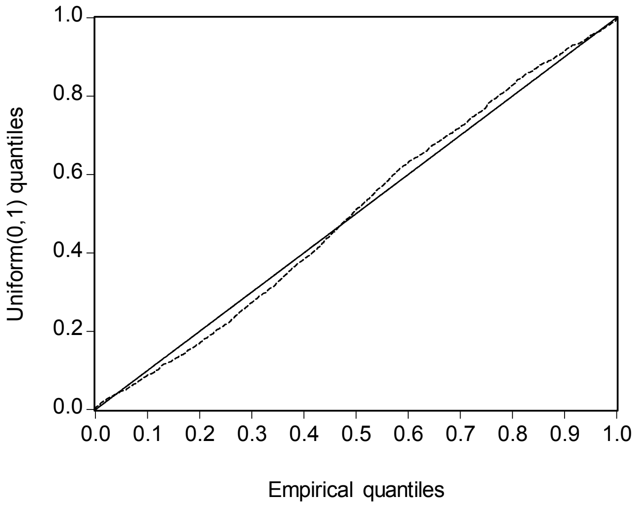
Figure 4.13:



Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the GARCH(1,1)-NIG model, that is, $(r_t - \mu)\,|I_{t-1} \sim NIG\left(\overline{\alpha}, 0, 0, \sigma_t^2\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\,(0,1)$ distribution. The standardized residuals are assumed to be $NIG_{\sigma^2}\,(1.705, 0, 0, 1)$.

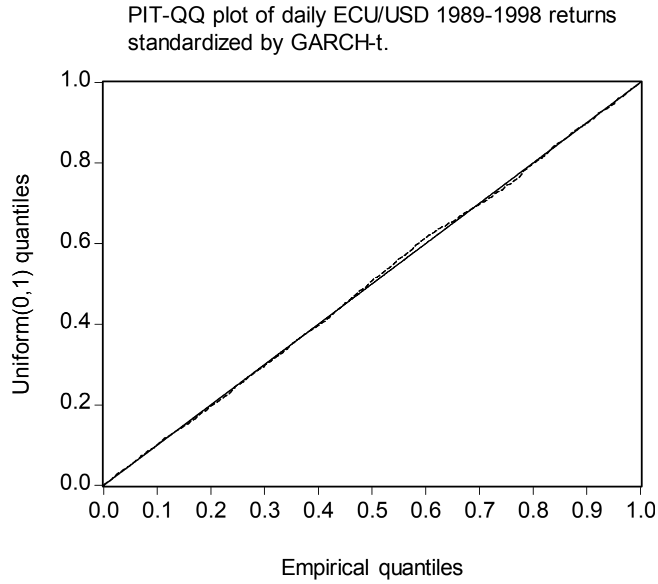
Figure 4.14:

PIT-QQ plot of daily ECU/USD 1989-1998 returns standardized by GARCH-N.



Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the GARCH(1,1)-N model, that is, $(r_t - \mu)\,|I_{t-1} \sim N\left(0, \sigma_t^2\right)$ where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\,(0,1)$ distribution. The standardized residuals are assumed to be $N\,(0,1)$.

Figure 4.15:



Notes: The figure shows a QQ-plot PIT of the ECU/USD 1989 - 1998 returns standardized by the GARCH(1,1)-model, that is, $(r_t - \mu)\,|I_{t-1} \sim t\left(\sigma_t^2, \nu\right)$, where $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$, against the quantiles of the $U\,(0,1)$ distribution. The standardized residuals are assumed to be $t\,(1, 5.992)$.

Chapter 5

Temporal aggregation of realized volatility and the inverse Gaussian distribution

In this chapter we use another approach to investigate the assumption that the realized volatility is well described by the inverse Gaussian distribution. We look at the aggregation properties of the realized volatility and the inverse Gaussian distribution. Recall that the daily realized volatility is the sum of the squared intraday returns. This implies, that the lower frequencies can be calculated as the sum of the higher frequencies, that is, for example, the weekly realized volatility is the sum of five daily realized volatilities.

We examine the ability of the inverse Gaussian distribution to capture the aggregation properties of the standardized realized volatility as in (4.18). However, here we use a slightly different parameterization of the conditional mean of the inverse Gaussian, than what we used in Chapter 4.

The inverse Gaussian distribution is closed under aggregation. That is, the sum of independent inverse Gaussian variables follows an inverse Gaussian distribution. In our parameterization, we have the following aggregation rules[1]: let

$$\begin{aligned} Z_i &\sim IG\left(\sigma^2, \overline{\alpha}\right), \\ E\left(Z_i\right) &= \sigma^2. \end{aligned}$$

If

$$Y = \Sigma_{i=1}^{N} Z_i,$$

then

$$Y \sim IG\left(N\sigma^2, N\overline{\alpha}\right), \tag{5.1}$$

[1] Here we present the aggregation results when we aggregate variates from the same (inverse Gaussian) distribution. For a more general presentation of the aggregation properties of the inverse Gaussian distribution, see Barndorff-Nielsen (1978).

where

$$E(Y) = N\sigma^2.$$

When we use the IG distribution to model the realized volatility, this means that if the daily variance is distributed according to $V_{Daily} \sim IG\left(\sigma^2_{Daily}, \overline{\alpha}_{Daily}\right)$, with expected value $E(V_{Daily}) = \sigma^2_{Daily}$. The weekly variance is then $V_{Weekly} \sim IG\left(5\sigma^2_{Daily}, 5\overline{\alpha}_{Daily}\right)$. Using the parameter estimates from the daily frequency, we can calculate the parameter values for the lower frequencies using these aggregation formulas. Below, we will analyze the (aggregation) implied moments for lower frequencies using the parameter estimates from the higher frequencies.

The mean and the variance of Y grow linearly with the number of variables aggregated, whereas the skewness and kurtosis converge to zero and three respectively. For the aggregated variables, the Central Limit Theorem kicks in and the sum will be normally distributed, i.e., the sum will have skewness zero and kurtosis three.

5.1 Fitting IG to realized volatility

Using the fact that the IG is closed under aggregation, we evaluate the distribution's ability to recreate the moments of the lower frequencies. If the IG is a good distribution for the realized volatility, using the aggregation results in (5.1), we can use the daily data to calculate the implied moments for the lower frequencies, such as the weekly, bi-weekly and monthly. The aggregation properties of the IG are valid only under i.i.d. observations, and, clearly, there is serial dependence in the realized volatility, as documented in Chapters 3 and 4. This we need to take into account when fitting the inverse Gaussian distribution to the realized volatility.

5.2 Fitting IG to standardized RV

One way of verifying (4.19) is first to fit the inverse Gaussian distribution to the daily realized volatility and use (5.1) to calculate the implied parameters for the lower frequencies, and then compare these "implied" moments with the sample moments and the parametric moments calculated by the parameter estimates from estimation of that lower frequency.

To take the serial correlation into account, we model the conditional mean of the RV using F_{t-1}, the information set consisting of lagged realized volatilities. We let the mean σ^2_t of IG be time varying and we assume that it follows an ARMA(1,1) like structure with lagged RV_t as explanatory variables. We assume that the daily RV_t is *conditionally IG distributed*, that is

$$RV_{t,\text{ Daily}}|F_{t-1} \sim IG_{\sigma^2}\left(\sigma_t^2, \overline{\alpha}\right),$$

where

$$\sigma_t^2 = \phi_0 + (\phi_1 + \theta_1) RV_{t-1,\text{Daily}} - \theta_1 \sigma_{t-1}^2. \tag{5.2}$$

and $RV_{t,\text{ Daily}}$ refers to the realized volatility at the daily frequency at time t. We will refer to this model as the ARMA(1,1)-IG model. We recreate the sequence of the conditional mean σ_t^2 and standardize the raw RV by $RV^*_{t,\text{ Daily}} = \frac{RV_t}{\widehat{\sigma}_t^2}$, where $\widehat{\sigma}_t^2$ is the estimated conditional mean, and $RV^*_{t,\text{ Daily}}$ is the daily RV standardized by the ARMA(1,1)-IG procedure, now distributed according to $RV^*_{t,\text{ Daily}} \sim IG_{\sigma^2}(1, \overline{\alpha})$.

The lower RV-standardized frequencies are derived by aggregating the daily standardized volatility. That is, the weekly standardized RV is actually the sum of five $RV^*_{t,\text{ Daily}}$, etc.

The ARMA(1,1)-IG standardization procedure is as follows:

- Fit ARMA(1,1)-IG where
$$\sigma_t^2 = \phi_0 + (\phi_1 + \theta_1) RV_{t-1} - \theta_1 \sigma_{t-1}^2. \tag{5.3}$$
- Standardize $RV_t^* = \frac{RV_t}{\sigma_t^2}$.
- RV_t^* is the standardized RV, $RV^*_{t,\text{ Daily}} \sim IG_{\sigma^2}(1, \overline{\alpha})$.

Having standardized the data, we can now fit the inverse Gaussian to RV_t^* and the temporal aggregation results in (5.1) should be valid.

Recall that the *mean* in the IG distribution is σ^2, and the *variance* in the NIG is σ^2, the same parameter. In the ARMA(1,1)-IG model, we model the conditional mean of the variance as in (5.2). In the GARCH-NIG we model the conditional variance of the returns, which has the time varying structure $\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \pi_1 \sigma_{t-1}^2$. So, using the ARMA(1,1)-IG, the standardization of the RV is very similar to when we model the conditional variance of the daily returns using the GARCH-NIG model.

Results of the standardization

The results of the standardization are given in Tables 5.1, 5.2 and 5.3 and in Figure 5.2. Table 5.1 shows the parameter estimates from the ARMA(1,1)-IG model. The AR-parameter ϕ is above 0.9, indicating that there is strong autocorrelation in the data. The plot of the sample autocorrelation function in Figure 5.2, shows that we have accounted for most of the serial dependence in the standardized RV.

The standardized RVs are used later to evaluate the aggregation properties of the IG, so it is important that these residuals are free from temporal dependence. Ljung-Box Q-statistics of the raw and standardized RV are reported in Table 5.2. The Q-statistics for test of serial correlation up to the first and fifth lag are not significant, but the 10th and 20th lag are significant.

The standardized weekly, bi-weekly and monthly RV, are obtained from the daily standardized RV by aggregation. Table 5.3 reports the descriptives of the standardized RV for the different frequencies. For the weekly frequency, the Q-statistic for one lag is significant at the 5% level. For the daily frequency, the Q-statistic for ten lags is significant. The Q-statistics of the remaining lags are not significant, indicating that most of the serial dependence is removed by the standardization.

5.2.1 Fitting IG to standardized RV results

We fit the IG to the different frequencies of the standardized RV. The results are reported in Figures 5.3 a,b and 5.4 a, b. In Figure 5.3a, b, we illustrate the fit of the IG to the different frequencies of the standardized RV. Overall, the IG distribution seems to give a good fit.

In Figure $4, a, b$ we present the QQ-plots of the PIT of the data against the quantiles of a $U(0,1)$ distribution. For all the frequencies, the empirical quantiles are very close to the 45 degree line, indicating a good fit of the IG distribution to the standardized RV.

5.2.2 Moment aggregation of the realized volatility

To evaluate the assumption of an IG distribution, we evaluate the ability of the IG to aggregate moments over the different frequencies. If the IG distribution is a good approximation of the RV, this should be seen in the temporal aggregation properties of the IG when compared with the data. We can estimate the parameters of the IG using the daily frequency, aggregate the parameter values to the monthly frequency, and calculate the "aggregation implied" moments using the aggregated parameter values. These implied moments should then be the same as those parametric moments calculated when the parameters are estimated on the monthly data. It should also match the sample moments of the monthly frequency. This idea is reflected in Tables 5.5a to d.

To explain the idea behind these aggregational tables, we first study Table 5.5a. This table reports the aggregated mean of the IG when fitted to the standardized RV. We do the following: We estimate the sample mean for the daily, weekly, bi-weekly and monthly standardized RV; the values obtained are reported in the first row, the "empirical mean". The part of the table "implied parametric means" is constructed as follows. For the first row: We estimate the parameters of the IG distribution on the daily frequency using the method of moments based on the first two moments.[2] Using the parameter estimates, we calculate the parametric mean (4.13) for the daily frequency. This number should be compared to the sample mean for the daily data. Next, we use the formulas for the temporal aggregation of the parameters (5.1) to calculate the "implied parameter value" for the weekly frequency. The implied parametric mean for the weekly frequency, using the parameter estimates from the daily frequency will be

$$\widehat{\sigma}^2_{\text{Weekly, implied from daily}} = 5\widehat{\sigma}^2_{\text{Daily}}.$$

This number $\left(\widehat{\sigma}^2_{\text{implied weekly from daily}}\right)$ is reported in the first row of the table "implied parametric mean" in the second column, and should be compared to the sample mean for the weekly frequency in the row above. Next, we aggregate the parameters from the daily frequency to the bi-weekly:

$$\widehat{\sigma}^2_{\text{Bi-weekly, implied from daily}} = 10\widehat{\sigma}^2_{\text{Daily}}.$$

This number is reported in the third column. Finally, we aggregate the parameter to obtain the monthly value $\widehat{\sigma}^2_{\text{Monthly, implied from daily}} = 20\widehat{\sigma}^2_{\text{Daily}}$.

When constructing the second row of the part of the table "implied parametric means", we start by estimating the parameters of the IG on the weekly raw RV, and in the second column the parametric mean $\widehat{\sigma}^2_{\text{Weekly}}$. We aggregate the weekly estimate to bi-weekly using $\widehat{\sigma}^2_{\text{Bi-weekly, implied from weekly}} = 2\widehat{\sigma}^2_{\text{Weekly}}$. This number is reported in the third column, we continue in the same manner for the row bi-weekly. The last row, monthly, has only one number and that is the parametric mean of the IG using the parameter estimate $\widehat{\sigma}^2_{\text{Monthly}}$, obtained by estimating the parameters of the IG on the monthly standardized RV.

If the RV is IG distributed, the number in the columns of Table 5.5*a* should be the same. Given the IG distribution, we should have the same parametric estimate of the mean of, say, the bi-weekly frequency if we estimate the parameters of the IG using the daily frequency $\left(\widehat{\sigma}^2_{\text{Daily}}\right)$and calculate the implied parametric mean for the bi-weekly $\left(10\widehat{\sigma}^2_{\text{Daily}}\right)$ or if we fit the parameter of the IG directly on the bi-weekly frequency $\left(\widehat{\sigma}^2_{\text{Bi-weekly}}\right)$.

The procedure for deriving the lower part of Tables 5.5*a* to *d* can be summarized as follows

1. To get the first row:

[2] I experimented with other estimation methods, such as the method of moments, and a minimum chi-square estimation, and found that they yielded similar estimates.

- Estimate the parameters of the IG on the daily frequency.
- Calculate the "implied parameter values" for the weekly, bi-weekly and monthly frequency using (5.1) .
- Calculate the "implied moments" of the weekly, bi-weekly and monthly frequency using (4.13) .

2. To get the second row:

   - Estimate the parameters using the weekly frequency, and calculate the parametric moment for the weekly frequency.
   - Calculate the "implied parameter values" for the bi-weekly and monthly frequencies using (5.1) .
   - Calculate the "implied moments" for the bi-weekly and monthly frequencies using (4.13) .
   - Compare these implied moments to the empirical moments calculated on the raw RV.

3. To get the third row:

   - Estimate the parameters using the bi-weekly frequency, and calculate the parametric moment for the bi-weekly frequency.
   - Calculate the "implied parameter values" for the monthly frequency using (5.1) .
   - Calculate the "implied moments" for the bi-weekly and monthly frequency using (4.13) .
   - Compare these implied moments to the empirical moments calculated on the raw RV.

4. To get the fourth row:

   - Estimate the parameters using the monthly frequency and calculate the parametric moment for the monthly frequency.

Tables 5b, c and d are produced in a similar manner, but now calculating the "aggregation implied" variance, skewness and kurtosis respectively, by using the aggregation results of (5.1) and the measures (4.13) .

5.2.3 Analytical aggregation of the moments of IG on standardized realized volatility

Here we present the results for the aggregation of the moments of the IG over the different frequencies of the standardized realized volatility. The results are presented in Tables 5.5*a* to *d*. Table 5.5*a* reports the aggregation of the mean of the IG distribution. For a given row, the numbers are almost identical across the different frequencies. This means, that we can use the parameter estimates from the daily frequency to get a good estimate of, say, the monthly (implied parametric) mean.

The aggregation seems to work fine for the variance (Table 5.5*b*) too. The monthly parametric variance implied from the daily ones is 5.868, to be compared to 6.026 for the empirical variance and 5.852 for the parametric variance, where the parameters are estimated on the monthly data. The numbers in the monthly column are between 5.852 and 6.543, indicating that the variance aggregates over the different frequencies.

The aggregation of the skewness (Table 5.5*c*) works reasonable well, although the parametric estimates are lower than the sample analogs. We can see that the empirical skewness starts at 3.419 for the daily frequency and decreases over the weekly and bi-weekly frequencies, and the empirical skewness for the monthly data is 0.556. This is an indication of the central limit theorem kicking in. As we aggregate more and more observations, the distribution of the aggregate should converge to the normal distribution. This is why the skewness goes to zero.

In Table 5.5*d*, the empirical kurtosis starts at 28.235 for the daily frequency, for the bi-weekly frequency is 3.835 and for the monthly 2.933. This is also an indication of the CLT kicking in. The kurtosis for the normal distribution is three, and the empirical kurtosis for the monthly frequency is 2.933. The parametric estimate of the kurtosis for the daily frequency is 7.396 compared to the sample analogue of 28.235, and in column 2 the weekly parametric kurtosis (3.879) and the weekly kurtosis implied from the daily parameter estimates (3.980) are below the sample estimate of 5.862, but otherwise the parametric kurtosis matches the sample kurtosis. Throughout the monthly column we have numbers about three, indicating that the IG distribution aggregates nicely over the different frequencies.

5.3 Conclusions

In this chapter, we have shown that the distribution of realized volatility, conditional on the information set F_{t-1}, consisting of lagged realized volatilities, is well described by the inverse Gaussian distribution in terms of aggregation properties. Empirically, for the ECU/USD 1989 - 1998 dataset, we have shown that when we aggregate the parameter estimates from the daily frequency to obtain the lower frequencies, we get a good fit of the implied moments for the lower frequencies.

5.4 Tables

Table 5.1:
Parameter estimates of the ARMA-IG(1,1) standardization of RV of ECU/USD 1989 - 1998.

	Estimates
$\overline{\alpha}$	3.416
	(0.071)
ϕ_0	0.026
	(0.003)
ϕ_1	0.940
	(0.008)
θ_1	-0.700
	(0.016)

Notes: The estimation is done by ML. Standard errors of the estimates are given in parenthesis. The ARMA(1,1)-IG standardization procedure is described in the text.

Table 5.2:
Ljung-Box Q-statistics of raw and standardized RV ECU/USD 1989 - 1998.
(standardized by ARMA(1,1)-IG).

	Raw RV	RV standardized by ARMA(1,1)-IG
$Q\,(1)$	667.9	1.560
	(0.000)	(0.211)
$Q\,(5)$	2133.9	10.637
	(0.000)	(0.059)
$Q\,(10)$	3147.0	20.574
	(0.000)	(0.024)
$Q\,(20)$	3914.6	47.072
	(0.000)	(0.000)

Notes:$Q\,(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized returns for up to lag (n). The p-values are given in parenthesis.

Table 5.3:
Descriptives of RV of ECU/USD 1989 - 1998, standardized by ARMA(1,1)-IG.

	Frequency			
	Daily	Weekly	Bi-weekly	Monthly
Mean	1.000	5.004	10.008	20.016
Median	0.875	4.761	9.880	19.842
Maximum	8.284	12.093	15.948	27.735
Minimum	0.071	2.057	6.249	15.014
Std	0.586	1.335	1.773	2.465
Skewness	3.419	1.252	0.785	0.556
Kurtosis	28.235	5.862	3.835	2.933
$Q(1)$	1.560	4.245	0.856	0.130
	(0.212)	(0.039)	(0.355)	(0.718)
$Q(10)$	20.574	18.140	17.386	8.815
	(0.024)	(0.053)	(0.066)	(0.550)

Notes: The daily RV is standardized by ARMA(1,1)-IG. The lower frequencies are aggregated from the daily standardized RV. $Q(n)$ denotes the Ljung-Box statistic for serial correlation in the standardized returns for up to lag (n). The p-values are given in parenthesis.

Table 5.4:
Estimates of IG fitted to RV of ECU/USD 1989 - 1998, standardized by ARMA(1,1)-IG.

	Frequency			
	Daily	Weekly	Bi-weekly	Monthly
σ^2	1.000	5.004	10.008	20.016
	(0.010)	(0.057)	(0.111)	(0.221)
$\overline{\alpha}$	3.411	15.308	33.365	68.463
	(0.070)	(0.898)	(2.939)	(9.980)

Notes: The RV has been standardized by the procedure ARMA(1,1)-IG as described in the text. The weekly, bi-weekly and monthly RV is aggregated from the daily standardized RV. Then the IG is fitted to the standardized RV, estimation method is ML, the standard errors are given in parenthesis.

Table 5.5a: Mean
Temporal aggregation of mean of IG fitted to standardized RV
of ECU/USD 1989 - 1998. RV is standardized by ARMA(1,1)-IG.

