UPPSALA
UNIVERSITET

Autoregressive Conditional Density

By Jacob Lindberg

Department of Statistics
Uppsala University

Supervisor: Per Johansson

2016

Abstract

We compare two time series models: an ARMA(1,1)-ACD(1,1)-NIG model against an ARMA(1,1)-GARCH(1,1)-NIG model. Their out-of-sample performance is of interest rather than their in-sample properties. The models produce one-day ahead forecasts which are evaluated using three statistical tests: VaR-test, VaRdur-test and Berkowitz-test. All three tests are concerned with the the tail events, since our time series models are often used to estimate downside risk.

When the two models are applied to data on Canadian stock market returns, our three statistical tests point in the direction that the ACD model and the GARCH model perform similarly. The difference between the models is small.

We finish with comments on the model uncertainty inherit in the comparison.

Keywords: time series, higher moments, risk, autoregressive conditional density, ACD, out-of-sample, monte carlo, bootstrap

Contents

Part I

Introduction & Theory

1 Introduction

This thesis will compare two models' out-of-sample performance with respect to downside risk. Market movements are difficult to foresee and rare. They also impact a lot of people. When applying statistical models to predict the impact of these extreme events, particular care must be taken in selecting appropriate models. Of equal importance is to evaluate these models using proper measures. Volatility—the second moment—is one measure of downside risk but higher moments such as skewness and kurtosis—the third and fourth moment—are of importance as well.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are frequently used when forecasting volatility and downside risk. In a standard GARCH model with skew Student-t distributed innovations, the mean and variance parameters vary over time but the parameters that control skewness and shape are assumed to be fixed over time. Implicit in this model specification is that the conditional distribution of $\tilde{\varepsilon}_t$ (the GARCH standardized residuals) is constant over time. There is no economic or financial argument for why it should be so. An extension of the GARCH model is the Autoregressive Conditional Density (ACD) model. It does, by contrast, allow for time variation in the higher moments skewness and kurtosis. The ACD model is more complex than the GARCH and is less used in practice.

The GARCH model of Bollerslev (1987) predates the ACD model of Hansen (1994). Some research has been done, which will be briefly mentioned in the following subsection. The results are not solid. It is not clear if incorporating the skewness and kurtosis lead to models that better fits the data (in-sample) or is better at prediction (out-of-sample). The latter has been of especially little focus, and this thesis will contribute to the literature by testing if the ACD model is better than the GARCH model at forecasting according to measures of financial downside risk.

1.1 Literature Review

The GARCH model was developed in Bollerslev (1987) as an extension to the ARCH model defined in Engle (1982). GARCH models have been used extensively in the literature. The ACD model used in this thesis is defined in equation (5) on page 10. The ACD model of Hansen (1994) extended the GARCH model by allowing for the higher moments skew and shape to vary over time. Hansen let the law of motion be that a weighted difference ($moment_t - d \cdot moment_{t-1}$) depended linearly on yesterday's white noise term z_{t-1} and z_{t-1}^2. Other laws of motion have been suggested, for example a piecewise linear dynamics so that negative shocks $z_{t-1} < 0$ impact the $moment_t$

differently than positive shocks does.

The current literature on ACD models is summarized in Table 0. It is an extensive table and should be self-explanatory. Nonetheless we comment a few things below.

Data, the third column, reveals that mostly equity indices (EI) have been used. Although some asset classes such as Foreign Exchange (FX) and bonds have also been used. It is logical since equity is a big and important asset class.

Data *frequency* matters. In the fourth column we see that daily is the most used frequency. Hansen (1994) who developed the first ACD model used monthly and weekly data on US Treasury and FX. Hansen found in-sample evidence for the ACD model. Rockinger and Jondeau (2002) used three equity indices based on weekly data but found no in-sample evidence. One year later, however, when they researched daily data, Jondeau and Rockinger (2003) found in-sample evidence of both skew and kurtosis. In addition, Bond and Patel (2003) and Harvey and Siddique (1999) found time varying skew in weekly and monthly data.