		Frequency			
		Daily	Weekly	Bi-weekly	Monthly
	Empirical mean	1.000	5.004	10.008	20.016
Implied	Daily	1.001	5.003	10.005	20.011
parametric	Weekly	-	5.004	10.008	20.017
means	Bi-weekly	-	-	10.008	20.017
	Monthly	-	-	-	20.016

Notes: RV has been standardized by ARMA(1,1)-IG as described in the text. The weekly, bi-weekly and monthly data is then aggregated from the daily standardized data. The row empirical mean is the sample mean of these (standardized and aggregated) data. Rest of the table, first row: We estimated the parameters of the IG using the daily standardized data, and then we calculate the parametric mean for the daily frequency.Using the parameter estimates (from the daily standardized data), we then used the aggregation formulas described in the text to calculate the implied parameter values for the weekly, bi-weekly and monthly frequency. Using these "aggregation-implied" parameter values we calculated the parametric mean of the weekly, bi-weekly and monthly frequencies. These are presented in the first row. The procedure is similar to create the second row, but now we estimated the parameters of the IG using the standardized weekly data, and then aggregated the parameter values to bi-weekly and monthly values, using these "aggregation-implied" parameter values to calculate the parametric mean. The numbers should be compared column-wise.

Table 5.5b: Variance
Temporal aggregation of variance of IG fitted to the standardized
RV of ECU/USD 1989 - 1998.
The RV is standardized by ARMA(1,1)-IG.

		Frequency			
		Daily	Weekly	Bi-weekly	Monthly
	Empirical variance	0.344	1.779	3.129	6.026
Implied	Daily	0.293	1.467	2.934	5.868
parametric	Weekly	-	1.636	3.272	6.543
variances	Bi-weekly	-	-	3.002	6.004
	Monthly	-	-	-	5.852

Notes: See notes for Table 5a, here we calculated the parametric variance instead of the parametric mean.

Table 5.5c: Skewness
Temporal aggregation of skewness of IG fitted to the standardized
RV of ECU/USD 1989 - 1998.
The RV is standardized by ARMA(1,1)-IG.

		Frequency			
		Daily	Weekly	Bi-weekly	Monthly
	Empirical skewness	3.419	1.252	0.785	0.556
Implied	Daily	1.624	0.726	0.514	0.363
parametric	Weekly	-	0.767	0.542	0.383
skewness	Bi-weekly	-	-	0.519	0.367
	Monthly	-	-	-	0.363

Notes: See notes for Table 5a, here we calculate the parametric skewness instead of the parametric mean.

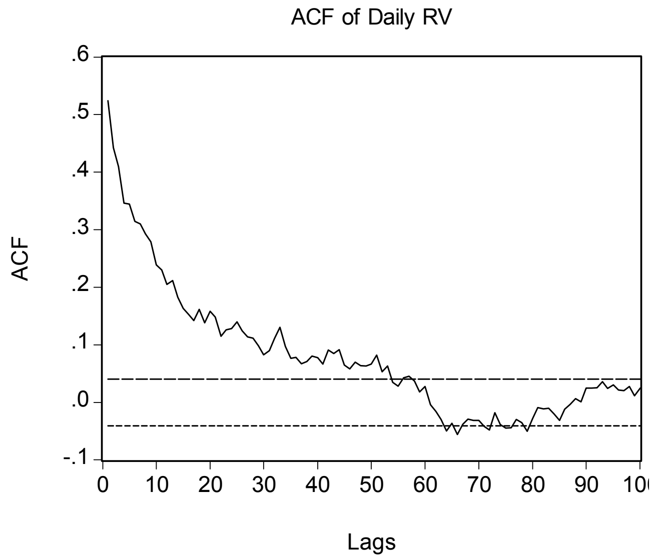
Table 5.5d: Kurtosis
Temporal aggregation of kurtosis of IG fitted to the standardized RV of ECU/USD 1989 - 1998.
The RV is standardized by ARMA(1,1)-IG.

		Frequency			
		Daily	Weekly	Bi-weekly	Monthly
	Empirical kurtosis	28.235	5.862	3.835	2.933
Implied	Daily	7.396	3.879	3.440	3.220
Parametric	Weekly	-	3.980	3.490	3.245
kurtosis	Bi-weekly	-	-	3.450	3.225
	Monthly	-	-	-	3.219

Notes: See notes for Table 5a, here we calculate the parametric kurtosis instead of the parametric mean.

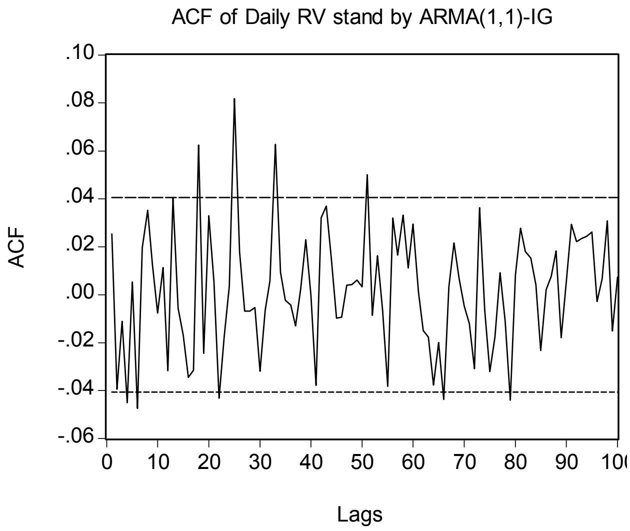
5.5 Figures

Figure 5.1:



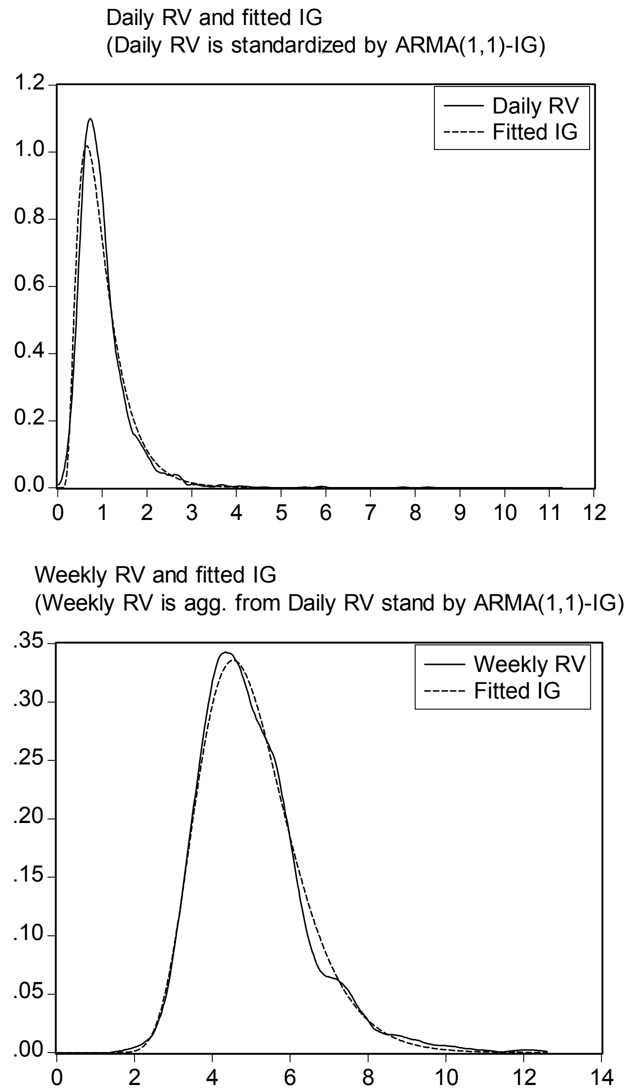
Notes: The figure shows the SACF of the daily realized volatility for the ECU/USD 1989 - 1998.

Figure 5.2:
Sample autocorrelation functions of the daily standardized RV of ECU/USD 1989 - 1998.



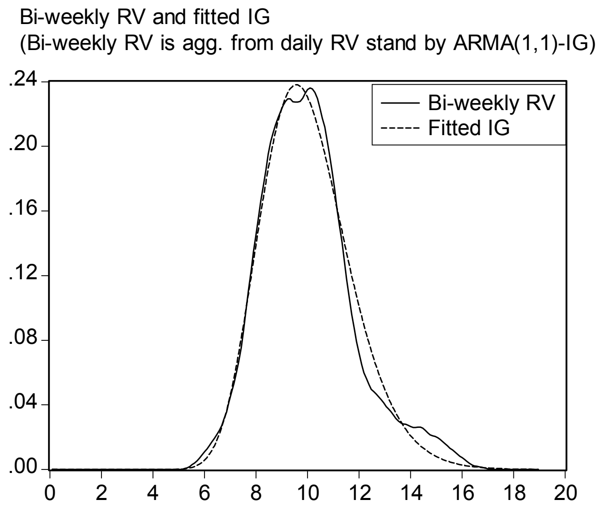
Notes: The RV is standardized by the ARMA(1,1)-IG as described in the text.

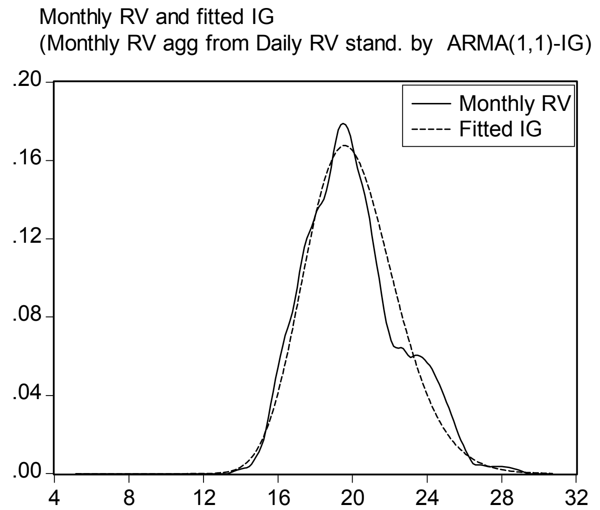
Figure 5.3a:
Standardized RV of ECU/USD 1989 - 1998 and fitted IG.



Notes: The RV is standardized by the ARMA(1,1)-IG procedure as described in the text.

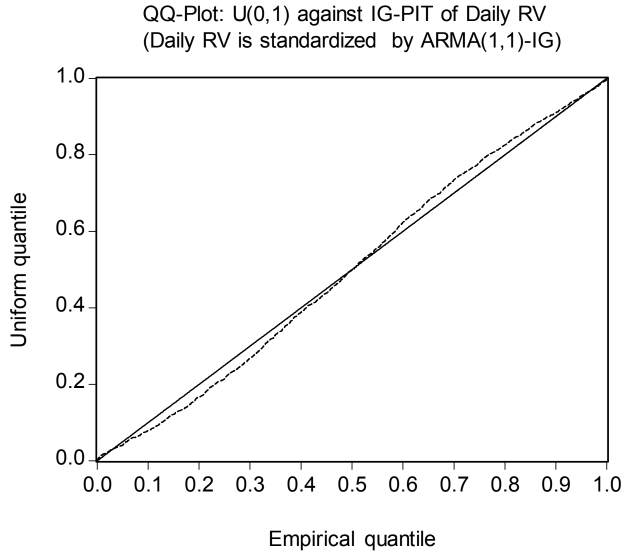
Figure 5.3b:
Standardized RV of ECU/USD 1989 - 1998 and fitted IG.

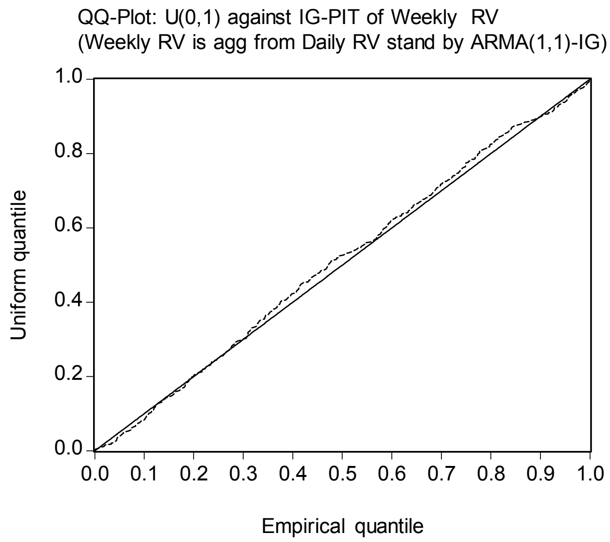




Notes: The RV is standardized by the ARMA(1,1)-IG procedure as described in the text.

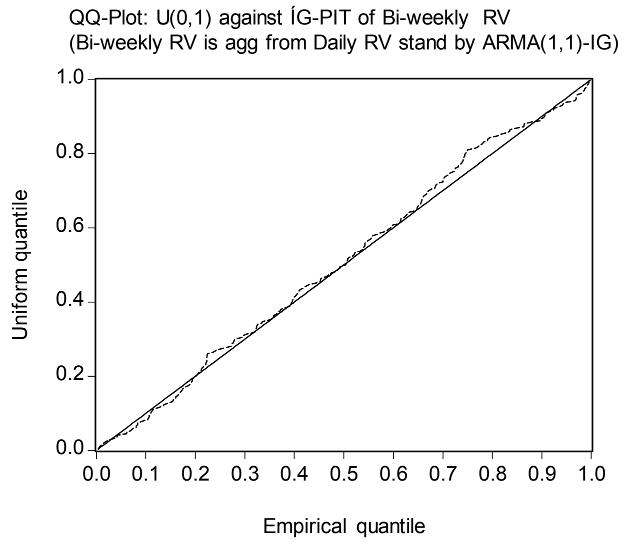
Figure 5.4a:
QQ-plots of IG-PIT of standardized RV of ECU/USD 1989 - 1998 against $U(0,1)$.

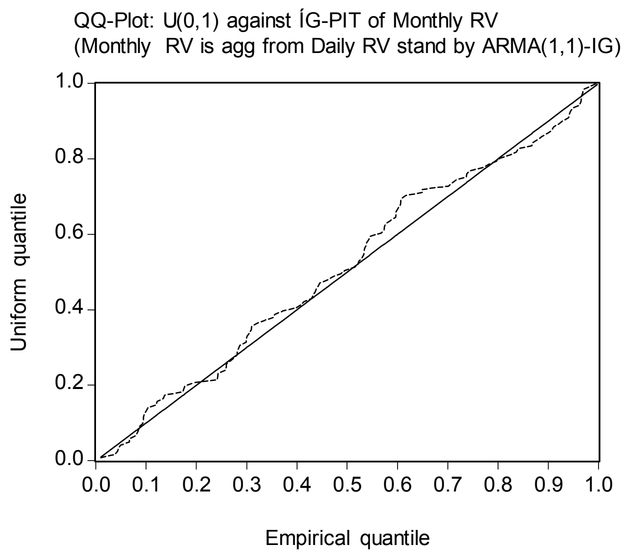




Notes: The IG-PIT is the probability integral transform of the data using the IG distribution. The daily data has been standardized using the ARMA(1,1)-IG procedure as described in the text. The lower frequencies have been derived by aggregating the daily standardized RV.

Figure 5.4b:
QQ-plots of IG-PIT of standardized RV of ECU/USD 1989 - 1998 against $U\left(0,1\right).$





Notes: The IG-PIT is the probability integral transform of the data using the IG distribution. The daily data has been standardized using the ARMA(1,1)-IG procedure as described in the text. The lower frequencies have been derived by aggregating the daily standardized RV.

Chapter 6

The General GARCH-NIG model

Thus far we have motivated the use of the NIG distribution for the conditional returns when modeling conditional variance in the financial markets. This was done by linking the conditional variance to the conditional realized volatility, and by showing that the inverse Gaussian distribution gives a good fit to the conditional variance. Still, we have only used the "plain vanilla" GARCH parameterization of the conditional variance. This parameterization captures the most prominent stylized facts concerning returns, the non-normality of the returns and the volatility clustering. However, numerous extensions of the standard GARCH model have been proposed in the literature to capture other stylized facts, such as the asymmetry and the Taylor effect, to mention just two.

In this chapter, we introduce a new General GARCH(p,q)-NIG model.[1] As special cases we get three models that are designed to capture the asymmetry, the Taylor effect and the asymmetry and the Taylor effect simultaneously. The model capturing the asymmetry effect is the Threshold GARCH(p,q)-NIG (T-GARCH(p,q)-NIG) model, where we model the conditional *variance*. The model for the Taylor effect, where we model the conditional *standard deviation*, is the Absolute Value GARCH(p,q)-NIG (AV-GARCH(p,q)) model, and the model capturing both the asymmetry and the Taylor effect is the Threshold Absolute Value GARCH(p,q)-NIG (TAV-GARCH(p,q)-NIG) model. In the last two models, we parameterize the conditional *standard deviation.* These models are generalizations of existing models in the literature. The

[1] A noted in the introduction: In a concurrent and independent work, Jensen and Lunde (2001) proposed a model they refer to as the NIG-S&ARCH model, which is the A-PARCH model of Ding, Granger and Engle (1993) used with the NIG distribution. This model has as special cases the GARCH-NIG, T-GARCH-NIG and the AV-GARCH-NIG. However, they do not give the moment structure of their model, or any special case, nor do they provide the analytical gradients or Hessian of the model.

existing models use the normality assumption for the standardized returns, and we generalize these models by assuming the NIG distribution for the standardized returns.

This chapter is organized as follows. First we give a short review of the asymmetry and the Taylor effect, and the models that are proposed in the literature to capture these effects. Then we formulate the General GARCH-NIG(p,q) model that will nest all the GARCH-NIG models proposed in this thesis. From The General GARCH-NIG(p,q) model, we derive the three models as special cases, and we go into some detail of the different models. For brevity, the moment structures, the gradients and the Hessians of the models are given in appendices.

6.1 Asymmetry

One of the stylized facts concerning asset returns, is that the conditional variance is sometimes asymmetric. This seems to have been noted first by Black (1976).[2] By this, we mean that the conditional variance at time t is bigger if the return at time $t-1$ is negative, than it is when the return at time $t-1$ is positive and of the same size. One could say that we have a non-zero correlation between r_t^2 and r_{t-j}. For the GARCH(p,q)-NIG model, i.e., $r_t = \sigma_t z_t$, where $z_t \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1)$, we have $E(r_t) = 0$ and $E(z_t) = 0$, so

$$\begin{aligned} cov\left(r_t^2, r_{t-j}\right) &= E\left(r_t^2 r_{t-j}\right) \\ &= E\left(\sigma_t^2 z_{t-j} \sigma_{t-j}\right). \end{aligned}$$

If σ_t^2 is an even function of z_{t-j}, this covariance is zero. The first GARCH model to take this asymmetry into account, was the Exponential GARCH(p,q) model of Nelson (1991). In the EGARCH(p,q) model we model the log of the variance, and the specification of the is asymmetric. Another asymmetric model was introduced by Glosten, Jagannathan and Runkle (1992), their model is refered to as the GJR-GARCH, and the (1,1) case can be written as

$$r_t = \sigma_t \varepsilon_t, \tag{6.1}$$

[2] Black(1976) introduced the term leverage effect. The name is due to when the value of a firm drops, i.e., the stock price falls, the financial leverage of the firm increases. However, Black (1976), Christie (1982) and Schwert (1989) argues that the leverage effect cannot explain all the asymmetry in the data. Another explanation to the asymmetry is the volatilty feedback story. According to the volatility feedback story the increased volatility comes first, raising required returns on the stock, causing the price to fall. For an excellent discussion of the origins of asymmetry, see Bekaert and Wu (2000).

where $\varepsilon_t \sim N(0,1)$ and

$$\sigma_t^2 = \rho_0 + \rho_1 r_{t-1}^2 + \omega_1 I(\varepsilon_{t-1}) r_{t-1}^2 + \pi_1 \sigma_{t-1}^2,$$

where

$$\begin{aligned} I(\varepsilon_{t-1}) &= 1 \text{ if } \varepsilon_{t-1} < 0 \\ &= 0 \text{ otherwise.} \end{aligned}$$

The indicator function $I(\varepsilon_{t-1})$ makes the response of the volatility to the lagged return asymmetric. If yesterday's return was negative, the conditional variance today is $\omega_1 r_{t-1}^2$ larger than it would have been if the return had been positive. Below, we will generalize the GJR-model to allow for NIG distributed errors.

Zakoïan (1994) introduces an asymmetric model similar to the GJR model, but in Zakoïan's specification, we model the conditional standard deviation. The model of Zakoïan (1994) is referred to as the Threshold-GARCH(p,q) model, or the T-GARCH model. The T-GARCH(p,q) model of Zakoïan can be written as (6.1) where $\varepsilon_t \sim N(0,1)$, but where we parameterize the conditional standard deviation instead of the conditional variance, that is

$$\sigma_t = \rho_0 + \rho_1 |r_{t-1}| + \omega_1 I(r_{t-1}) |r_{t-1}| + \pi_1 \sigma_{t-1}. \tag{6.2}$$

As seen in (6.2), this model uses the absolute value of the lagged returns instead of the conditional variance. Since we model the conditional standard deviation, and we square the conditional standard deviation to get the conditional variance, Zakoïan claims that we can allow the conditional standard deviation to be negative.

6.2 Modeling the conditional standard deviation

Taylor (1986) introduced the Taylor effect, which refers to the empirical fact that the strongest temporal dependence seems to be in the standard deviation, not in the variance of the returns; see also Ding et al (1993). To capture this effect, we develop two new models, the Absolute Value GARCH(p,q)-NIG and the Threshold Absolute Value GARCH(p,q)-NIG. In the next section, we outline a general GARCH(p,q)-NIG model, that nests all the models proposed in this chapter. Then, using this general GARCH(p,q)-NIG model, we derive the new models.

6.3 A general GARCH-NIG(p,q) model

To set the stage for the new models, we introduce a general GARCH(p,q)-NIG model, that nests all the GARCH-NIG models considered in this chapter. Let

x_t be the return at time t, which we decompose as

$$x_t = \sigma_t z_t, \tag{6.3}$$

where

$$z_t \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1),$$

and where we model σ_t^k as

$$\sigma_t^k = \rho_0 + \Sigma_{i=1}^q \rho_i |x_{t-i}|^k + \Sigma_{i=1}^q \omega_i I(x_{t-i} < 0) |x_{t-i}|^k + \Sigma_{j=1}^p \pi_j \sigma_{t-j}^k. \tag{6.4}$$

where k is an integer, $k = 1, 2$ and

$$\begin{aligned} I(z_t) &= 1 \text{ if } z_t < 0 \\ &= 0 \text{ otherwise,} \end{aligned}$$

is an indicator function. $I(z_t)$ captures the asymmetry in the conditional volatility. Sufficient restrictions of the parameters to guarantee positivity of the σ_t^k process are $\rho_i > 0$, $\rho_i + \omega_i > 0$ $i = 1, ..., q$ and $\pi_j > 0$ for $j = 1, ..., p$

are The integer k determines whether we model the conditional variance or conditional standard deviation. With $k = 2$, we model the conditional variance and for $k = 1$ we model the conditional standard deviation. To get the GARCH(p,q)-NIG model using (6.4), we choose $k = 2$ and we set $\omega_i = 0$ for $i = 1, ..., q$.

6.3.1 The Threshold-GARCH(p,q)-NIG model

Here, we generalize the GJR model to have NIG errors: we will refer to this model as the Threshold GARCH(p,q)-NIG model.[3] To get the Threshold-GARCH(p,q)-NIG model, using (6.4), we let $k = 2$ and x_t follow (6.3), and we model σ_t^2 as

$$\sigma_t^2 = \rho_0 + \Sigma_{i=1}^q \rho_i x_{t-i}^2 + \Sigma_{i=1}^q \omega_i I(x_{t-i} < 0) x_{t-i}^2 + \Sigma_{j=1}^p \pi_j \sigma_{t-j}^2. \tag{6.5}$$

The indicator function with the extra parameter ω_i will give the process an increase in the conditional variance when the lagged returns are negative.

To see the effect of the different signs of the lagged residual on the σ_t^2, we study the T-GARCH(1,1) process and we write out the process given the

[3] In the literature, the name Threshold GARCH(p,q)-*Normal* model is that given to the model of Zakoïan (1994), which models the conditional *standard deviation*. However, we choose to refer to this model as the Threshold GARCH(p,q)-NIG model to get consistency in the naming of the new models, rather than retaining the inconsistent historical names. The generalization of Zakoïan'ss model will be referred to as the Absolute Value GARCH(p,q)-NIG, since it uses the absolute value of the lagged returns as explanatory variable in the equation of the conditional standard deviation.

different signs of the lagged observation. For $x_{t-1} > 0$, $I(x_{t-1}) = 0$ and the conditional variance becomes

$$\sigma_t^2 = \rho_0 + \rho_1 x_{t-1}^2 + \pi_1 \sigma_{t-1}^2,$$

for $x_{t-1} < 0$, $I(x_{t-1}) = 1$ and we have

$$\sigma_t^2 = \rho_0 + (\rho_1 + \omega_1) x_{t-1}^2 + \pi_1 \sigma_{t-1}^2.$$

Note the additional parameter ω_1, which kicks in when the lagged return is negative. The moment structure of the T-GARCH(p,q)-NIG model is given in Appendix A and B.

6.3.2 The Absolute Value-GARCH(p,q)-NIG model

Using the general formulation in (6.4), we get a symmetric model for the conditional standard deviation by choosing $k = 1$, and by setting $\omega_i = 0$ for $i = 1, .., q$. This will give us a model that has a symmetric response to the sign of the lagged return, and where we use the absolute value of the lagged returns as explanatory variables in the equation for the conditional standard deviation. We will call this model the Absolute Value GARCH(p,q)-NIG model, and it is given by

$$x_t = \sigma_t z_t, \tag{6.6}$$

where

$$z_t \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1),$$

and we model the conditional standard deviation by

$$\sigma_t = \rho_0 + \Sigma_{i=1}^q \rho_i |x_{t-i}| + \Sigma_{j=1}^p \pi_j \sigma_{t-j}. \tag{6.7}$$

This model is a direct generalization of the Absolute Value GARCH(p,q)-normal model of Hentschel (1991). The moment structure of the AV-GARCH(p,q)-NIG model is given in Appendix A and B.

6.3.3 The Threshold Absolute Value GARCH(p,q)-NIG model

In order to model the conditional standard deviation while taking into account the conditional asymmetry, and still using the NIG distribution, we derive the TAV-GARCH(p,q)-NIG model. Using the general formulation in (6.4), we get an asymmetric model for the conditional standard deviation by choosing $k = 1$, and we estimate ω_i, $i = 1, .., q$. This will result in the Threshold

Absolute Value GARCH(p,q)-NIG model.[4] Again, let x_t be the returns at time t, then

$$x_t = \sigma_t z_t,$$

where $z_t \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1)$ and we model the conditional standard deviation by

$$\sigma_t = \rho_0 + \Sigma_{i=1}^{q} \left[\rho_{i1} + \omega_1 I\left(x_{t-1} < 0\right)\right] \left|x_{t-1}\right| + \Sigma_{j=1}^{p} \pi_j \sigma_{t-j}. \tag{6.8}$$

To see the effect of different signs of the lagged residual, on the conditional standard deviation σ_t, we study the property of the process for the different signs on the lagged observation. We study the (1,1) case i.e.,

$$\sigma_t = \rho_0 + \rho_1 \left|x_{t-1}\right| + \omega_1 I\left(x_{t-1} < 0\right) \left|x_{t-1}\right| + \pi_1 \sigma_{t-1}.$$

For $x_{t-1} < 0$, $I\left(x_{t-1}\right) = 1$ and we have

$$\sigma_t = \rho_0 + \rho_1 \left|x_{t-1}\right| + \omega_1 \left|x_{t-1}\right| + \pi_1 \sigma_{t-1},$$

hence

$$\sigma_t = \rho_0 + \left(\rho_1 + \omega_1\right) \left|x_{t-1}\right| + \pi_1 \sigma_{t-1},$$

and, for $x_{t-1} > 0$, $I\left(x_{t-1}\right) = 0$ and we find

$$\sigma_t = \rho_0 + \rho_1 \left|x_{t-1}\right| + \pi_1 \sigma_{t-1}.$$

It is clear from the above that when the lagged return is negative, the effect of the lagged return on the conditional standard deviation is larger $\left(\rho_1 + \omega_1\right)$ than when the lagged return is positive $\left(\rho_1\right)$. Therefore, the model is expected to capture the asymmetric behaviour of the data. The moment structure of the TAV-GARCH(p,q)-NIG model is given in Appendix A and B.

6.4 Estimation of the models

Estimation of the parameters of the different models is straightforward using maximum likelihood. The log likelihood for the sample for any of the GARCH-NIG models is given by

$$\begin{aligned} \ln l_t \;=\;& \left(T - \max\left(p, q\right)\right) \left(\frac{1}{2} \ln \overline{\alpha} - \ln\left(\pi\right)\right) - \Sigma_{\max(p,q)}^{T} \left(\frac{1}{2} \ln \sigma_t^2 + \overline{\alpha}\right) \\ & - \Sigma_{\max(p,q)}^{T} \frac{1}{2} \ln\left(1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}}\right) + \Sigma_{\max(p,q)}^{T} \ln K_1 \left(\overline{\alpha}\left(1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}}\right)^{1/2}\right), \end{aligned} \tag{6.9}$$

[4] This model can be viewed as a generalization of the Threshold-GARCH(p,q) model of Zakoïan (1994), where we use the NIG distribution instead of the normal distribution. Since this new model uses the absolute value of the returns as explanatory variables in the equation for the conditional standard deviation, it makes more sense to use the name Threshold Absolute Value GARCH(p,q)-NIG for this model.

where T is the number of observations and where σ_t^2 is given by the model choosen i.e., by one of equations (2.5) , (6.5) , (6.7) or (6.8). Note that σ_t^2 has to be derived recursively. The expression for the gradients and Hessians of the different models are given in Appendix C. In the next chapter, we evaluate the properties of the maximum likelihood estimator for these models.

6.5 Concluding remarks

In this chapter we introduced the General GARCH(p,q)-NIG model that nests the four models GARCH(p,q)-NIG, T-GARCH(p,q)-NIG, AV-GARCH(p,q)-NIG and TAV-GARCH(p,q)-NIG. These models are introduced to capture the asymmetry effect, the Taylor effect, and both effects simultaneously. We give the moment structure for the (1,1) cases and also the analytical gradients and Hessians for the models.

6.6 Further work

A natural extension of this chapter is to develop tests for the models. That is, tests for ARCH-effects, given the assumption of a NIG distribution. In addition, one could develop tests for asymmetry in the conditional variance given the NIG assumption. Maybe it would be possible use the realized volatility and the IG distribution to develop tests for asymmetry. As in Chapter 4, one could assume that the realized volatility follows an IG distribution and use the RV-GARCH-IG model, and inlude an indicator function for the sign of the lagged return. Using realized volatility it should be possible to develop more powerful tests, since the realized volatility contains more information than just the squared return. Of course, one could use other distributional assumptions for the realized volatility in this context. When building models aimed for capturing the Taylor effect, one could use the idea of realized power variation, as introduced in Barndorff-Nielsen and Shephard (2001c).

There is a need to develop model tests for the GARCH-NIG models. The standard tests for ARCH effects assumes normality for the standardized returns, and the normality assumption simplifies the test procedure. With the normality assumption, in the GARCH case, we can run a regression of r_t^2 on lagged returns, r_{t-i}^2 and test whether T times the R^2 from this regression is zero. This simplified version of the LM test is due to the normality of the standardized returns. When assuming othter distributions for the standardized returns are NIG, we need to get back to basic concerning the LM test methods. The basic issue here is that we need to estimate the (nuisance) parameter $\overline{\alpha}$ and take the uncertinty of the estimate into account. Alternatively, we can use tests based on moment conditions as outlined in Newey (1985) and Tauchen (1985).

Chapter 7

Simulation study

In this chapter we evaluate the performance of the maximum likelihood estimator for the models: the GARCH(1,1)-NIG, T-GARCH(1,1)-NIG, AV-GARCH(1,1) and the TAV-GARCH(1,1).[1] We investigate the normality of the estimates, the bias and the standard errors. We evaluate the empirical coverage rates of 95 percent confidence intervals.

Since the analytical derivatives of the GARCH-NIG models, as given in Appendix C, are quite cumbersome to implement, one might ask if the numerical derivatives perform equally well. Thus, we also make a comparison of the above measures using analytical and numerical derivatives.

7.1 Simulation setup

We simulate the four different models for different parameter setups and different sample sizes. We use three different parameter combinations or parameter sets, outlined below. The number of observations for each parameter set is n=250, 500, 1000, 2000 and 4000, in the case of daily data, this would correspond to one, two, four, eight and sixteen years of daily data, respectively,[2]. For stocks that have a short history on the stock market, the smallest sample size is of interest. How well can we estimate the parameters of a GARCH-NIG model for a stock that has been traded for only a year? Is one year of data enough to get good estimates of the parameters, what about two years...? Do we need analytical gradients, or can we rely on the numerical gradients. For the larger sample sizes, we expect the estimates to be close to the true

[1] I would like to thank Johan Lyhagen at the Stockholm School of Economics, for assisting in the implemenation (i.e., running the simulations) for this study.

[2] The number of trading days for the stock markets is around 250 per year. For foreign exchange (FX) markets, the trading can include weekends and holidays, so the number of trading days for a year can be higher than 250.

For simplicity, we refer to the number of trading days as if we were dealing with data from the stock market.

values. One might discuss if the assumed GARCH process is stable over a period of eight or sixteen years. There might of course be regime shifts (structural breaks) in the parameters during such a long period of time. However, the question of occurance of regime shifts in the process is beyond the scope of this study, where we assume the parameters to be stable over the sample period.

Since the (1,1)-case is the most commonly encountered in the literature, we focus our attention on this case. For each model, we examine three different sets of parameters, these parameter combinations are given in Tables 7.1 a,b and 7.2 a,b. The parameter sets were chosen in such a way that, for Set 1, the unconditional variance and the unconditional fourth moment of the models from which data were generated exist. In Set 2, the parameters were chosen such that the condition for the fourth moment (i.e., γ_{c4} for the models modeling the conditional variance, the GARCH(1,1)-NIG and T-GARCH(1,1)-NIG models, and γ_{c2} for the models modeling the conditional variance GARCH(1,1)-NIG and T-GARCH(1,1)-NIG models) would be close to one. In Set 3, the unconditional fourth moment does not exist for the models.

The setup for the simulation was as follows. The simulation study was implemented in GAUSS. For each model, data was generated for the different sample sizes. The model was estimated, the estimates saved, and then the next replication was generated and so on. The number of replications was 1000 for every setup.[3] Once all the replications had been completed, the evaluation criteria were calculated. The simulation of the GARCH-NIG variates was done as follows.[4] To start up the recursion of the GARCH equation (4.17), the initial values of the r_{t-1} and σ_{t-1}^2 were set to zero. Using the parameter values for that setup, the conditional variance (σ_1^2) was calculated for the first observation. Then an IG variate was generated given the value of $\overline{\alpha}$ and σ_1^2, that is $z_t \sim IG_{\sigma^2}\left(\sigma_1^2, \overline{\alpha}\right)$.[5] Then an $x_t \sim N(0,1)$ variate was generated, and the first simulated GARCH-NIG observation is given by $r_t = \sqrt{z_t} x_t$. Now, given the (observed) return r_t, the variance Equation (4.17) is updated. To reduce the impact of the starting values ($r_t = 0$ and $\sigma^2 = 0$), the first 500 observations were discarded. This procedure was repeated until the necessary number of observations was generated.

[3] This number of replications for each setup was chosen as a compromize between the number of setups used, and the notion that "the more replicates the better", given limited computer resourses.

[4] Here, we present the case of generating the GARCH(1,1)-NIG variates, the data from the other models was generated with the obvious adjustments implemented in the procedure.

[5] How to generate data from the IG distribution is explained in Michael, Schucany and Haas (1976).

The estimation of the parameters of the models were performed as follows. We use the BHHH algorithm.[6] This algorithm uses the outer product of the gradients (OPG) as an estimate of the information matrix, which we invert to find the asymptotic covariance matrix of the parameters. Numerical integration was performed to evaluate the Bessel functions of the analytical gradient.

The starting values of the estimations were set to the true values of the parameters. Of course, in practice, this is not realistic. Some sensitivity analyses were made, and it was found that the starting values do not matter very much as long as they imply a stationary process. The true values were chosen to make the simulation results more comparable across parameter sets and models.

7.2 Evaluation measures

For all the simulations, a number of evaluation measures were calculated. We present these measures here. For all parameter estimates, the bias was calculated as

$$Bias = \Sigma_{i=1}^{1000} \left(\widehat{\theta}_i - \theta_{True} \right) /M,$$

where $\widehat{\theta}_i$ is the parameter estimate for the i:th replication, θ_{True} is the true value of the parameter, and M is the number of replications. To calculate the standard errors of the parameter estimate, the usual formula is used

$$Se = \sqrt{\frac{\Sigma_{i=1}^{M} \left(\widehat{\theta}_i - \overline{\widehat{\theta}} \right)^2}{M-1}},$$

where $\overline{\widehat{\theta}}$ is the sample mean of the estimates.

The bias measure does not take into account the relative error of the estimate, therefore, we also calculate the Mean Absolute Percentage Error (MAPE) as

$$MAPE = 100 * \frac{\Sigma_{i=1}^{M} \left(\left| \widehat{\theta}_i - \theta_{True} \right| /\theta_{True} \right)}{M}.$$

We also evaluate the empirical size of the 95 % confidence intervals. For each replication, and for each parameter, a confidence interval was calculated as

$$\widehat{\theta}_i \pm 1.96 * \sqrt{Var\left(\widehat{\theta}_i\right)},$$

[6] See Berndt, Hall, Hall and Hausman (1974).

where $Var\left(\widehat{\theta}_i\right)$ is the variance of the estimate of $\widehat{\theta}_i$. In this simulation setup, $Var\left(\widehat{\theta}_i\right)$ is the diagonal element of the (asymptotic) covariance matrix, obtained by inverting the outer product of the gradient. The empirical coverage probability $\widehat{p}$ corresponds to the number of times the confidence interval covers the true value of the parameter, divided by the number of replications. The standard error of the coverage probability is given by the Bernoulli formula $\sqrt{p\left(1-p\right)/M}$, and using this standard error, we calculate a test region of the null that size of the confidence interval is 0.95, that is

$$0.95 \pm 1.96 * \sqrt{0.95\left(1-0.95\right)/M},$$

so, the standard error is calculated under the null. Empirical sizes of the 95% confidence interval of the falling outside of this region are written in boldface in the tables.

To test for normality of the estimates, we perform the Jarque-Bera test (Jarque and Bera, 1987) of normality, which is given by

$$JB = \frac{M}{6}\left(S^2 + \frac{\left(K-3\right)^2}{4}\right),$$

where S is the skewness of the estimates, K is the kurtosis of the estimates and M is the number of replications. Under the null of normality, the JB-statistic is distributed as χ^2_2.

7.3 Results of the simulations

The results of the simulations are found in Tables 7.3 through 7.18.

7.3.1 Results for GARCH-NIG

The results for the GARCH(1,1)-NIG are found in Tables 7.3 to 7.6. For the bias and standard errors there seems to be no difference between the analytical and numerical gradients, with the exception of the parameter $\overline{\alpha}$, Set 2 and 3 and sample size 250. For Set 2, the bias for the numerical gradient is 3.186 against 0.437 for the analytical gradient. Similarly, for Set 3, the bias for the numerical gradient is 3.09, while it is 0.433 for the analytical gradient. For the other parameter sets, sample sizes and parameters, there seems to be no difference in performance.

Turning to the coverage rate for the estimates (Table 7.5), there seems to be no clear pattern in the difference between the analytical and numerical gradients. Even though the empirical size is significantly different from the nominal size, the deviation is not severe in any case.

The normality test of the estimates of the GARCH(1,1)-NIG is presented in Table 7.6. As seen in the table, the value of the JB-test statistic decreases over the sample sizes, but normality cannot be rejected only for the estimates for of ρ_1 and π_1, using the analytical gradient for Set 3, n=4000. Otherwise, the value of the JB-statistic seems to be lower or about the same, for the numerical gradients, exception of some entries, such as the estimates of $\overline{\alpha}$ for Sets 2 and 3 with n=250 and n=500.

7.3.2 Results for T-GARCH-NIG

The results for the T-GARCH(1,1)-NIG are reported in Tables 7.7 to 7.10. As seen in Table 7.7, the bias and standard errors for the parameter estimates of the T-GARCH(1,1)-NIG model are very similar for the estimation using analytical or numerical gradients. This is confirmed in Table 7.8, which reports the MAPE for the parameter estimates of the T-GARCH(1,1)-NIG.

Turning to the confidence interval coverage in Table 7.9, we see that for the estimate of $\overline{\alpha}$, the coverage rate is better for the estimator using the numerical gradients, although the empirical coverage rates for the estimates using analytical gradients are not very far off. For the estimates of ρ_1, the empirical coverage rates are underestimated for the estimator using analytical and numerical gradients. The empirical sizes increase with the sample size, but do not reach the nominal size of 95%. For the estimates of the asymmetry parameter ω_1, the empirical size is lower than the nominal size, at least for Set 1, and for the smaller sample sizes (n=250 and n=500) for Sets 2 and 3. All of the confidence interval coverage of the estimates for π_1 have been underestimated, for all sets of parameters, with the exception of the sample size of n=1000.

The JB-test statistic for the estimates of the T-GARCH(1,1)-NIG is reported in Table 7.10. For the estimates of $\overline{\alpha}$ and ρ_0 we reject normality for all parameter sets and all sample sizes. For the estimate of ρ_1 we reject normality at the one percent level for all settings except for the sample size n=4000 for Sets 2 and 3, using the five percent level we cannot reject the null of normality. For the asymmetry parameter ω_1 we cannot reject normality for the sample size n=4000 for any of the three parameter sets. The overall impression is that, for the smaller sample sizes, the estimates are far from normal, while the for the larger sample sizes, the value of the JB-statistic decreases, i.e., the estimate are, at least, less nonnormal.

7.3.3 Results for AV-GARCH-NIG

The results for the AV-GARCH(1,1)-NIG are given in Tables 7.11 to 7.14. In Table 7.11 the bias and standard errors of the estimates of the AV-GARCH(1,1)-

NIG are reported. For the estimate of $\overline{\alpha}$ we see that the bias is larger for the smaller sample sizes when we use the numerical gradients. For Set 1, n=250, the bias is 3.654 when using the numerical gradients and 0.392 when using analytical gradients, i.e., the values are similar as for the GARCH(1,1)-NIG. For Set 2, sample sizes n=250, 500 and 1000 the standard error of the estimates using numerical gradients are also considerably higher than the estimates that use analytical gradients. For n=250 the standard error is 22.412 and for n=500 it is 10.731, indicating that for small sample sizes it is difficult to estimate the parameter $\overline{\alpha}$.

When estimating π_1, a negative bias is obtained, indicating that we underestimate the parameter. The larger bias of the estimate of $\overline{\alpha}$ using numerical gradients is confirmed in Table 7.12 where the MAPE of the estimates is reported.

The confidence interval coverage reported in Table 7.13 indicates that, at least for Set 3, the estimates using analytical gradients overestimate the nominal size of the confidence interval for the parameter $\overline{\alpha}$. For the estimates of the other parameters, no general pattern can be seen in the differences between analytical and numerical gradients.

The Jarque-Bera test of the normality of the estimates of $\overline{\alpha}$ shows that, for the smaller sample sizes, the estimates using numerical gradients have a higher value of the JB-statistics than the estimates using analytical gradients. For the larger sample sizes, the values are about the same. We reject normality for all the sets and sample sizes for the estimates of $\overline{\alpha}$ and ρ_0. Looking at the estimates of ρ_1 and π_1, for Set 2, we see that we cannot reject normality at the 5 percent level for the estimate of ρ_1, n=4000 using analytical gradients, and of π_1 n=4000 using numerical gradients. For Set 3, we cannot reject normality for ρ_1 for n=1000 and n=2000, both for the estimates using analytical and numerical gradients. This is also true for the estimates of π_1 for n=2000 and n=4000.

7.3.4 Results for TAV-GARCH-NIG

The results for the TAV-GARCH(1,1)-NIG are reported in Tables 7.15 to 7.19. Again, as seen in Table 7.15, we have problems with the estimates of $\overline{\alpha}$ when we use the numerical gradients. For Set 2 and 3 the standard error is much larger than for the estimates using analytical estimates.

There are no differences in the performance with respect to the bias and standard error between the estimate using analytical and numerical gradients. This is confirmed in by the MAPE measure in Table 7.16. Turning to the confidence interval coverage in Table 7.17, we see that the coverage rates of the confidence intervals around $\overline{\alpha}$ are overestimated. The coverage rate of the confidence intervals for ρ_1, ω_1 and π_1 are severely underestimated for all parameter sets and sample sizes. A closer look at this indicate that the reason might be that the OPG was not invertible, and therefore, the standard errors of the estimates are missing values. One might consider scaling the parameters in the estimation procedure in order to avoid singularity of the covariance matrix, however, this was not done in this study.

The JB-test of normality in Table 7.18 indicates that we reject normality for all the parameter sets and sample sizes.

7.4 Concluding remarks

The study indicates that the ML estimator works reasonably well for the larger sample sizes often available in practice, i.e., n=1000 and larger. Normality of the estimates seems to be a problem for the sample sizes studied here, even though we cannot reject normality in some instances. Extrapolating the results, it is possible that the we cannot reject normality for sample sizes of say 5000 or higher. This would correspond to daily data for 20 years.

The confidence interval coverage works well for the GARCH(1,1)-NIG and the T-GARCH(1,1). The poor performance of the coverage rates for the AV-GARCH(1,1) and the TAV-GARCH(1,1) models seems to be a result of that the estimated OPG matrix was not inveertible, which means that we did not get any estimate of the asymptotic covariance matrix. This might be due to numerical problems with the Bessel functions, or a scaling problem. Using other ways of evaluating the Bessel functions, and/ or scale the parameters might improve the situation.

The overall impression of the results of the simulation study is that, for smaller sample sizes, it matters whether we use numerical or analytical gradients. For small sample sizes,the estimates of $\overline{\alpha}$ are very uncertain when we use numerical gradients. When working with sample sizes of n=250 and n=500, one should use the analytical gradients. When working with daily data, this corresponds to one and two years, respectively. If we use weekly data or even lower frequencies, as in Engle (1982), one might want to use the GARCH models to model uncertainty in macro data, collected on, say monthly or even lower frequencies. In these cases, it is preferably to use analytical gradients.

There are reasons to suspect that the numerical procedures used in GAUSS when implementing the Bessel functions, are suboptimal and it would be of interest of implement the simulation study using different software.

7.5 Tables

Table 7.1a:
Parameter sets for the GARCH(1,1)-NIG.

		Set 1	Set 2	Set 3
Parameters	$\overline{\alpha}$	1.0	1.5	1.5
	ρ_0	0.03	0.01	0.0043
	ρ_1	0.070	0.04	0.04665
	ω_1	-	-	-
	π_1	0.90	0.95	0.949
Moment conditions	γ_{c1}	0.9654	0.9900	0.9956
	γ_{c2}	1.0000	0.9865	1.0000
Implied moments	Var	1.0000	1.0000	1.0000
	Kurt	10.2486	7.3704	-

Notes: The table reports the parameter sets that are used in the simulation study for the GARCH(1,1)-NIG model.
γ_{c1} has to be less than one for the unconditional variance to exist for models modeling conditional variance.
γ_{c2} has to be less than one for the unconditional fourth moment to exist for models modeling conditional variance.
- indicates that the quantity does not exist.

Table 7.1b:
Parameter sets for the T-GARCH(1,1)-NIG.

		Set 1	Set 2	Set 3
Parameters	$\overline{\alpha}$	1.0	1.2	1.2
	ρ_0	0.03	0.013	0.01725
	ρ_1	0.05	0.03	0.042
	ω_1	0.04	0.05	0.0715
	π_1	0.9	0.932	0.905
Moment	γ_{c1}	0.970	0.987	0.98275
conditions	γ_{c2}	0.96780	0.99122	1.00003
Implied	Var	1.00000	1.00000	1.00000
moments	Kurt	11.01242	16.17931	-

Notes: The table reports the parameter sets that are used in the simulation study for Threshold-GARCH(1,1)-NIG model.
γ_{c1} has to be less than one for the unconditional variance to exist for models modeling conditional variance.
γ_{c2} has to be less than one for the unconditional fourth moment to exist for models modeling conditional variance.
- indicates that the quantity does not exist.

Table 7.2a:
Parameter sets for the AV-GARCH(1,1).

		Set 1	Set 2	Set 3
Parameters	$\overline{\alpha}$	1.5000	1.5000	1.000
	ρ_0	0.062365	0.0036951	0.0069569
	ρ_1	0.0500	0.081	0.113
	ω_1	-	-	-
	π_1	0.900	0.935	0.90915
Moment	γ_{c1}	0.8797	0.9939	0.9888
conditions	γ_{c2}	0.7776	0.9987	1.0001
Implied	Var	1.000	1.000	1.000
moments	Kurt	5.1813	10.8439	-

Notes: The table reports the parameter sets that are used in the simulation study for the AV-GARCH(1,1)-NIG model.
γ_{c2} has to be less than one for the unconditional variance to exist for models modeling conditional standard deviation.
γ_{c4} has to be less than one for the unconditional fourth moment to exist for models modeling conditional standard deviation.
- indicates that the quantity does not exist.

Table 7.2b:
Parameter sets for the TAV-GARCH(1,1)-NIG.

		Set 1	Set 2	Set 3
Parameters	$\overline{\alpha}$	1.000	1.200	1.500
	ρ_0	0.0521335	0.0121069	0.010308
	ρ_1	0.05	0.032	0.031
	ω_1	0.03	0.0535	0.0689
	π_1	0.90	0.944	0.940
Moment	γ_{c1}	0.89958	0.97706	0.98099
conditions	γ_{c2}	0.85669	0.99183	1.00036
Implied	Var	1.000	1.000	1.000
moments	Kurt	8.35956	37.31482	-

Notes: The table reports the parameter sets that are used in the simulation study for the TAV-GARCH(1,1)-NIG model.
γ_{c2} has to be less than one for the unconditional variance to exist for models modeling conditional standard deviation.
γ_{c4} has to be less than one for the unconditional fourth moment to exist for models modeling conditional standard deviation.
- indicates that the quantity does not exist.

Table 7.3a:
Bias and standard errors for the estimates of the GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.295	0.263	0.126	0.065
		(0.710)	(0.679)	(0.133)	(0.135)
	500	0.156	0.134	0.084	0.023
		(0.432)	(0.432)	(0.060)	(0.057)
	1000	0.080	0.061	0.069	0.009
		(0.249)	(0.260)	(0.025)	(0.025)
	2000	0.054	0.033	0.064	0.004
		(0.160)	(0.167)	(0.012)	(0.012)
	4000	0.039	0.017	0.062	0.002
		(0.107)	(0.111)	(0.008)	(0.008)
Set 2	250	0.437	3.186	0.12	0.093
		(1.034)	(16.504)	(0.208)	(0.206)
	500	0.231	0.815	0.053	0.034
		(0.678)	(7.624)	(0.095)	(0.109)
	1000	0.135	0.114	0.028	0.008
		(0.464)	(0.445)	(0.025)	(0.019)
	2000	0.062	0.066	0.023	0.003
		(0.279)	(0.288)	(0.007)	(0.007)
	4000	0.039	0.03	0.062	0.001
		(0.107)	(0.189)	(0.008)	(0.004)
Set 3	250	0.433	3.09	0.072	0.058
		(1.031)	(16.227)	(0.168)	(0.156)
	500	0.284	0.711	0.025	0.020
		(0.753)	(6.971)	(0.035)	(0.067)
	1000	0.089	0.200	0.013	0.005
		(0.417)	(3.143)	(0.009)	(0.011)
	2000	0.055	0.057	0.010	0.002
		(0.301)	(0.289)	(0.003)	(0.004)
	4000	0.027	0.025	0.009	0.001
		(0.191)	(0.189)	(0.002)	(0.002)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}}\left(\widehat{\theta}_i - \theta_{True}\right)/n$.
Standard errors of the estimates are given in parenthesis.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.3b:
Bias and standard errors for the estimates of the GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.013	0.011	-0.096	-0.091
		(0.057)	(0.055)	(0.201)	(0.199)
	500	0.004	0.002	-0.032	-0.029
		(0.033)	(0.033)	(0.088)	(0.083)
	1000	0.001	0.001	-0.012	-0.011
		(0.022)	(0.021)	(0.039)	(0.039)
	2000	0.000	0.000	-0.005	-0.004
		(0.016)	(0.016)	(0.024)	(0.023)
	4000	0.000	0.000	-0.002	-0.002
		(0.011)	(0.011)	(0.016)	(0.016)
Set 2	250	0.011	0.006	-0.131	-0.146
		(0.044)	(0.042)	(0.259)	(0.284)
	500	0.005	0.003	-0.043	-0.048
		(0.027)	(0.024)	(0.123)	(0.149)
	1000	0.002	0.001	-0.012	-0.010
		(0.014)	(0.014)	(0.048)	(0.029)
	2000	0.001	0.000	-0.005	-0.004
		(0.009)	(0.010)	(0.014)	(0.014)
	4000	0.000	0.000	-0.002	-0.002
		(0.007)	(0.006)	(0.009)	(0.009)
Set 3	250	0.009	0.004	0.109	-0.121
		(0.046)	(0.044)	(0.234)	(0.257)
	500	0.004	0.002	-0.028	-0.036
		(0.025)	(0.024)	(0.074)	(0.118)
	1000	0.001	0.001	-0.008	-0.009
		(0.015)	(0.015)	(0.023)	(0.036)
	2000	0.000	0.000	-0.003	-0.003
		(0.010)	(0.010)	(0.011)	(0.011)
	4000	0.000	0.000	-0.002	-0.002
		(0.007)	(0.007)	(0.007)	(0.007)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$.
Standard errors of the estimates are given in parenthesis.
* means that analytical gradients has been used in the estimation.
** means that numerical gradients has been used in the estimation.

Table 7.4a:
MAPE for the estimates of the GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	48.281	47.915	247.216	244.887
	500	29.996	30.455	106.068	102.761
	1000	18.561	19.036	51.428	51.321
	2000	12.805	13.079	31.536	31.399
	4000	8.795	8.799	20.309	20.26
Set 2	250	51.447	235.743	1035.881	971.92
	500	33.805	73.156	351.188	366.754
	1000	23.411	22.192	102.013	103.614
	2000	14.23	15.026	51.118	50.174
	4000	10.132	10.154	29.177	30.329
Set 3	250	51.532	229.689	1525.338	1379.659
	500	36.996	67.329	411.709	474.692
	1000	21.339	28.823	119.968	134.483
	2000	15.102	15.047	54.357	60.083
	4000	9.992	10.198	32.225	34.801

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=$100 * \Sigma_{i=1}^{\#\mathrm{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients has been used in the estimation.

** means that numerical gradients has been used in the estimation.

Table 7.4b:
MAPE for the estimates of the GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	57.645	55.782	13.13	12.666
	500	37.251	36.552	5.900	5.635
	1000	24.252	24.046	3.063	3.023
	2000	17.652	17.66	2.040	2.018
	4000	12.274	12.262	1.383	1.376
Set 2	250	76.744	73.029	15.316	16.919
	500	48.824	45.44	5.549	6.135
	1000	26.624	27.682	2.047	1.952
	2000	18.783	19.36	1.134	1.109
	4000	12.785	12.491	0.711	0.721
Set 3	250	69.268	66.662	12.976	14.178
	500	40.453	40.342	3.75	4.636
	1000	24.167	25.52	1.528	1.700
	2000	16.511	17.361	0.896	0.951
	4000	11.8	11.048	0.609	0.611

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=$100 * \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients has been used in the estimation.

** means that numerical gradients has been used in the estimation.

Table 7.5a:
Confidence interval coverage for the estimates of the GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.957	0.944	**0.920**	**0.914**
	500	**0.973**	0.953	0.955	0.956
	1000	**0.968**	0.949	0.963	0.962
	2000	**0.970**	**0.963**	0.944	0.947
	4000	**0.966**	0.957	0.942	0.950
Set 2	250	**0.938**	**0.898**	**0.803**	**0.782**
	500	0.948	0.939	0.957	0.946
	1000	0.953	0.940	**0.976**	**0.976**
	2000	0.946	0.952	0.951	0.955
	4000	**0.914**	0.944	**0.921**	0.949
Set 3	250	**0.935**	**0.898**	**0.820**	**0.805**
	500	0.948	**0.933**	**0.965**	0.956
	1000	0.953	**0.932**	**0.990**	**0.979**
	2000	0.953	0.946	**0.976**	**0.968**
	4000	**0.914**	0.945	**0.930**	0.955

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.5b:
Confidence interval coverage for the estimates of the GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.942	0.937	**0.919**	**0.920**
	500	**0.928**	**0.929**	0.954	0.952
	1000	0.941	0.944	0.955	0.957
	2000	**0.921**	**0.925**	**0.933**	0.938
	4000	**0.928**	0.937	**0.926**	**0.927**
Set 2	250	**0.874**	**0.837**	**0.832**	**0.817**
	500	**0.926**	**0.925**	0.941	0.946
	1000	0.945	**0.933**	0.961	0.957
	2000	**0.933**	**0.928**	0.944	**0.890**
	4000	**0.895**	**0.925**	**0.908**	**0.890**
Set 3	250	**0.886**	**0.856**	**0.856**	**0.845**
	500	**0.930**	**0.929**	0.959	0.949
	1000	0.938	**0.924**	0.954	0.948
	2000	**0.926**	**0.926**	0.944	**0.917**
	4000	**0.897**	**0.928**	**0.911**	**0.890**

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.6a:
Jarque-Bera test of normality of the estimates for the GARCH(1,1)-NIG.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	3973.732	1468.649	5548.607	7482.419
		(0.000)	(0.000)	(0.000)	(0.000)
	500	4983.597	2194.165	27033.063	32601.058
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2349.777	1988.368	89621.661	97934.709
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	159.702	138.703	471.494	461.511
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	59.629	61.944	1027.758	1018.317
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	501.571	41028.315	8488.405	12798.794
		(0.000)	(0.000)	(0.000)	(0.000)
	500	514.273	1079915.889	62467.645	158339.825
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	492.538	812.119	3125981.062	319237.276
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	656.818	124.229	1281.519	9357.689
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	101.982	32.679	522.185	230.948
		(0.000)	(0.000)	(0.000)	(0.000)
Set 3	250	408.351	44254.401	96607.25	84767.008
		(0.000)	(0.000)	(0.000)	(0.000)
	500	620.074	1552727.607	103230.768	737038.2
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	160.029	38187409.92	251783.737	1937024.438
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	826.715	120.245	7357.923	6201.675
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	105.143	30.765	430.086	349.687
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.6b:
Jarque-Bera test of normality of the estimates for the GARCH(1,1)-NIG.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	1021.185	1186.504	3118.446	3639.54
		(0.000)	(0.000)	(0.000)	(0.000)
	500	137.681	115.575	20799.922	22904.177
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	35.95	29.126	13074.292	14525.236
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	24.335	23.72	97.461	91.399
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	18.021	17.302	71.973	74.421
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	1332.722	1416.592	1616.659	1126.174
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1591.028	277.698	39870.182	29608.068
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	124.367	58.643	3639030.163	30274.717
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	50.536	34.029	293.228	322.339
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	9.269	9.107	31.384	19.744
		(0.010)	(0.011)	(0.000)	(0.000)
Set 3	250	986.946	1090.835	2902.476	2059.337
		(0.000)	(0.000)	(0.000)	(0.000)
	500	622.782	181.556	272479.491	83524.289
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	170.272	55.867	292335.764	8275547.709
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	81.68	34.677	45.038	60.923
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	3.106	28.945	0.488	11.711
		(0.212)	(0.000)	(0.784)	(0.003)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.7a:
Bias and standard errors for the estimates of the T-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.320	0.336	0.063	0.061
		(0.684)	(1.403)	(0.132)	(0.127)
	500	0.169	0.144	0.022	0.023
		(0.437)	(0.434)	(0.067)	(0.068)
	1000	0.087	0.067	0.007	0.008
		(0.252)	(0.261)	(0.022)	(0.022)
	2000	0.057	0.036	0.003	0.003
		(0.161)	(0.168)	(0.012)	(0.012)
	4000	0.041	0.018	0.001	0.002
		(0.107)	(0.112)	(0.008)	(0.008)
Set 2	250	0.369	0.324	0.057	0.060
		(0.819)	(0.784)	(0.128)	(0.140)
	500	0.207	0.190	0.018	0.018
		(0.561)	(0.573)	(0.057)	(0.057)
	1000	0.103	0.088	0.005	0.005
		(0.336)	(0.348)	(0.013)	(0.013)
	2000	0.061	0.047	0.002	0.002
		(0.208)	(0.215)	(0.006)	(0.006)
	4000	0.038	0.022	0.001	0.001
		(0.135)	(0.140)	(0.004)	(0.004)
Set 3	250	0.376	0.324	0.041	0.042
		(0.849)	(0.797)	(0.089)	(0.090)
	500	0.207	0.187	0.013	0.013
		(0.565)	(0.566)	(0.032)	(0.032)
	1000	0.097	0.089	0.005	0.005
		(0.324)	(0.347)	(0.011)	(0.011)
	2000	0.062	0.047	0.002	0.002
		(0.207)	(0.214)	(0.006)	(0.006)
	4000	0.038	0.022	0.001	0.001
		(0.135)	(0.140)	(0.004)	(0.004)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$. Standard errors of the estimates are in parenthesis.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.7b:
Bias and standard errors for the estimates of the T-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	-0.003	-0.003	0.029	0.028	-0.091	-0.086
		(0.051)	(0.051)	(0.085)	(0.085)	(0.199)	(0.194)
	500	-0.004	-0.004	0.011	0.011	-0.028	-0.028
		(0.034)	(0.034)	(0.050)	(0.051)	(0.095)	(0.095)
	1000	-0.002	-0.002	0.004	0.004	-0.009	-0.009
		(0.024)	(0.024)	(0.032)	(0.032)	(0.037)	(0.037)
	2000	-0.001	-0.001	0.001	0.001	-0.004	-0.004
		(0.017)	(0.017)	(0.024)	(0.024)	(0.023)	(0.023)
	4000	-0.001	0.000	0.001	0.001	-0.002	-0.002
		(0.012)	(0.012)	(0.016)	(0.016)	(0.016)	(0.016)
Set 2	250	0.001	0.001	0.020	0.020	-0.086	-0.088
		(0.040)	(0.040)	(0.080)	(0.080)	(0.192)	(0.196)
	500	-0.001	-0.001	0.006	0.006	-0.024	-0.024
		(0.027)	(0.027)	(0.042)	(0.042)	(0.085)	(0.085)
	1000	-0.001	-0.001	0.003	0.003	-0.007	-0.007
		(0.018)	(0.018)	(0.027)	(0.027)	(0.026)	(0.026)
	2000	-0.001	-0.001	0.001	0.001	-0.002	-0.002
		(0.012)	(0.012)	(0.019)	(0.019)	(0.015)	(0.015)
	4000	0.000	0.000	0.001	0.001	-0.002	-0.002
		(0.008)	(0.008)	(0.013)	(0.013)	(0.010)	(0.010)
Set 3	250	0.000	0.001	0.020	0.020	-0.069	-0.070
		(0.048)	(0.047)	(0.089)	(0.090)	(0.166)	(0.166)
	500	-0.002	-0.002	0.006	0.007	-0.018	-0.018
		(0.032)	(0.032)	(0.053)	(0.053)	(0.063)	(0.063)
	1000	-0.001	-0.001	0.004	0.004	-0.006	-0.006
		(0.022)	(0.022)	(0.034)	(0.034)	(0.027)	(0.027)
	2000	-0.001	-0.001	0.001	0.001	-0.002	-0.002
		(0.015)	(0.015)	(0.024)	(0.024)	(0.018)	(0.018)
	4000	0.000	0.000	0.002	0.002	-0.001	-0.002
		(0.010)	(0.010)	(0.017)	(0.017)	(0.012)	(0.012)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$. Standard errors of the estimates are given in parenthesis.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.8a:
MAPE for the estimates of the T-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	49.299	53.661	242.265	234.328
	500	30.68	30.915	105.194	106.522
	1000	18.877	19.243	47.257	47.596
	2000	12.891	13.129	30.212	30.368
	4000	8.871	8.844	19.725	19.756
Set 2	250	50.276	48.853	468.063	492.864
	500	32.915	33.488	161.361	161.623
	1000	20.324	20.737	62.392	62.412
	2000	13.671	14.028	34.327	34.45
	4000	9.093	9.244	23.100	23.359
Set 3	250	51.181	48.918	267.023	272.725
	500	32.998	33.283	97.33	97.726
	1000	19.935	20.763	46.516	46.735
	2000	13.667	13.942	28.073	28.32
	4000	9.072	9.228	19.463	19.562

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=$100 * \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.8b:
MAPE for the estimates of the T-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	78.632	78.602	144.923	144.281	12.833	12.420
	500	55.554	55.790	94.602	95.085	5.848	5.878
	1000	38.753	38.822	64.242	64.325	2.918	2.913
	2000	26.186	26.254	47.663	47.814	1.994	1.993
	4000	18.887	19.036	32.437	32.531	1.358	1.361
Set 2	250	99.932	99.721	106.119	106.876	10.959	11.185
	500	72.513	72.959	64.337	64.195	4.217	4.211
	1000	50.458	50.296	43.396	43.269	1.973	1.970
	2000	33.056	33.231	30.474	30.757	1.251	1.256
	4000	22.302	22.582	20.785	20.866	0.850	0.851
Set 3	250	87.211	86.317	89.654	90.127	10.071	10.109
	500	62.333	62.489	57.02	57.029	4.217	4.210
	1000	43.352	43.175	37.747	37.775	2.292	2.286
	2000	28.372	28.513	27.245	27.460	1.535	1.543
	4000	19.997	20.066	18.438	18.530	1.076	1.078

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE$=100 * \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.9a:
Confidence interval coverage for the estimates of the T-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	**0.964**	0.949	**0.906**	**0.909**
	500	**0.975**	0.959	0.939	0.940
	1000	**0.971**	0.953	0.950	0.950
	2000	**0.969**	**0.964**	0.941	0.945
	4000	0.962	0.959	**0.932**	0.943
Set 2	250	0.954	0.940	**0.908**	**0.908**
	500	**0.972**	0.953	0.958	0.957
	1000	**0.964**	0.948	0.954	0.956
	2000	**0.966**	0.958	0.946	0.948
	4000	0.955	0.958	0.940	0.957
Set 3	250	0.955	0.943	**0.927**	**0.930**
	500	**0.971**	0.954	0.962	**0.963**
	1000	0.960	0.951	0.954	0.959
	2000	**0.965**	0.956	0.940	0.948
	4000	0.956	0.958	0.937	0.953

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.9b:
Confidence interval coverage for the estimates of the T-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.*	Anal.*	Num.**
Set 1	250	**0.734**	**0.737**	**0.707**	**0.702**	**0.907**	**0.908**
	500	**0.858**	**0.859**	**0.773**	**0.774**	**0.924**	**0.925**
	1000	**0.895**	**0.897**	**0.868**	**0.869**	0.951	0.951
	2000	**0.927**	**0.929**	**0.933**	0.937	**0.927**	**0.930**
	4000	**0.928**	**0.931**	0.939	0.952	**0.915**	**0.919**
Set 2	250	**0.640**	**0.646**	**0.775**	**0.771**	**0.911**	**0.913**
	500	**0.777**	**0.774**	**0.875**	**0.879**	**0.934**	**0.933**
	1000	**0.893**	**0.895**	0.946	0.948	0.952	0.955
	2000	**0.924**	**0.925**	0.942	0.945	**0.923**	**0.919**
	4000	**0.923**	0.938	**0.928**	0.947	**0.922**	**0.917**
Set 3	250	**0.712**	**0.713**	**0.807**	**0.809**	**0.917**	**0.924**
	500	**0.848**	**0.846**	**0.912**	**0.913**	**0.933**	**0.933**
	1000	**0.900**	**0.912**	0.944	0.951	0.945	0.952
	2000	**0.929**	**0.932**	0.935	0.944	**0.919**	**0.920**
	4000	**0.914**	**0.931**	**0.930**	0.947	**0.914**	**0.923**

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.10a:
Jarque-Bera test of normality of the T-GARCH(1,1)-NIG estimates.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	1505.769	12938544.52	5385.185	5119.504
		(0.000)	(0.000)	(0.000)	(0.000)
	500	3637.52	1966.595	126284.35	124300.707
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	1990.068	1707.348	51417.779	52761.625
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	199.208	181.903	481.084	478.098
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	65.333	65.125	373.04	387.965
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	677.302	486.627	20235.625	41103.788
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1947.094	2098.517	735166.867	732105.82
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	4884.127	6381.39	76334.054	75521.437
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	120.521	113.18	520.619	538.485
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	39.127	45.436	132.784	184.095
		(0.000)	(0.000)	(0.000)	(0.000)
Set 3	250	893.872	696.525	10251.69	9614.055
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1840.787	1661.639	142854.585	142486.598
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	3835.804	6321.941	9939.746	9747.199
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	130.939	121.111	296.078	296.522
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	44.896	45.858	111.603	108.15
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.10b:
Jarque-Bera test of normality of the T-GARCH(1,1)-NIG estimates.

		ρ_1		ω_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	1127.464	1130.071	9600.804	9783.384
		(0.000)	(0.000)	(0.000)	(0.000)
	500	89.653	96.815	1201.606	1198.868
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	13.651	13.258	108.211	110.391
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	25.861	26.335	15.410	15.347
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	19.727	18.443	1.133	1.182
		(0.000)	(0.000)	(0.568)	(0.554)
Set 2	250	2463.501	2455.566	14092.905	13873.305
		(0.000)	(0.000)	(0.000)	(0.000)
	500	241.703	258.036	1291.83	1356.464
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	34.995	34.868	36.007	37.710
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	25.258	23.639	5.010	3.960
		(0.000)	(0.000)	(0.082)	(0.000)
	4000	7.653	6.257	0.145	0.064
		(0.022)	(0.044)	(0.930)	(0.969)
Set 3	250	1077.273	1100.488	5203.454	4957.619
		(0.000)	(0.000)	(0.000)	(0.000)
	500	107.868	106.754	451.006	464.913
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	22.985	20.664	18.881	19.955
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	21.895	22.065	3.754	3.358
		(0.000)	(0.000)	(0.153)	(0.187)
	4000	7.212	6.484	0.172	0.190
		(0.027)	(0.039)	(0.918)	(0.909)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.10c:
Jarque-Bera test of normality of the T-GARCH(1,1)-NIG estimates.

		π_1	
Par. Set	n	Anal.*	Num.**
Set 1	250	2991.384	3393.666
		(0.000)	(0.000)
	500	59153.96	57327.495
		(0.000)	(0.000)
	1000	6893.354	7083.597
		(0.000)	(0.000)
	2000	169.379	170.055
		(0.000)	(0.000)
	4000	35.055	35.464
		(0.000)	(0.000)
Set 2	250	4774.534	4453.25
		(0.000)	(0.000)
	500	160580.479	159811.827
		(0.000)	(0.000)
	1000	4279.027	4264.638
		(0.000)	(0.000)
	2000	66.392	66.149
		(0.000)	(0.000)
	4000	6.902	7.123
		(0.032)	(0.028)
Set 3	250	7150.088	7089.334
		(0.000)	(0.000)
	500	103416.45	104460.789
		(0.000)	(0.000)
	1000	405.65	415.114
		(0.000)	(0.000)
	2000	36.061	35.994
		(0.000)	(0.000)
	4000	4.596	4.265
		(0.100)	(0.119)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.11a:
Bias and standard errors for the estimates of the AV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.392	3.654	0.184	0.179
		(0.962)	(0.298)	(0.296)	(17.679)
	500	0.285	0.835	0.117	0.101
		(0.727)	(7.627)	(0.225)	(0.218)
	1000	0.143	0.320	0.046	0.048
		(0.458)	(4.421)	(0.132)	(0.124)
	2000	0.063	0.070	0.014	0.020
		(0.287)	(0.286)	(0.046)	(0.061)
	4000	0.038	0.033	0.008	0.007
		(0.192)	(0.189)	(0.03)	(0.029)
Set 2	250	0.430	5.721	0.062	0.044
		(1.06)	(22.412)	(0.129)	(0.105)
	500	0.224	1.381	0.023	0.014
		(0.727)	(10.731)	(0.04)	(0.039)
	1000	0.103	0.288	0.013	0.005
		(0.458)	(4.421)	(0.007)	(0.007)
	2000	0.051	0.048	0.010	0.002
		(0.275)	(0.282)	(0.003)	(0.003)
	4000	0.025	0.02	0.009	-0.004
		(0.193)	(0.185)	(0.002)	(0.002)
Set 3	250	0.301	0.969	0.054	0.040
		(0.713)	(8.265)	(0.095)	(0.096)
	500	0.138	0.213	0.026	0.011
		(0.426)	(3.156)	(0.022)	(0.022)
	1000	0.071	0.049	0.018	0.004
		(0.250)	(0.258)	(0.007)	(0.007)
	2000	0.044	0.024	0.016	0.002
		(0.160)	(0.166)	(0.004)	(0.004)
	4000	0.035	0.013	0.015	0.001
		(0.106)	(0.111)	(0.002)	(0.002)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$.
Standard errors of the estimates are given in parenthesis.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.11b:
Bias and standard errors for the estimates of the AV-GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.023	0.015	-0.206	-0.228
		(0.057)	(0.057)	(0.320)	(0.340)
	500	0.013	0.007	-0.128	-0.113
		(0.039)	(0.036)	(0.239)	(0.237)
	1000	0.003	0.003	-0.048	-0.052
		(0.026)	(0.026)	(0.141)	(0.137)
	2000	0.001	0.002	-0.015	-0.021
		(0.017)	(0.019)	(0.054)	(0.071)
	4000	0.001	0.001	-0.009	-0.007
		(0.013)	(0.012)	(0.037)	(0.036)
Set 2	250	0.008	-0.002	-0.086	-0.120
		(0.051)	(0.051)	(0.184)	(0.252)
	500	0.004	0.002	0.024	-0.034
		(0.029)	(0.030)	(0.064)	(0.117)
	1000	0.003	0.002	-0.009	-0.010
		(0.020)	(0.020)	(0.018)	(0.045)
	2000	0.002	0.001	-0.004	-0.004
		(0.014)	(0.014)	(0.011)	(0.011)
	4000	0.001	0.001	-0.002	-0.002
		(0.009)	(0.009)	(0.007)	(0.007)
Set 3	250	0.008	0.004	-0.065	-0.066
		(0.059)	(0.058)	(0.135)	(0.151)
	500	0.004	0.003	-0.02	-0.019
		(0.035)	(0.034)	(0.04)	(0.046)
	1000	0.002	0.002	-0.008	-0.007
		(0.024)	(0.023)	(0.02)	(0.020)
	2000	0.001	0.001	-0.004	-0.003
		(0.017)	(0.017)	(0.014)	(0.014)
	4000	0.001	0.001	-0.002	-0.002
		(0.011)	(0.011)	(0.009)	(0.009)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\mathrm{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$.
Standard errors of the estimates are given in parenthesis.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.12a:
MAPE for the estimates of the AV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	47.345	266.844	331.621	327.306
	500	35.542	74.111	218.13	197.1
	1000	22.308	35.293	108.489	107.701
	2000	14.65	14.900	50.216	59.336
	4000	10.038	10.144	34.304	33.963
Set 2	250	51.126	404.821	1500.261	1227.429
	500	34.377	112.451	418.176	399.153
	1000	21.907	34.985	145.019	142.97
	2000	14.181	14.664	67.697	67.453
	4000	10.19	10.001	36.176	37.273
Set 3	250	49.235	118.702	594.736	590.287
	500	29.389	39.815	179.247	176.459
	1000	18.356	18.83	78.004	77.384
	2000	12.591	13.015	42.673	42.499
	4000	8.659	8.79	25.619	25.397

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=100 $* \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.12b:
MAPE for the estimates of the AV-GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	89.759	85.67	26.475	28.686
	500	61.112	57.153	17.163	15.901
	1000	40.287	40.406	8.661	8.746
	2000	27.06	29.702	4.254	5.002
	4000	19.808	19.666	2.980	2.933
Set 2	250	48.115	48.661	10.488	14.14
	500	28.891	29.232	3.364	4.423
	1000	19.708	19.391	1.626	1.773
	2000	13.404	13.273	1.012	0.995
	4000	8.615	8.518	0.620	0.627
Set 3	250	39.274	38.661	8.558	8.879
	500	24.166	23.717	3.244	3.209
	1000	16.802	16.556	1.845	1.791
	2000	11.951	11.818	1.252	1.228
	4000	7.770	7.756	0.815	0.809

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=100 $* \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.13a:
Confidence interval coverage for estimates of the
the AV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.955	0.945	**0.804**	0.955
	500	0.953	0.941	**0.909**	**0.890**
	1000	**0.964**	0.946	**0.907**	**0.920**
	2000	0.945	0.955	**0.902**	**0.914**
	4000	**0.930**	0.950	**0.897**	**0.888**
Set 2	250	0.942	**0.875**	**0.893**	**0.840**
	500	0.952	**0.924**	**0.982**	**0.969**
	1000	0.941	**0.934**	**0.986**	**0.989**
	2000	0.958	0.952	0.947	0.955
	4000	**0.927**	0.945	**0.932**	0.945
Set 3	250	0.961	0.936	0.949	0.937
	500	**0.972**	0.948	**0.990**	**0.989**
	1000	**0.969**	0.945	**0.975**	**0.977**
	2000	**0.969**	0.958	0.960	0.962
	4000	**0.969**	0.957	0.952	0.960

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.13b:
Confidence interval coverage for the estimates of the AV-GARCH(1,1)-NIG model.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	**0.873**	**0.928**	**0.802**	**0.890**
	500	**0.925**	**0.928**	**0.903**	**0.890**
	1000	**0.928**	0.941	**0.916**	**0.926**
	2000	**0.928**	**0.921**	**0.916**	**0.905**
	4000	**0.900**	**0.929**	**0.898**	**0.876**
Set 2	250	**0.929**	**0.871**	**0.909**	**0.862**
	500	0.946	**0.933**	0.963	0.954
	1000	**0.920**	**0.920**	0.945	0.943
	2000	**0.902**	**0.904**	**0.912**	**0.912**
	4000	**0.926**	**0.907**	**0.922**	**0.870**
Set 3	250	**0.928**	**0.921**	0.940	0.942
	500	0.935	0.939	**0.963**	**0.968**
	1000	0.936	0.939	0.959	**0.966**
	2000	**0.922**	**0.924**	**0.932**	0.936
	4000	**0.932**	**0.931**	**0.925**	**0.922**

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.14a:
Jarque-Bera test of normality of the estimates of AV-GARCH(1,1)-NIG.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	568.306	29257.825	363.696	388.165
		(0.000)	(0.000)	(0.000)	(0.000)
	500	929.736	1074865.852	1651.825	2430.822
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	930.07	9865338.702	17008.999	15883.476
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	352.41	141.668	5197.055	18206.14
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	99.098	36.305	4464.821	4537.384
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	653.846	9774.879	90377.277	60248.612
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1207.166	264478.163	628395.795	764739.791
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2769.45	9878765.061	10813.671	11186.439
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	149.592	119.608	8238.964	6903.283
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	41.46	27.909	257.025	227.875
		(0.000))	(0.000)	(0.000))	(0.000)
Set 3	250	2263.598	791782.208	496086.056	477734.971
		(0.000)	(0.000)	(0.000)	(0.000)
	500	5071.627	38490315.32	289672.311	298866.721
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2241.127	1971.27	2224.514	2283.252
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	125.002	112.037	2210.671	1868.274
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	54.246	53.176	443.205	476.906
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.14b:
Jarque-Bera test of normality of the estimates of AV-GARCH(1,1)-NIG.

		ρ_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	286.499	447.019	258.77	211.482
		(0.000)	(0.000)	(0.000)	(0.000)
	500	194.566	145.918	1123.115	1559.568
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	308.483	126.865	12052.218	9697.554
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	37.877	70.978	2724.796	11713.339
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	25.648	8.392	2650.436	2049.818
		(0.000)	(0.015)	(0.000)	(0.000)
Set 2	250	235.001	144.451	6981.872	2174.832
		(0.000)	(0.000)	(0.000)	(0.000)
	500	39.841	16.622	459987.623	99628.054
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	9.691	13.131	208.695	5326826.096
		(0.008)	(0.001)	(0.000)	(0.000)
	2000	13.079	10.727	11.28	8.912
		(0.001)	(0.005)	(0.004)	(0.012)
	4000	5.799	13.158	17.291	3.952
		(0.055)	(0.001)	(0.000)	(0.139)
Set 3	250	181.461	137.745	16563.94	14900.249
		(0.000)	(0.000)	(0.000)	(0.000)
	500	34.439	24.156	28271.227	378850.644
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2.113	0.363	43.145	32.255
		(0.348)	(0.834)	(0.000)	(0.000)
	2000	2.330	1.757	3.152	2.317
		(0.312)	(0.415)	(0.207)	(0.314)
	4000	9.604	8.513	4.771	3.735
		(0.008)	(0.014)	(0.092)	(0.155)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.15a:
Bias and standard errors for the estimates of the TAV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	0.296	0.282	0.237	0.235
		(0.668)	(0.691)	(0.318)	(0.320)
	500	0.168	0.141	0.156	0.158
		(0.440)	(0.433)	(0.261)	(0.266)
	1000	0.089	0.069	0.109	0.113
		(0.253)	(0.262)	(0.202)	(0.212)
	2000	0.059	0.038	0.074	0.072
		(0.161)	(0.168)	(0.147)	(0.142)
	4000	0.042	0.019	0.039	0.039
		(0.106)	(0.112)	(0.068)	(0.069)
Set 2	250	0.353	0.416	0.180	0.184
		(0.825)	(3.212)	(0.281)	(0.286)
	500	0.215	0.196	0.103	0.109
		(0.574)	(0.575)	(0.191)	(0.205)
	1000	0.104	0.095	0.058	0.057
		(0.329)	(0.349)	(0.115)	(0.105)
	2000	0.065	0.052	0.035	0.036
		(0.206)	(0.214)	(0.052)	(0.053)
	4000	0.039	0.024	0.026	0.026
		(0.134)	(0.140)	(0.019)	(0.019)
Set 3	250	0.382	0.441	0.153	0.147
		(0.965)	(2.525)	(0.247)	(0.247)
	500	0.250	0.286	0.092	0.089
		(0.682)	(1.449)	(0.172)	(0.170)
	1000	0.136	0.128	0.051	0.050
		(0.445)	(0.466)	(0.088)	(0.088)
	2000	0.077	0.067	0.034	0.033
		(0.282)	(0.292)	(0.037)	(0.031)
	4000	0.038	0.026	0.028	0.028
		(0.181)	(0.194)	(0.016)	(0.017)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$.
Standard errors of the estimates are given in parenthesis.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.15b:
Bias and standard errors for the estimates of the TAV-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	-0.013	-0.015	0.020	0.018	-0.241	-0.236
		(0.053)	(0.051)	(0.062)	(0.062)	(0.343)	(0.344)
	500	-0.021	-0.022	0.004	0.003	-0.145	-0.144
		(0.035)	(0.034)	(0.043)	(0.042)	(0.277)	(0.279)
	1000	-0.023	-0.024	-0.005	-0.005	-0.091	-0.094
		(0.027)	(0.026)	(0.028)	(0.028)	(0.213)	(0.220)
	2000	-0.025	-0.026	-0.010	-0.010	-0.053	-0.05
		(0.020)	(0.020)	(0.020)	(0.020)	(0.157)	(0.151)
	4000	-0.025	-0.026	-0.013	-0.013	-0.016	-0.016
		(0.014)	(0.014)	(0.014)	(0.014)	(0.075)	(0.076)
Set 2	250	-0.009	-0.009	-0.003	-0.005	-0.178	-0.181
		(0.043)	(0.042)	(0.056)	(0.058)	(0.304)	(0.309)
	500	-0.013	-0.013	-0.015	-0.015	-0.090	-0.095
		(0.027)	(0.026)	(0.037)	(0.037)	(0.207)	(0.219)
	1000	-0.016	-0.016	-0.021	-0.021	-0.040	-0.038
		(0.019)	(0.019)	(0.024)	(0.024)	(0.123)	(0.114)
	2000	-0.017	-0.017	-0.024	-0.024	-0.014	-0.014
		(0.013)	(0.013)	(0.017)	(0.017)	(0.059)	(0.060)
	4000	-0.016	-0.017	-0.025	-0.025	-0.004	-0.005
		(0.009)	(0.009)	(0.011)	(0.011)	(0.024)	(0.024)
Set 3	250	-0.009	-0.009	-0.014	-0.017	-0.147	-0.138
		(0.040)	(0.043)	(0.054)	(0.055)	(0.274)	(0.272)
	500	-0.013	-0.013	-0.025	-0.026	-0.074	-0.071
		(0.027)	(0.027)	(0.038)	(0.037)	(0.189)	(0.187)
	1000	-0.015	-0.016	-0.030	-0.031	-0.029	-0.027
		(0.018)	(0.018)	(0.024)	(0.024)	(0.097)	(0.097)
	2000	-0.016	-0.016	-0.033	-0.032	-0.009	-0.009
		(0.013)	(0.013)	(0.017)	(0.016)	(0.044)	(0.038)
	4000	-0.016	-0.016	-0.033	-0.033	-0.003	-0.003
		(0.009)	(0.009)	(0.011)	(0.011)	(0.020)	(0.021)

Notes: The table reports the bias and the standard and error of the estimates for all replicates. Bias is calculated as $\Sigma_{i=1}^{\#\text{Rep}} \left(\widehat{\theta}_i - \theta_{True}\right)/n$. Standard errors of the estimates are given in parenthesis.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.16a:
MAPE for the estimates of the TAV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	47.362	48.449	485.467	482.85
	500	30.310	30.388	331.34	333.847
	1000	18.882	19.190	233.726	241.165
	2000	12.989	13.181	164.684	160.239
	4000	8.836	8.854	89.006	89.18
Set 2	250	48.774	56.27	1518.312	1547.913
	500	33.158	33.188	863.704	915.314
	1000	20.087	20.66	487.244	474.182
	2000	13.581	14.041	293.881	296.211
	4000	8.996	9.254	213.787	215.416
Set 3	250	47.260	53.051	1509.051	1446.817
	500	33.538	38.231	898.542	868.599
	1000	21.958	23.442	496.164	482.845
	2000	14.722	14.942	324.168	316.668
	4000	9.617	10.282	266.089	264.194

Notes: The table reports the Mean Absolute Percentage error of the estimates, MAPE=$100 * \Sigma_{i=1}^{\#\text{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.16b:
MAPE for the estimates of the TAV-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	89.442	88.234	148.745	146.628	30.613	30.225
	500	71.162	71.301	105.351	104.05	20.370	20.360
	1000	61.897	62.208	74.891	74.942	13.993	14.370
	2000	57.336	57.795	62.291	62.157	9.816	9.533
	4000	52.436	52.584	53.843	54.114	5.124	5.155
Set 2	250	104.346	104.345	79.815	81.544	21.065	21.446
	500	80.708	79.969	61.808	62.357	11.451	12.034
	1000	66.858	66.966	51.737	51.863	6.008	5.830
	2000	58.896	59.262	48.999	48.861	3.227	3.238
	4000	53.335	53.429	47.769	47.889	1.767	1.773
Set 3	250	103.536	106.785	64.727	66.494	17.914	17.169
	500	82.481	82.187	55.954	56.288	9.964	9.577
	1000	67.338	67.98	49.311	50.100	4.943	4.776
	2000	59.209	59.147	48.123	46.996	2.662	2.552
	4000	53.531	53.15	47.683	48.192	1.576	1.605

Notes: The table reports the Mean Absolute Percentage error of the estimates,
MAPE$=100 * \Sigma_{i=1}^{\#\mathrm{Rep}} \left(\left| \widehat{\theta}_i - \theta_{True} \right| / \theta_{True} \right) / n$.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.17a:
Confidence interval coverage for the estimates of the TAV-GARCH(1,1)-NIG model.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	**0.969**	0.949	**0.786**	**0.786**
	500	**0.975**	**0.963**	**0.860**	**0.864**
	1000	**0.974**	0.949	**0.909**	**0.904**
	2000	**0.970**	0.962	**0.910**	**0.913**
	4000	**0.974**	**0.963**	0.950	0.942
Set 2	250	0.963	0.939	**0.794**	**0.792**
	500	**0.971**	0.959	**0.928**	**0.917**
	1000	**0.964**	0.953	**0.970**	**0.975**
	2000	**0.967**	0.958	0.951	0.950
	4000	0.961	0.959	**0.803**	**0.791**
Set 3	250	0.958	**0.936**	**0.843**	**0.825**
	500	0.963	**0.926**	0.939	0.937
	1000	0.954	0.957	**0.977**	**0.979**
	2000	**0.969**	0.955	**0.902**	**0.893**
	4000	0.946	**0.923**	**0.503**	**0.489**

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval $0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.17b:
Confidence interval coverage for the estimates of the
TAV-GARCH(1,1)-NIG model.

		ρ_1		ω_1		π_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.*	Anal.*	Num.**
Set 1	250	**0.518**	**0.516**	**0.643**	**0.635**	**0.764**	**0.766**
	500	**0.629**	**0.630**	**0.642**	**0.642**	**0.828**	**0.832**
	1000	**0.672**	**0.663**	**0.717**	**0.713**	**0.880**	**0.873**
	2000	**0.600**	**0.602**	**0.748**	**0.744**	**0.862**	**0.864**
	4000	**0.462**	**0.456**	**0.742**	**0.740**	**0.900**	**0.890**
Set 2	250	**0.407**	**0.417**	**0.706**	**0.688**	**0.786**	**0.784**
	500	**0.538**	**0.545**	**0.732**	**0.734**	**0.891**	**0.879**
	1000	**0.636**	**0.635**	**0.739**	**0.745**	**0.916**	**0.919**
	2000	**0.623**	**0.625**	**0.583**	**0.589**	**0.895**	**0.894**
	4000	**0.507**	**0.519**	**0.356**	**0.374**	**0.906**	**0.895**
Set 3	250	**0.409**	**0.394**	**0.755**	**0.728**	**0.830**	**0.816**
	500	**0.510**	**0.517**	**0.730**	**0.736**	**0.891**	**0.895**
	1000	**0.629**	**0.642**	**0.657**	**0.638**	**0.922**	**0.915**
	2000	**0.648**	**0.653**	**0.458**	**0.457**	**0.900**	**0.890**
	4000	**0.532**	**0.527**	**0.196**	**0.184**	**0.896**	**0.834**

Notes: The table reports the empirical coverage of the 95% confidence intervals of the estimates. Empirical sizes that are outside the inteval
$0.95 \pm 1.96\sqrt{0.95\,(1-0.95)\,/1000}$ are written in bold face.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.18a:
Jarque-Bera test of normality of the TAV-GARCH(1,1)-NIG estimates.

		$\overline{\alpha}$		ρ_0	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	1447.561	1569.101	205.903	210.496
		(0.000)	(0.000)	(0.000)	(0.000)
	500	6581.475	2965.17	751.162	764.677
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2612.359	2288.417	2812.351	2677.011
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	153.244	140.174	9635.963	9971.776
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	57.634	61.209	68437.056	67459.699
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	1033.305	32346619.74	771.613	704.236
		(0.000)	(0.000)	(0.000)	(0.000)
	500	2609.844	2690.202	5142.781	4474.377
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	4993.501	12193.942	50211.155	46694.094
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	127.309	111.944	211389.889	208236.671
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	47.711	45.134	5917.551	5427.252
		(0.000)	(0.000)	(0.000)	(0.000)
Set 3	250	535.428	1551125.494	1336.522	1574.512
		(0.000)	(0.000)	(0.000)	(0.000)
	500	844.85	14042738.86	9202.383	8348.255
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	2052.968	511.515	105628.189	96096.895
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	152.200	437.788	473681.242	52791.561
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	51.534	35.559	3403.874	17406.364
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.

* means that analytical gradients.has been used in the estimation.

** means that numerical gradients.has been used in the estimation.

Table 7.18b:
Jarque-Bera test of normality of the TAV-GARCH(1,1)-NIG estimates.

		ρ_1		ω_1	
Par. Set	n	Anal.*	Num.**	Anal.*	Num.**
Set 1	250	1322.88	1482.811	1421.24	1565.788
		(0.000)	(0.000)	(0.000)	(0.000)
	500	537.826	594.594	1205.51	1188.517
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	513.069	459.987	704.632	1201.169
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	360.428	380.115	301.021	301.536
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	58.089	59.912	236.641	240.233
		(0.000)	(0.000)	(0.000)	(0.000)
Set 2	250	4830.932	4431.961	1473.634	2323.245
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1200.321	1521.435	1189.421	1171.199
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	634.157	662.885	468.975	532.009
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	451.407	431.21	159.859	143.243
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	68.958	58.127	56.069	56.340
		(0.000)	(0.000)	(0.000)	(0.000)
Set 3	250	4039.119	7929.707	1062.129	1633.326
		(0.000)	(0.000)	(0.000)	(0.000)
	500	1826.554	2140.778	723.273	539.531
		(0.000)	(0.000)	(0.000)	(0.000)
	1000	581.554	785.558	432.78	172.582
		(0.000)	(0.000)	(0.000)	(0.000)
	2000	541.859	138.891	46.802	85.29
		(0.000)	(0.000)	(0.000)	(0.000)
	4000	30.612	22.666	51.054	14.579
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Table 7.18c:
Jarque-Bera test of normality of the TAV-GARCH(1,1)-NIG estimates.

		π_1	
Par. Set	n	Anal.*	Num.**
Set 1	250	168.554	175.178
		(0.000)	(0.000)
	500	602.186	601.999
		(0.000)	(0.000)
	1000	2206.974	2050.645
		(0.000)	(0.000)
	2000	7630.818	7735.419
		(0.000)	(0.000)
	4000	42781.77	42790.59
		(0.000)	(0.000)
Set 2	250	519.559	487.394
		(0.000)	(0.000)
	500	4160.307	3579.114
		(0.000)	(0.000)
	1000	36086.466	34148.855
		(0.000)	(0.000)
	2000	119513.933	118650.626
		(0.000)	(0.000)
	4000	3339.034	3053.212
		(0.000)	(0.000)
Set 3	250	901.308	1087.704
		(0.000)	(0.000)
	500	7325.622	6403.841
		(0.000)	(0.000)
	1000	67060.403	68027.858
		(0.000)	(0.000)
	2000	202894.719	21400.02
		(0.000)	(0.000)
	4000	1485.418	7109.667
		(0.000)	(0.000)

Notes: The table reports the Jarque-Bera test statistic of normality, in parenthesis the p-value of the test.
* means that analytical gradients.has been used in the estimation.
** means that numerical gradients.has been used in the estimation.

Remarks and further work

There is much to do in the area of financial econometrics. The intraday data literature is still in its cradle, screaming for attention. The use of realized volatility is a vibrant area, and one can only speculate on which direction this line of research is going.

One obvious thing to do in the area of realized volatility is to model the realized volatility directly for the purpose of asset pricing and risk management. Some steps have been taken in this direction in ABDL (2001), where they assume that the realized volatility follows a lognormal distribution, (they take logs of the RV and model it with an ARMA(p,q)-model). Of course, one could model the RV directly by assuming other distributions, such as the generalized inverse Gaussian family of distributions. In addition, one might introduce long memory models for the realized volatility. Barndorff-Nielsen and Shephard (2001b, 2002b) discuss the asymptotic distribution of the measurement error of the realized volatility, but what about the small sample properties of the error? An important task for future research is to find ways to deal with the measurement error in a "good" way. Future models for realized volatility would need to take into account the sampling error of the realized volatility.

In the GARCH-models, we model the conditional variance, but there is reason to believe that higher moments of the returns, such as the fourth moment, are also time varying. As in the NIGSV(p,q) model of Andersson (2001), we need to develop models to take into account the possible variation in the higher moments. Maybe we could build a model where we specify a separate "GARCH-equation" for the fourth moment to let the (conditional) second moment follow one process, and the conditional fourth moment another process.

When dealing with asymmetry in the conditional variance, one could develop a News Impact Curve (NIC) in the same spirit as the NIC of Engle and Ng (1993), but where we use the realized volatility instead of the squared daily returns. This would also give rise to new tests for different types of asymmetry, such as shift and rotation asymmetry. Maybe it is possible to use the realized volatiltiy to shed some light on the origins of the asymmetry.

Can realized volatility help us to answer the question: Is asymmery caused by the leverage or by the volatility feedback effect, or a mix of both? Can we link the (technical) types of asymmetry: shift and rotation, to economic theory?

Another important area of research is how to price options and other derivatives using the realized volatility. How should we deal with the sampling error in the realized volatility when pricing options? To what extent does this error affect the results? Can models and methods based on realized volatility outperform traditional option pricing methods that only use daily data? It would be interesting to conduct an extensive study of this topic, using realistic loss functions. Is it be possible to have access to intraday option prices in the future? If so, we could use these intraday option prices to better estimate the risk neutral density of the returns.

In the GARCH literature, more and more distributional assumptions are proposed. There seems to be a need for formal test procedures to test the different distributional assumptions against each other. That is, to (jointly) test which distribution, the normal, t, NIG or any other distribution for the standardized residuals, gives the best fit. We can investigate the impact of the different distributional assumptions in a practical situation, for example, by asking: What is the impact in terms of Value-at-Risk (VaR) related loss functions, such as "dollars lost", when using the different distributional assumptions?

Is it possible to have regime shifts, not just in the parameters of the GARCH equation, but also in the distribution of the standardized return? If so, can we build a framework to test this? The generalized hyperbolic distribution could be worth testing in this situation.

The Lévy processes introduced in financial econometrics by Barndorff-Nielsen and Shephard (2001a) is also likely to generate interesting research. A Lévy process is nothing more than a subordinated Brownian motion, or, if you like, Brownian motion where time is random. One special type of Lévy process uses the IG distribution as subordinator. In discrete time, this would correspond to a discrete process with normally distributed increments where the variance at a given time follows an IG distribution, i.e., the GARCH-NIG model. To estimate the parameters of a Lévy process, one could use indirect inference and use the GARCH-NIG model as an auxiliary model. One needs to find the theoretical links, if any, between the Lévy processes and discrete time GARCH processes.

Another use of the IG distribution could be in the Autoregressive Conditional Duration (ACD) models pioneered by Engle and Russel (1993). The underlying idea in these models is to model the times between trades. It would be interesting to try the IG (or GIG) distribution for the waiting times, and compare the resulting models to the ones in the literature.

The flip side of modeling times between trades is to model the number of trades in a given interval of time. This is the purpose of the BIN-model by Rydberg and Shephard (1999). They use a Possion distribution where they put a GARCH-like structure on the intensity parameter. If we assume that the intensity parameter of the Poisson distribution is drawn from an IG distribution, the number of trades in that BIN would be Sichel distributed (Sichel, 1974). So, a natural exension of the BIN model of Rydberg and Shephard (1999) is to assume that the number of trades in a BIN is Sichel distributed. The Sichel distribution exists in a multivariate setting, which would come in handy when modeling portfolios of assets.

Of course, the ideas outlined above are just ideas. Some might be bad, or even infeasible to implement, however others might be worthwhile investigating further. Surely, some results are already on the way, but results do not come by themselves. So, let's get to work!

References

ANDERSEN, T. and T. BOLLERSLEV (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885-905.

ANDERSEN, T. and T. BOLLERSLEV (1998). DM-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer-run dependencies. *Journal of Finance* 53, 219-265.

ANDERSEN, T.G. (1996). Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility. *Journal of Finance* 51, 169-204.

ANDERSEN, T.G., T. BOLLERSLEV and F.X. DIEBOLD (2002). Parametric and Nonparametric Volatility Measurement. In L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics.* Amsterdam: North-Holland, forthcoming

ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and H. EBENS (2001). The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61, 43-76.

ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and P. LABYS (2000). Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian. *Multinational Finance Journal* 4, 159-179.

ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and P. LABYS (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, 42-55.

ANDERSEN, T.G., T. BOLLERSLEV, F.X. DIEBOLD and P. LABYS (2002). Modeling and Forecasting Realized Volatility. Manuscript, Department of Economics, Duke University.

ANDERSSON, J. (2001). On the Normal Inverse Gaussian Stochastic Volatility Model. *Journal of Business and Economic Statistics* 19, 44-54.

ANDREOU, E. and E. GHYSELS (2002). Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation and Empirical Results. *Journal of Business and Economic Statistics,* forthcoming.

ANE, T. and H. GEMAN (2000). Order Flow, Transaction Clock, and Normality of Asset Returns. *Journal of Finance* 60, 2259-2284.

BARNDORFF-NIELSEN, O.E. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian journal of Statistics* 5, 151-157.

BARNDORFF-NIELSEN, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24, 1-13.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (2001a). Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics, (with discussion) *Journal of the Royal Statistical Society, Series B* 63, 167-241.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (2001b). Estimating Quadratic Variation Using Realized Volatility. Manuscript, Nuffield College, Oxford University.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (2002a). Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society, Series B* 64, forthcoming.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (2002b). How Accurate is the Asymptotic Approximation to the Distribution of Realized Volatility? In D.W.F. Andrews, J.L. Powell, P.A. Ruud, and J.H. Stock (eds.), *Identification and Inference for Econometric Models,* forthcoming.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (1998). Aggregation and model construction for volatility models. Working paper, Nuffield College, Oxford University.

BARNDORFF-NIELSEN, O.E. and N. SHEPHARD (2001c). Realised power variation and stochastic volatility models. Working paper, Nuffield College, Oxford University.

BEKAERT, G. and G. WU (2000). Asymmetric Volatility and Risk in Equity Markets. *The Review of Financial Studies* 13, c.

BERNDT, E., B. HALL, R. HALL and J. HAUSMAN (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement* 3/4, 653-665.

Black, F. (1976). Studies in stock volatility changes. *Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association*, 177-181.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307-327.

Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542-547.

Bollerslev, T. and H.O. Mikkelsen (1999). Long-term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics* 92, 75-99.

Bollerslev, T., R.Y. Chou and K.F. Kroner (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5-59.

Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115-127.

Christie, A.A. (1982). The stochastic behavior of common stock variances-value, leverage and interest rate effets. *Journal of Financial Economics* 10, 407-432.

Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135-155.

Diebold, F.X., T. 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.

Drost, F.C, and T.E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica* 61, 909-927.

Engle, R. and V.K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749-1778.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* 50, 987-1008.

Engle, R.F. and J.R. Russel (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66, 1127-1162.

Epps, T.W. and M.L. Epps (1976). The Stochastic Dependence of Security Prices Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis. *Econometrica* 44, 305-321.

Fama, E.F. (1965). The behaviour of stock market prices. *Journal of Business* 38, 34-105.

Gallant, A.R., P.E. Rossi and G.E. Tauchen (1991). Ch8 of Nonparametric and semiparameric methods in econometrics and statistics: proceedings of the fifth international syposium in economic theory and econometrics, edited by William A. Barnett, James Powell and George E. Tauchen, Cambridge University Press 1991.

Ghysels, E., A. Harvey and E. Renault (1996). Stochastic volatility. *in Handbook of Statistics Vol. 14, Statistical Methods in Finance, (Es. G.S. Maddala)*, Amsterdam: North Hollands.

Glosten, L.R., R. Jagannathan and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1802.

Glosten, L.R., R. Jagannathan and D.E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1801.

He, C. (1997). *Statistical properties of GARCH procesess.* Doctoral thesis, Stockholm School of Economics.

Hentschel, L. (1991). The Absolute Value GARCH model and the volatility of U.S. stock returns. Mimeo, Princeton University.

Hsieh, D.A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance* 46, 1839-1877.

Jarque, C.M. and A.K. Bera (1987). A test for normality of observations and regression residuals. *International Statistical Review* 55, 163-172.

Jensen, M.B. and A. Lunde (2001). The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroscedastic volatility model. *Econometrics Journal* 4, 319-342.

Kim, S., N. Shephard and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* 65, 361-393.

Maheu, J.M. and T. H. McCurdy (2002). Nonlinear Features of Realized FX Volatility. *Review of Economics and Statistics*, forthcoming.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business* 36, 394-419.

Michael, J.R., W.R. Schucany and W. Haas (1976). Generating random variables using transformations with multiple roots. *The American Statistician* 30, 88-90.

Nelson, D (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* 59, 347- 370.

Newey, W.K (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047-1070.

Rydberg, T. and N. Shephard (1999). BIN models for trade-by-trade data. Working paper, Nuffield College, Oxford University.

Schwert, G.W. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1115-1153.

Schwert, G.W. (1990a). Stock market volatility. *Financial Analysts Journal* May-June, 23-34.

Schwert, G.W. (1990b). Stock volatility and the crash of '87. *Review of Financial Studies* 3, 77-102.

Seshadri, V. (1993). *The inverse Gaussian distribution.* Clarendon Press, Oxford.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, O.E. Barndorff-Nielsen and D.V. Hinkley, eds. *Time series models in econometrics, finance and other fields* 1-67. London: Chapman and Hall.

Sichel, H.S. (1974). On a distribution representing sentence-length in written prose. *Journal of Royal Statistical Society, Series A* 137, 25-34.

Tauchen, G.E. (1985). Diagnostic testing end evaluation of maximum likelihood models. *Journal of Econometrics* 30, 415-443.

Tauchen, G.E., and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica* 51, 485-505.

Taylor, S.J. (1982). Financial Returns Modeled by the Product of two Stochastic Processes - a Study of the Daily Sugar Prices 1961-75. In O.D. Anderson (ed.), *Time Series Analysis: Theory and Practice*, 203-226. Amsterdam: North-Holland.

Taylor, S.J. (1986). *Modelling financial time series.* Chichester: John Wiley and Sons.

Watson, G.N. (1922). *A treatise on the theory of Bessel functions*, Cambridge University Press.

Zakoïan, J-M (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931-955.

Appendix A

Moments of the models

A.1 Moment structure of the GARCH(1,1)-NIG

We use the results of He (1997) to derive the moment structures of the (1,1) case of the model. First, we need some notation, let

$$\nu_k = E\left|z_t\right|^k, \; k = 1, ..., 4, \tag{A.1}$$

that is, ν_k is the expected value of the absolute value of the standardized return, note that z_t is $NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, 1\right)$. As shown in Appendix B, we find that the absolute moments of $z_t \sim NIG_{\sigma^2}\left(\overline{\alpha}, 0, 0, 1\right)$ are

$$\begin{aligned} \nu_1 &= \frac{K_0\left(\overline{\alpha}\right) 2\sqrt{a} \exp\left(\overline{\alpha}\right)}{\pi}, \\ \nu_2 &= 1, \\ \nu_3 &= \frac{4\sqrt{\overline{\alpha}} K_1\left(\overline{\alpha}\right) \exp\left(\overline{\alpha}\right)}{\pi}, \end{aligned} \tag{A.2}$$

and

$$\nu_4 = 3 + \frac{3}{\overline{\alpha}},$$

where $K_1\left(x\right)$ denotes the modified Bessel function of order three and index one, and $K_0\left(x\right)$ denotes the modified Bessel function of order three and index zero.

Further we need the following definitions

$$\gamma_{c1} = \pi_1 + \rho_1 \nu_2,$$

and

$$\gamma_{c2} = \pi_1^2 + 2\pi_1\rho_1 + \rho_1^2 \nu_4.$$

Now, we can state that the second order moment of the observed x_t exists if and only if

$$\pi_1 + \rho_1 \nu_2 < 1. \tag{A.3}$$

If $(A.3)$ holds, then the unconditional variance of x_t is given by

$$Ex_t^2 = \frac{\rho_0}{[1-\gamma_{c1}]}. \tag{A.4}$$

The fourth order unconditional moment of the observed process, exists if and only if

$$\gamma_{c2} < 1,$$

i.e.

$$\pi_1^2 + 2\pi_1\rho_1\nu_2 + \rho_1^2\nu_4 < 1. \tag{A.5}$$

If $(A.5)$ holds, then the fourth order moment is given by

$$Ex_t^4 = \mu_4 = \left[\frac{\nu_4}{[1-\gamma_{c2}]}\right]\left[\left[\frac{2\rho_0^2\gamma_{c1}}{[1-\gamma_{c1}]}\right] + \rho_0\right],$$

$$\mu_4 = \left[\frac{\nu_4}{\left[1-\left[\pi_1^2 + 2\pi_1\rho_1\nu_2 + \rho_1^2\nu_4\right]\right]}\right]\left[\frac{2\rho_0^2\left[\pi_1+\rho_1\nu_2\right]}{\left[1-\left[\pi_1+\rho_1\nu_2\right]\right]} + \rho_0\right]. \tag{A.6}$$

If $(A.5)$ holds, then the kurtosis of the observed process is given by

$$\kappa_4^0 = \frac{\kappa_4(z_t)\left[[1-\gamma_{c1}] + 2\gamma_{c1}\right][1-\gamma_{c1}]}{[1-\gamma_{c2}]},$$

where $\kappa_4(z_t) = \frac{\nu_4}{\nu_2^2}$ is the kurtosis of the standardized process. If $(A.5)$ holds, then the autocorrelation function of the squared observed process is given by

$$\rho_1^0 = \frac{\nu_2\overline{\gamma}_c\left[1-\gamma_{c1}^2\right] - \gamma_{c1}\left[1-\gamma_{c2}\right]}{\nu_4\left[1-\gamma_{c1}^2\right] - \nu_2^2\left[1-\gamma_{c2}\right]}, \tag{A.7}$$

and

$$\rho_n^0 = \rho_1^0\gamma_{c1}^{n-1}. \tag{A.8}$$

where

$$\overline{\gamma}_c = \rho_1\nu_4 + \pi_1.$$

A.2 Moment structure of the T-GARCH(1,1)-NIG

Again, we use the results from He (1997) and for the T-GARCH(1,1) model we have that

$$\gamma_{c1} = \pi_1 + \left[\rho_1 + \frac{\omega}{2}\right],$$

$$\gamma_{c2} = \left(\frac{\omega^2}{2} + \rho_1^2 + \omega\rho_1\right)\nu_4 + (\pi_1\omega + 2\pi_1\rho_1) + \pi_1^2.$$

The second order moment of the observed ε_t exists if and only if

$$\pi_1 + \left[\rho_1 + \frac{\omega}{2}\right] \nu_2 < 1. \tag{A.9}$$

If $(A.9)$ holds, then the unconditional variance of ε_t is given by

$$Ex_t^2 = \frac{\rho_0 \nu_2}{\left[1 - \left(\pi_1 + \left[\rho_1 + \frac{\omega}{2}\right] \nu_2\right)\right]}. \tag{A.10}$$

The fourth order unconditional moment of the observed process, exists if and only if

$$\left(\frac{\omega^2}{2} + \rho_1^2 + \omega\rho_1\right) \nu_4 + (\pi_1\omega + 2\pi_1\rho_1) + \pi_1^2 < 1. \tag{A.11}$$

If $(A.11)$ holds, then the fourth order moment is given by

$$\mu_4 = \left[\frac{\nu_4}{[1 - \gamma_{c2}]}\right] \left[\left[\frac{2\rho_0^2\gamma_{c1}}{[1 - \gamma_{c1}]}\right] + \rho_0\right],$$

$$\begin{aligned} \mu_4 \;\; = \;\; & \left[\frac{\nu_4}{\left[1 - \left(\frac{\omega^2}{2} + \rho_1^2 + \omega\rho_1\right) \nu_4 - (\pi_1\omega + 2\pi_1\rho_1)\, \nu_2 - \pi_1^2\right]}\right] \\ & * \left[\frac{2\rho_0^2 \left[\pi_1 + \left[\rho_1 + \frac{\omega}{2}\right] \nu_2\right]}{\left[1 - \left[\pi_1 + \left[\rho_1 + \frac{\omega}{2}\right] \nu_2\right]\right]} + \rho_0\right]. \end{aligned} \tag{A.12}$$

If $(A.11)$ holds, then the kurtosis of the observed process is given by

$$\kappa_4^0 = \frac{\kappa_4(z_t) \left[\rho_0^2 \left[1 - \gamma_{c1}\right] + 2\rho_0^2\gamma_{c1}\right] \left[1 - \gamma_{c1}\right]}{\rho_0^2 \left[1 - \gamma_{c2}\right]}.$$

If $(A.11)$ holds, then the autocorrelation function of the squared observed process is given by

$$\rho_1^0 = \frac{\overline{\gamma}_c \left[1 - \gamma_c^2\right] - \gamma_c \left[1 - \gamma_{c2}\right]}{\nu_4 \left[1 - \gamma_{c1}^2\right] - \left[1 - \gamma_{c2}\right]},$$

where

$$\overline{\gamma}_{c1} = \rho_1\nu_4 + \frac{\omega\nu_4}{2} + \pi_1,$$

and for $n > 1$

$$\begin{aligned} \rho_n^0 \;\; &= \;\; \rho_1^0 \gamma_{c1}^{n-1}, \\ \rho_n^0 \;\; &= \;\; \rho_1^0 \left[\pi_1 + \left[\rho_1 + \frac{\omega}{2}\right]\right]^{n-1}. \end{aligned}$$

A.3 Moment structure of the AV-GARCH(1,1)-NIG

We use the results in He (1997) to derive the moment structure of the (1,1) case. The moment structure for the general (p,q) case can be derived by using a system of equation methods as outlined in Andersson (2001).

For the AV-GARCH(1.1)-NIG the second order moment of the observed process x_t exists if and only if

$$\pi_1^2 + \rho_1^2\nu_2 + 2\pi_1\rho_1\nu_1 < 1. \tag{A.13}$$

If $(A.13)$ holds, then the unconditional variance of x_t is given by

$$Ex_t^2 = \frac{\nu_2}{\left[1 - \left[\pi_1^2 + \rho_1^2\nu_2 + 2\pi_1\rho_1\nu_1\right]\right]}\left[\frac{2\rho_0^2\left[\pi_1 + \rho_1\nu_1\right]}{\left[1 - \left[\pi_1 + \rho_1\nu_1\right]\right]} + \rho_0^2\right]. \tag{A.14}$$

The fourth order unconditional moment of the observed process, exists if and only if

$$\pi_1^4 + 6\rho_1^2\nu_2\pi_1^2 + 4\pi_1^3\rho_1\nu_1 + \rho_1^4\nu_4 + 4\rho_1^3\nu_3\pi_1 < 1 \tag{A.15}$$

If $(A.15)$ holds, then the fourth order moment is given by

$$\mu_4 = \nu_4\rho_0^4\Delta_4^0\Pi_{i=1}^4\left[1 - \gamma_{ci}\right]^{-1}. \tag{A.16}$$

where

$$\begin{aligned}
\gamma_{c1} &= \pi_1 + \rho_1\nu_1, \\
\gamma_{c2} &= \pi_1^2 + \rho_1^2\nu_2 + 2\pi_1\rho_1\nu_1, \\
\gamma_{c3} &= \pi_1^3 + 3\rho_1^2\nu_2\pi_1 + 3\rho_1\nu_1\pi_1^2 + \rho_1^3\nu_3,
\end{aligned}$$

and

$$\gamma_{c4} = \pi_1^4 + 6\rho_1^2\nu_2\pi_1^2 + 4\pi_1^3\rho_1\nu_1 + \rho_1^4\nu_4 + 4\rho_1^3\nu_3\pi_1.$$

If $(A.15)$ holds, then the kurtosis of the observed process is given by

$$\kappa_4^+ = \frac{\kappa_4(z)\Delta_4^0\left[1 - \gamma_{c1}\right]\left[1 - \gamma_{c2}\right]}{\left[1 + \gamma_{c1}\right]^2\left[1 - \gamma_{c3}\right]\left[1 - \gamma_{c4}\right]}. \tag{A.17}$$

The ACF of the squared observations of the AV-GARCH(1,1)-NIG model is given by

$$\begin{aligned}
\rho_1^0 \quad = \quad & \frac{\nu_2\left(1 - \gamma_{c1}\right)\left(1 - \gamma_{c2}\right)\left[2\overline{\gamma}_c\left(1 - \gamma_{c4}\right)\Delta_3^0 + \overline{\gamma}_{c2}\Delta_4^0\right]}{\Delta^0} \qquad \text{(A.18)} \\
& - \frac{\nu_2^2\left(1 + \gamma_{c1}\right)\left(1 - \gamma_{c3}\right)\left(1 - \gamma_{c4}\right)\left[2\gamma_c + \gamma_{c2}\left(1 - \gamma_{c1}\right)\right]}{\Delta^0},
\end{aligned}$$

and

$$\rho_n^+ = \gamma_{c2}\rho_{n-1}^+ + \theta^0\gamma_{c1}^{n-1}, \tag{A.19}$$

where

$$\Delta^0 = \nu_4 \Delta_4^0 (1-\gamma_{c1})(1-\gamma_{c2}) - \nu_2^2 (1+\gamma_{c1})^2 (1-\gamma_{c3})(1-\gamma_{c4}),$$

$$\theta^0 = \frac{1}{\Delta^0}\left[2\nu_2 (1-\gamma_{c2})(1-\gamma_{c4})\left[\Delta_3^0 \overline{\gamma}_c (1-\gamma_{c1})\right] - \nu_2 \gamma_{c1} (1+\gamma_{c1})(1-\gamma_{c3})\right],$$

$$\Delta_3^0 = 1 + 2\gamma_{c1} + 2\gamma_{c2} + \gamma_c \gamma_{c2},$$

and

$$\Delta_4^0 = 1 + 3\gamma_c + 5\gamma_{c2} + 3\gamma_{c3} + 3\gamma_c\gamma_{c2} + 5\gamma_c\gamma_{c3} + 3\gamma_{c2}\gamma_{c3} + \gamma_c\gamma_{c2}\gamma_{c3}.$$

A.4 Moment structure of the TAV-GARCH(1,1)-NIG

The second order moment of the observed x_t exists if and only if

$$\nu_2\rho_1^2 + \frac{2\nu_2\rho_1\omega}{2} + 2\nu_1\rho_1\pi_1 + \frac{\nu_2\omega^2}{2} + \frac{2\nu_1\omega\pi_1}{2} + \pi_1^2 < 1. \tag{A.20}$$

If $(A.20)$ holds then the second order unconditional moment of x_t can be expressed as

$$\begin{aligned} Ex_t^2 \;\; = \;\; & 2\nu_2\rho_0^2 * \\ & \frac{4\rho_0\rho_1\nu_1 + 2\rho_0\omega\nu_1 + 4\rho_0\pi_1 + 2 - 2\rho_1\nu_1 - \omega\nu_1 - 2\pi_1}{\left(-2 + 2\nu_2\rho_1^2 + 2\nu_2\rho_1\omega + 4\nu_1\rho_1\pi_1 + \nu_2\omega^2 + 2\nu_1\omega\pi_1 + 2\pi_1^2\right)} \\ & * \frac{1}{(-2 + 2\rho_1\nu_1 + \omega\nu_1 + 2\pi_1)}. \end{aligned}$$

Fourth moment of the observed process $E\left(x_t^4\right)$

$$\gamma_{c4} < 1 \tag{A.21}$$

where

$$\begin{aligned} \gamma_{c4} \;\; = \;\; & 4\rho_1^3\nu_4\omega\frac{1}{2} + 4\rho_1^3\nu_3\pi_1 + \pi_1^4 + 4\omega\frac{1}{2}\nu_1\pi_1^3 + 4\omega^3\frac{1}{2}\nu_3\pi_1 + \\ & 6\omega^2\frac{1}{2}\nu_2\pi_1^2 + 12\rho_1\nu_2\omega\frac{1}{2}\pi_1^2 + 12\rho_1\nu_3\omega^2\frac{1}{2}\pi_1 + 4\rho_1\nu_4\omega^3\frac{1}{2} \\ & + \rho_1^4\nu_4 + 6\rho_1^2\nu_4\omega^2\frac{1}{2} + 12\rho_1^2\nu_3\omega\frac{1}{2}\pi_1 + \omega^4\frac{1}{2}\nu_4 + 4\rho_1\nu_1\pi_1^3 + 6\rho_1^2\nu_2\pi_1^2 \end{aligned}$$

If $(A.21)$ holds, then the fourth order moment of x_t is given by

$$\mu_4^* = \nu_4\rho_0^4\Delta_4^0\Pi_{i=1}^4\left[1-\gamma_{ci}\right]^{-1},$$

and the kurtosis by

$$\kappa_4^+ = \frac{\kappa_4(z)\Delta_4^0\,[1-\gamma_c]\,[1-\gamma_{c2}]}{[1+\gamma_c]^2\,[1-\gamma_{c3}]\,[1-\gamma_{c4}]},$$

where $\kappa_4(z)$ is the kurtosis of the standardized process and

$$\begin{aligned}
\Delta_4^0 &= 1+3\gamma_c+5\gamma_{c2}+3\gamma_{c3}+3\gamma_c\gamma_{c2}+5\gamma_c\gamma_{c3}+3\gamma_{c2}\gamma_{c3}+\gamma_c\gamma_{c2}\gamma_{c3},\\
\gamma_{c1} &= \rho_1\nu_1+\frac{\omega}{2}\nu_1+\pi_1,\\
\gamma_{c2} &= \nu_2\rho_1^2+\frac{2\nu_2\rho_1\omega}{2}+2\nu_1\rho_1\pi_1+\frac{\nu_2\omega^2}{2}+\frac{2\nu_1\omega\pi_1}{2}+\pi_1^2,
\end{aligned}$$

and

$$\begin{aligned}
\gamma_{c3} &= \rho_1^3\nu_3+3\rho_1^2\nu_3\omega\frac{1}{2}+3\rho_1^2\nu_2\pi_1+3\rho_1\nu_3\omega^2\frac{1}{2}+6\rho_1\nu_2\omega\frac{1}{2}\pi_1+3\rho_1\nu_1\pi_1^2\\
&\quad+\omega^3\frac{1}{2}\nu_3+3\omega^2\frac{1}{2}\nu_2\pi_1+3\omega\frac{1}{2}\nu_1\pi_1^2+\pi_1^3.
\end{aligned}$$

The ACF of the squared observations from the TAV-GARCH(1,1)-NIG model is given by

$$\begin{aligned}
\rho_1^0 &= \frac{\nu_2\,(1-\gamma_c)\,(1-\gamma_{c2})\,\left[2\overline{\gamma}_c\,(1-\gamma_{c4})\,\Delta_3^0+\overline{\gamma}_{c2}\Delta_4^0\right]}{\Delta^0}\\
&\quad-\frac{\nu_2^2\,(1+\gamma_c)\,(1-\gamma_{c3})\,(1-\gamma_{c4})\,[2\gamma_c+\gamma_{c2}\,(1-\gamma_c)]}{\Delta^0},
\end{aligned}$$

and

$$\rho_n^+ = \gamma_{c2}\rho_{n-1}^+ + \theta^0\gamma_c^{n-1},$$

where

$$\Delta^0 = \nu_4\Delta_4^0\,(1-\gamma_c)\,(1-\gamma_{c2}) - \nu_2^2\,(1+\gamma_c)^2\,(1-\gamma_{c3})\,(1-\gamma_{c4})\,,$$

$$\theta^0 = \frac{1}{\Delta^0}\left[2\nu_2\,(1-\gamma_{c2})\,(1-\gamma_{c4})\,\left[\Delta_3^0\overline{\gamma}_c\,(1-\gamma_c)\right]-\nu_2\gamma_c\,(1+\gamma_c)\,(1-\gamma_{c3})\right],$$

$$\Delta_3^0 = 1+2\gamma_c+2\gamma_{c2}+\gamma_c\gamma_{c2},$$

$$\Delta_4^0 = 1+3\gamma_{c1}+5\gamma_{c2}+3\gamma_{c3}+3\gamma_{c1}\gamma_{c2}+5\gamma_{c1}\gamma_{c3}+3\gamma_{c2}\gamma_{c3}+\gamma_{c1}\gamma_{c2}\gamma_{c3}.$$

Appendix B

Absolute moments of NIG

In this appendix we derive the expectation of the absolute value of a normal inverse Gaussian distributed variable.

We derive the quantity

$$\nu_k = E\left|z_t\right|^k,$$

where $z \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1)$, for $k = 1, ..., 4$.

From Barndorff-Nielsen and Shephard (1998) we know that if $z \sim NIG_{\sigma^2}(\overline{\alpha}, 0, 0, 1)$ we can write

$$z_t = \sqrt{\eta_t}\varepsilon_t,$$

where $\varepsilon_t \sim N(0,1)$ and $\eta_t \sim IG_{\sigma^2}(1, \overline{\alpha})$. Now, we can write

$$E(\left|z_t\right|^i) = E(\left|\sqrt{\eta_t}\varepsilon_t\right|^i),$$

The variable η_t is always positive so

$$E\left(\left|\sqrt{\eta_t}\varepsilon_t\right|^i\right) = E\left(\eta_t^{\frac{i}{2}}\left|\varepsilon_t\right|^i\right),$$

η_t and ε_t are independent

$$E\left(\eta_t^{\frac{i}{2}}\left|\varepsilon_t\right|^i\right) = E\left(\eta_t^{\frac{i}{2}}\right)E\left(\left|\varepsilon_t\right|^i\right). \tag{B.1}$$

The problem reduces to calculate $E\left(\eta_t^{\frac{i}{2}}\right)$ and $E\left(\left|\varepsilon_t\right|^i\right)$ for $i = 1, 2, 3, 4$. Following Seshadri (1993), we have

$$\begin{aligned} E\left(\eta_t^{1/2}\right) &= \frac{\sqrt{2\overline{\alpha}}K_0(\overline{\alpha})\exp(\overline{\alpha})}{\pi^{1/2}}, \\ E\left(\eta_t^{2/2}\right) &= 1, \\ E\left(\eta_t^{3/2}\right) &= \frac{\sqrt{2\overline{\alpha}}K_1(\overline{\alpha})\exp(\overline{\alpha})}{\pi^{1/2}}, \end{aligned}$$

and

$$E\left(\eta_t^{4/2}\right) = 1 + \frac{1}{\alpha}.$$

If $\varepsilon_t \sim N(0,1)$, then it is straightforward to show that

$$\begin{aligned} E\left(|\varepsilon_t|^1\right) &= \sqrt{\frac{2}{\pi}}, \\ E\left(|\varepsilon_t|^2\right) &= 1, \\ E\left(|\varepsilon_t|^3\right) &= \sqrt{\frac{8}{\pi}}, \end{aligned}$$

and

$$E\left(|\varepsilon_t|^4\right) = 3.$$

The moments in $(A.2)$ are given by multiplication as in $(B.1)$.

Appendix C

Gradients and Hessians of the models

In this appendix, we derive the gradient and Hessian for the models GARCH(p,q)-NIG, T-GARCH(p,q)-NIG, AV-GARCH(p,q)-NIG and TAV-GARCH(p,q)-NIG. These models all rely on the NIG distribution, so the gradients and Hessians differ only in the derivative with respect to the parameters in the conditional variance (standard deviation) equation, i.e., the term $\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}$. Therefore, we start by deriving an expression for a General GARCH-NIG gradient and Hessian, leaving the derivative with respect to the parameters in the variance equation unspecified.

For all the GARCH-NIG models, the log likelihood for one observation is given by

$$\ln l_t = \frac{1}{2}\ln \overline{\alpha} - \ln(\pi) - \frac{1}{2}\ln \sigma_t^2 + \overline{\alpha} - \frac{1}{2}\ln\left(1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}}\right) + \ln K_1\left(\overline{\alpha}\left(1 + \frac{x_t^2}{\sigma_t^2\overline{\alpha}}\right)^{1/2}\right). \tag{C.1}$$

For ease of derivation, we split up the gradient in two parts. Let $\boldsymbol{\psi}$ be the vector of all parameters of the model, which we can split up into $\boldsymbol{\psi} = (\overline{\alpha}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ contains the parameters for the variance equation. We write

$$\frac{\partial \ln l_t}{\partial \boldsymbol{\psi}} = \begin{pmatrix} \frac{\partial \ln l_t}{\partial \overline{\alpha}} \\ \\ \frac{\partial \ln l_t}{\partial \boldsymbol{\theta}} \end{pmatrix}.$$

The derivative of the log likelihood with respect to (wrt) to the parameter $\overline{\alpha}$, i.e., $\frac{\partial \ln l_t}{\partial \overline{\alpha}}$ is common to all the models[1]. The gradient of the log likelihood

[1] In the derivation of the gradients and Hesssians, we will make use of the derivative of a

function for the General GARCH-NIG is given by

$$\frac{\partial \ln l_t}{\partial \boldsymbol{\psi}} = \begin{pmatrix} \frac{1}{2\overline{\alpha}} + 1 + \frac{1}{2}\frac{x_t^2}{\overline{\alpha}\left(\sigma_t^2\overline{\alpha}+x_t^2\right)} - \frac{1}{2}\left[\left[\frac{K_0(\tau_t)\tau_t}{K_1(\tau_t)} + 1\right]\frac{\left(2\sigma_t^2\overline{\alpha}+x_t^2\right)}{\overline{\alpha}\left(\sigma_t^2\overline{\alpha}+x_t^2\right)}\right] \\ \frac{1}{2}\left(\frac{x_t^2}{\sigma_t^2\left(\sigma_t^2\overline{\alpha}+x_t^2\right)} - \frac{1}{\sigma_t^2} + \frac{x_t^2}{\sigma_t^4}\left(\frac{K_0(\tau_t)}{K_1(\tau_t)} + \frac{1}{\tau_t}\right)\left(1+\frac{x_t^2}{\sigma_t^2\overline{\alpha}}\right)^{-1/2}\right)\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}} \end{pmatrix}, \tag{C.2}$$

where

$$\tau_t = \overline{\alpha}\left(1+\frac{x_t^2}{\sigma_t^2\overline{\alpha}}\right)^{1/2}.$$

We have left the terms σ_t^2 and $\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}$ unspecified since these depend on the different models, see below.

To derive the Hessian, we take the second derivative of $(C.1)$,

$$\frac{\partial^2 \ln l_t}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'} = \begin{pmatrix} \frac{\partial^2 \ln l_t}{\partial \overline{\alpha}\partial \overline{\alpha}} \\ \frac{\partial^2 \ln l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \end{pmatrix}.$$

We start with the second derivative with respect to $\overline{\alpha}$

$$\begin{aligned} \frac{\partial^2 \ln l_t}{\partial \overline{\alpha}\partial \overline{\alpha}} &= \left(\frac{1}{2}\frac{\left(-2a\sigma_t^2 - x_t^2\right)}{\sigma_t^2} + \frac{K_0(\tau_t)\left(4\sigma_t^2 a + 2x_t^2\right)}{\tau_t K_1(\tau_t)\sigma_t^2} + \frac{K_0(\tau_t)^2\left(2a\sigma_t^2 + x_t^2\right)}{K_1(\tau_t)^2\sigma_t^2}\right) \\ &\quad + \left(-\frac{2\sigma_t^4 a^2 + 2\sigma_t^2 a x_t^2 + x_t^4}{\left(\sigma_t^2 a + x_t^2\right)^2 a^2}\right) \\ &\quad - \frac{1}{2}\left(\frac{K_0(\tau_t)\tau_t}{K_1(\tau_t)} + 1\right)\left(-\frac{2\sigma_t^4 a^2 + 2\sigma_t^2 a x_t^2 + x_t^4}{\left(\sigma_t^2 a + x_t^2\right)^2 a^2}\right) \\ &\quad - \frac{1}{2}\left[\frac{\left(2\sigma_t^2 a + x_t^2\right)}{a\left(\sigma_t^2 a + x_t^2\right)}\left(\frac{1}{2}\frac{\left(-2a\sigma_t^2 - x_t^2\right)}{\sigma_t^2} + \frac{K_0(\tau_t)\left(4\sigma_t^2 a + 2x_t^2\right)}{\tau_t K_1(\tau_t)\sigma_t^2}\right)\right] \\ &\quad - \frac{1}{2}\left[\frac{\left(2\sigma_t^2 a + x_t^2\right)}{a\left(\sigma_t^2 a + x_t^2\right)}\left(\frac{K_0(\tau_t)^2\left(2a\sigma_t^2 + x_t^2\right)}{K_1(\tau_t)^2\sigma_t^2}\right)\right]. \end{aligned}$$

To find the second order derivative wrt the parameter of the conditional variance equation we rewrite the second order derivative wrt the parameters of conditional variance equation as

$$\frac{\partial^2 \ln l_t}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \frac{\partial\left(\frac{1}{2}(AB)C\right)}{\partial \boldsymbol{\theta}'},$$

Bessel function. The derivative of a Bessel function with index ν s given by

$$\frac{d\left(K_\nu(x)\right)}{dx} = -\frac{K_{\nu+1}(x)x - \nu K_\nu(x)}{x}.$$

where

$$A = \left(\frac{x_t^2}{\sigma_t^2 \left(\sigma_t^2 \overline{\alpha} + x_t^2 \right)} - \frac{1}{\sigma_t^2} + \frac{x_t^2}{\sigma_t^4} \left(\frac{K_0(\tau_t)}{K_1(\tau_t)} + \frac{1}{\tau_t} \right) \right),$$

$$B = \left(1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}} \right)^{-1/2},$$

and

$$C = \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}}.$$

The second order derivative is given by

$$\frac{\partial^2 \ln l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = BC \frac{\partial A}{\partial \boldsymbol{\theta}'} + AC \frac{\partial B}{\partial \boldsymbol{\theta}'} + AB \frac{\partial C}{\partial \boldsymbol{\theta}'}, \tag{C.3}$$

where

$$\begin{aligned} \frac{\partial A}{\partial \boldsymbol{\theta}'} &= \left(-x_t^2 \left(\frac{\overline{\alpha}}{\phi^2} + \frac{1}{\phi} \frac{1}{\sigma_t^4} \right) + \frac{1}{\sigma_t^4} - \frac{2K_0(\tau_t)}{K_1(\tau_t)} \frac{x_t^4}{\sigma_t^6} \right) \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} \\ &\quad + \frac{1}{2} \frac{x_t^4}{\sigma_t^8 \sqrt{1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}}}} \left(1 - \left(\frac{K_0(\tau_t)^2}{K_1(\tau_t)^2} + \frac{K_0(\tau_t)}{K_1(\tau_t)\tau_t} \right) \right) \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} \\ &\quad + \left(\frac{x_t^4}{\phi} \frac{1}{2\tau_t} \frac{1}{\sigma_t^6} - \frac{1}{\tau_t} \frac{2x_t^2}{\sigma_t^6} \right) \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'} + \frac{x_t^2}{\sigma_t^2}, \end{aligned}$$

$$\frac{\partial B}{\partial \boldsymbol{\theta}'} = \frac{1}{2} \left(1 + \frac{x_t^2}{\sigma_t^2 \overline{\alpha}} \right)^{-3/2} \frac{x_t^2}{\sigma_t^4 \overline{\alpha}} \frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'},$$

and

$$\frac{\partial C}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

To obtain the gradient and Hessian of the different models, we specify the derivative wrt the parameter of the variance equation $\frac{\partial \sigma_t^2}{\partial \boldsymbol{\theta}'}$, this is done below for the four different models.

C.1 GARCH(p,q)-NIG

To find the gradient of the GARCH(p,q)-NIG model, we need to find the derivative of the variance equation wrt the parameters. Let us start with some convenient notation. Let $y_{t,GARCH} = \left(1, x_{t-1}^2, ..., x_{t-q}^2, \sigma_{t-1}^2, ..., \sigma_{t-p}^2\right)'$ be a data vector and $\boldsymbol{\psi}_{GARCH} = (\overline{\alpha}, \rho_0, \rho_1, ..., \rho_q, \pi_1, ..., \pi_p)$. In $\boldsymbol{\psi}_{GARCH}$ we have the parameters for the GARCH(p,q)-NIG model. Now, partition

$\boldsymbol{\psi}_{GARCH} = \left(\overline{\alpha}, \boldsymbol{\theta}'_{GARCH}\right)'$ where $\boldsymbol{\theta}_{GARCH} = \left(\rho_0, \rho_1, ..., \rho_q, \pi_1, ..., \pi_p\right)'$ contains the parameters in the GARCH equation. The variance equation now becomes

$$\sigma^2_{t,GARCH} = \rho_0 + \rho_1 x^2_{t-1} + ... + \rho_q x^2_{t-q} + \pi_1 \sigma^2_{t-1} + ... + \pi_p \sigma^2_{t-p}, \qquad \text{(C.4)}$$

or

$$\sigma^2_{t,GARCH} = \mathbf{y}'_{t,GARCH} \boldsymbol{\theta}_{GARCH}.$$

For the GARCH(p,q)-NIG, the derivative of the parameters wrt the variance equation is given by the recursion

$$\frac{\partial \sigma^2_{t,GARCH}}{\partial \boldsymbol{\theta}_{GARCH}} = \mathbf{y}_{t-1,GARCH} + \Sigma^p_{i=1} \pi_i \frac{\partial \sigma^2_{t-i,GARCH}}{\partial \boldsymbol{\theta}_{GARCH}}. \qquad \text{(C.5)}$$

which we plug into $(C.2)$. And the Hessian of the GARCH(p,q)-NIG is given by $(C.3)$ where we plug in $\sigma^2_{t,GARCH}$ in the place for σ^2_t and

$$\frac{\partial^2 \sigma^2_{t,GARCH}}{\partial \boldsymbol{\theta}_{GARCH} \partial \boldsymbol{\theta}'_{GARCH}} = \frac{\partial \mathbf{y}_{t-1,GARCH}}{\partial \boldsymbol{\theta}'_{GARCH}} + \frac{\partial \left(\Sigma^p_{i=1} \pi_i \partial \sigma^2_{t-i,GARCH}\right)}{\partial \boldsymbol{\theta}_{GARCH} \partial \boldsymbol{\theta}'_{GARCH}}.$$

in the place of $\frac{\partial^2 \sigma^2_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$.

C.2 T-GARCH(p,q)-NIG

For the T-GARCH(p,q)-NIG model choose the data vector as $\mathbf{y}_{t,T-GARCH} = \left(1, x^2_{t-1}, ..., x^2_{t-q}, I\left(x_{t-1} < 0\right) x^2_{t-1}, ..., I\left(x_{t-q} < 0\right) x^2_{t-q}, \sigma^2_{t-1}, ..., \sigma^2_{t-p}\right)'$ and the parameter vector $\boldsymbol{\psi}_{T-GARCH} = \left(\overline{\alpha}, \rho_0, \rho_1, ..., \rho_q, \omega_1, ..., \omega_q, \pi_1, ..., \pi_p\right)'$. Partition $\boldsymbol{\psi}_{T-GARCH} = \left(\overline{\alpha}, \boldsymbol{\theta}'_{T-GARCH}\right)'$ where $\boldsymbol{\theta}_{T-GARCH} = \left(\rho_0, \rho_1, ..., \rho_q, \omega_1, ..., \omega_q, \pi_1, ..., \pi_p\right)'$. The conditional variance equation for the T-GARCH(p,q)-NIG becomes

$$\begin{aligned}
\sigma^2_{t,T-GARCH} &= \mathbf{y}'_{t,T-GARCH} \boldsymbol{\theta}_{T-GARCH} \\
\sigma^2_{t,T-GARCH} &= \rho_0 + \rho_1 x^2_{t-1} + ... + \rho_q x^2_{t-q} + I\left(x_{t-1} < 0\right) \omega x^2_{t-1} + ... \\
&\quad + I\left(x_{t-q} < 0\right) \omega x^2_{t-q} + \pi_1 \sigma^2_{t-1} + ... + \pi_p \sigma^2_{t-p}.
\end{aligned}$$

For the T-GARCH(p,q)-NIG, the derivative of the parameters wrt the variance equation is given by the recursion

$$\frac{\partial \sigma^2_{t,T-GARCH}}{\partial \boldsymbol{\theta}_{T-GARCH}} = \mathbf{y}_{t-1,T-GARCH} + \Sigma^p_{i=1} \pi_i \frac{\partial \sigma^2_{t-i,T-GARCH}}{\partial \boldsymbol{\theta}_{T-GARCH}}. \qquad \text{(C.6)}$$

which we plug into $(C.2)$.

The Hessian of the GARCH(p,q)-NIG is given by $(C.3)$ where we plug in $\sigma^2_{t,T-GARCH}$ in the place of σ^2_t and

$$\frac{\partial^2 \sigma^2_{t,T-GARCH}}{\partial \boldsymbol{\theta}_{T-GARCH} \partial \boldsymbol{\theta}'_{T-GARCH}} = \frac{\partial \mathbf{y}_{t-1,T-GARCH}}{\partial \boldsymbol{\theta}'_{T-GARCH}} + \frac{\partial \left(\Sigma_{i=1}^{p} \pi_i \partial \sigma^2_{t-i,T-GARCH} \right)}{\partial \boldsymbol{\theta}_{T-GARCH} \partial \boldsymbol{\theta}'_{T-GARCH}},$$

in the place of $\frac{\partial^2 \sigma^2_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$.

C.3 AV-GARCH(p,q)-NIG

For the Absolute-Value GARCH(p,q)-NIG we model the conditional standard deviation that is σ_t, instead of the conditional variance $\left(\sigma^2_t\right)$as in the GARCH(p,q)-NIG and the T-GARCH(p,q)-NIG.

For the AV-GARCH(p,q)-NIG let the data vector be $\mathbf{y}_{t,AV-GARCH} = (1, |x_{t-1}|, ..., |x_{t-q}|, \sigma_{t-1}, ..., \sigma'_{t-p})$ and $\boldsymbol{\psi}_{AV-GARCH} = \left(\overline{\alpha}, \rho_0, \rho_1, ..., \rho_q, \pi_1, ..., \pi_p\right)'$ be the vector of parameters. Now, partition $\boldsymbol{\psi}_{AV-GARCH} = \left(\overline{\alpha}, \boldsymbol{\theta}'_{AV-GARCH}\right)'$ where $\boldsymbol{\theta}_{AV-GARCH} = \left(\rho_0, \rho_1, ..., \rho_q, \pi_1, ..., \pi_p\right)'$ is the parameters specific to the equation for the conditional standard deviation of the AV-GARCH(p,q)-NIG model.

The equation for the conditional standard deviation for the AV-GARCH(p,q)-NIG becomes

$$\sigma_{t,AV-GARCH} = \rho_0 + \rho_1 |x_{t-1}| + ... + \rho_1 |x_{t-q}| + ... + \pi_1 \sigma_{t-1} + ... + \pi_p \sigma_{t-p},$$

or

$$\sigma_{t,AV-GARCH} = \mathbf{y}'_{t,AV-GARCH} \boldsymbol{\theta}_{AV-GARCH}.$$

Therefore we need to adjust the term $\frac{\partial \sigma^2_t}{\partial \boldsymbol{\theta}}$ in $(C.2)$. We need the derivative wrt the parameter in σ_t, we have

$$\begin{aligned} \frac{\partial \sigma^2_t}{\partial \boldsymbol{\theta}} &= \frac{\partial \sigma^2_t}{\partial \sigma_t} \frac{\partial \sigma_t}{\partial \boldsymbol{\theta}} \\ &= 2\sigma_t \frac{\partial \sigma_t}{\partial \boldsymbol{\theta}}. \end{aligned}$$

To get the gradient of the AV-GARCH(p,q)-NIG we replace $\frac{\partial \sigma^2_t}{\partial \boldsymbol{\theta}}$ in $(C.2)$ by $2\sigma_{t,AV-GARCH} \frac{\partial \sigma_{t,AV-GARCH}}{\partial \boldsymbol{\theta}_{AV-GARCH}}$, and we get the gradient for AV-GARCH(p,q)-

NIG

$$\frac{\partial \ln l_t}{\partial \boldsymbol{\psi}_{AV-GARCH}} = \tag{C.7}$$
$$\left(\begin{array}{c} \frac{1}{2\overline{\alpha}} + 1 + \frac{1}{2}\frac{x_t^2}{\overline{\alpha}\left(\sigma_t^2\overline{\alpha}+x_t^2\right)} - \frac{1}{2}\left[\left[\frac{K_0(\tau_t)\tau_t}{K_1(\tau_t)} + 1\right]\frac{\left(2\sigma_t^2\overline{\alpha}+x_t^2\right)}{\overline{\alpha}\left(\sigma_t^2\overline{\alpha}+x_t^2\right)}\right] \\ \left(\frac{x_t^2}{\sigma_t\left(\sigma_t^2\overline{\alpha}+x_t^2\right)} - \frac{1}{\sigma_t} + \frac{x_t^2}{\sigma_t^3}\left(\frac{K_0(\tau_t)}{K_1(\tau_t)} + \frac{1}{\tau_t}\right)\left(1 + \frac{x_t^2}{\sigma_t^2\overline{\alpha}}\right)^{-1/2}\right)\frac{\partial \sigma_{t,AV-GARCH}}{\partial \boldsymbol{\theta}_{AV-GARCH}} \end{array} \right).$$

and the recursion is given by

$$\frac{\partial \sigma_{t,v}}{\partial \boldsymbol{\theta}_{AV-GARCH}} = \mathbf{y}_t + \Sigma_{i=1}^{p}\pi_i \frac{\partial \sigma_{t-i,AV-GARCH}}{\partial \boldsymbol{\theta}_{AV-GARCH}}, \tag{C.8}$$

which we plug into $(C.2)$.

The Hessian of the GARCH(p,q)-NIG is given by $(C.3)$ where we plug in

$$\frac{\partial^2 \sigma_t}{\partial \boldsymbol{\theta}_{AV-GARCH}\partial \boldsymbol{\theta}'_{AV-GARCH}} = \frac{\partial \mathbf{y}_{t-1}}{\partial \boldsymbol{\theta}'_{AV-GARCH}} + \frac{\partial\left(\Sigma_{i=1}^{p}\pi_i \partial \sigma_{t-i}\right)}{\partial \boldsymbol{\theta}_{AV-GARCH}\partial \boldsymbol{\theta}'_{AV-GARCH}},$$

in the place of $\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$.

C.4 TAV-GARCH(p,q)-NIG

To derive the gradient of TAV-GARCH(p,q)-NIG define the following vectors $\mathbf{y}_t = \left(1, |x_{t-1}|, ..., |x_{t-q}|, I(x_{t-1} < 0)|x_{t-1}|, ..., I(x_{t-q} < 0)|x_{t-q}|, \sigma_{t-1}, ..., \sigma_{t-p}\right)'$ and $\boldsymbol{\psi} = \left(\overline{\alpha}, \rho_0, \rho_1, ..., \rho_q, \omega_1, ..., \omega_q, \pi_1, ..., \pi_p\right)'$. As before partition $\boldsymbol{\psi} = \left(\overline{\alpha}, \boldsymbol{\theta}'\right)'$ where $\boldsymbol{\theta} = \left(\rho_0, \rho_1, ..., \rho_q, \omega_1, ..., \omega_q, \pi_1, ..., \pi_p\right)'$.

For the TAV-GARCH(p,q)-NIG model, the equation for the conditional standard deviation becomes

$$\begin{aligned} \sigma_{t,TAV-GARCH} &= \mathbf{y}'_{t,TAV-GARCH}\boldsymbol{\theta}_{TAV-GARCH} \\ \sigma_{t,TAV-GARCH} &= \rho_0 + \rho_1|x_{t-1}| + ... + \rho_1|x_{t-q}| + I(x_{t-1} < 0)\,\omega\,|x_{t-1}| + ... \\ &\quad + I(x_{t-q} < 0)\,\omega\,|x_{t-q}| + \pi_1\sigma_{t-1} + ... + \pi_p\sigma_{t-p} \end{aligned}$$

For the TAV-GARCH(p,q)-NIG, the derivative of the parameters wrt the conditional standard deviation equation is given by the recursion

$$\frac{\partial \sigma_{t,TAV-GARCH}}{\partial \boldsymbol{\theta}_{TAV-GARCH}} = \mathbf{y}_{t-1,TAV-GARCH} + \Sigma_{i=1}^{p}\pi_i \frac{\partial \sigma_{t-i,TAV-GARCH}}{\partial \boldsymbol{\theta}_{TAV-GARCH}}. \tag{C.9}$$

which we plug into $(C.2)$. The Hessian of the TAV-GARCH(p,q)-NIG is given by $(C.3)$ where we plug in

$$\begin{aligned}\frac{\partial^2 \sigma_t}{\partial \boldsymbol{\theta}_{TAV-GARCH} \partial \boldsymbol{\theta}'_{TAV-GARCH}} &= \frac{\partial \mathbf{y}_{t-1}}{\partial \boldsymbol{\theta}'_{TAV-GARCH}} + \\ &\quad \frac{\partial \left(\Sigma_{i=1}^{p} \pi_i \partial \sigma_{t-i} \right)}{\partial \boldsymbol{\theta}_{TAV-GARCH} \partial \boldsymbol{\theta}'_{TAV-GARCH}},\end{aligned}$$

in the place of $\frac{\partial^2 \sigma_t^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$.

There are no ends, only beginnings.