The two columns *Distribution* and *Moments* reveals that a lot of different distributions have been used with varying number of moments. Researchers have explored different laws of motion for skew and shape parameters, as well as different distributional assumptions. Notably, the NIG is a general distribution often used, as it incorporates the normal as a special case.

The results have not been solid regarding the moments. Brooks et al. (2005) found no in-sample evidence for the ACD model, whereas Premaratne and Bera (2000) found in-sample evidence for skew and kurtosis.

The last row of Table 0 states what we will do in this thesis and how it connects to the current literature. Most of the research have been regarding in-sample inference, therefore this thesis will contribute to the literature by doing an out-of-sample comparison. Although the interested reader can infer some in-sample differences from Appendix section 11.1 we will, due to page constraints, only discuss out-of-sample performance.

1.2 Research Question and Method

The GARCH model is heavily used in the financial literature and in practice by financial institutions. The ACD model extend the GARCH model by allowing for a time varying skewness and shape. *In this thesis we will compare the predictive abilities—according to measures of downside risk—of the GARCH and ACD models in an out-of-sample setting.*

The method will be as follows. We begin by looking at descriptives of the data to see what models can be appropriate. After that, we fit one ACD model and one GARCH model to our data-set. Then we discuss if the models are adequate enough to use in an out-of-sample setting. Lastly, we produce forecasts using a rolling window and backtest these with three statistical tests. From these tests we can draw some conclusions to answer our research question, and we comment the conclusions from the tests with a short discussion on parameter uncertainty.

Table 0: Literature.

Year	Authors	Data [a]	Frequency	Disitribution [b]	Moments [c]	Comparison [d]
1994	Hansen	US Treasury & FX	M & W	G skew t	4	in-sample
1999	Harvey and Siddique	Several EI	D & W & M	t	3	in-sample
2000	Premaratne and Bera	NYSE	Daily	PIV	4	in-sample
2002	Rockinger and Jondeau	Three EI	Weekly	Entropy Density	4	in-sample
2003	Jondeau and Rockinger	Currency & EI	Daily	G skew t	4	in-sample
2003	Bond and Patel	Real Estate	Monthly	G skew t	3	in-sample
2003	Brannas and Nordman	NYSE	Daily	PIV & LGG	3	in-sample
2005	Leon et al.	FX & EI	Daily	GC Normal	4	in-sample
2005	Brooks et al.	EI & Bonds	Daily	t	3	in-sample
2009	Wilhelmsson	SP500	Daily	NIG	4	in-s & out-s
2010	Ergun and Jun	SP500	5min	G skew t	3	in-s & out-s
2016	Lindberg	Canadian EI	Daily	NIG	4	out-of-sample

[a] The type of data used. Abbreviations are the following.
FX: Foreign Exchange.
EI: Equity Index.
NYSE: New York Stock Exchange (an EI).
SP500: S&P500 (an EI).

[b] Stands for the assumed distribution for the white noise term. Abbreviations are the following.
NIG: Normal Inverse Gaussian.
t: Student's t.
G skew t: Generalized skew Student's t.
PIV: Pearson Type IV.
GC Normal: Gram-Charlier Normal Expansion.

[c] Refers to how many moments are time-varying, either 3 if skewness included or 4 if skewness and kurtosis included.

[d] States if the research regarded in-sample inference or out-of-sample or both. Most research have done in-sample comparison so in this thesis we do out-of-sample comparison.

2 Time Series Background

In this section we define some basic concepts in time series analysis.

2.1 Prices and Returns

Let P_t be the price of an asset at time t. The return is given by

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

The log return is used instead of the fraction $\frac{P_t}{P_{t-1}}$ because of its statistical properties. (Tsay, 2013:4).

2.2 Information Set

The information set is

$$\mathcal{F}_{t-1} = \{r_1, ..., r_{t-1}\}$$

and this is used in the definition of conditional mean

$$\mu_t = E[r_t|\mathcal{F}_{t-1}]$$

and the conditional variance

$$\sigma_t^2 = \text{var}(r_t|\mathcal{F}_{t-1}).$$

We will often drop the prefix "conditional" and simply refer to them as the mean and variance. It is understood from the context if variance σ_t^2 or unconditional variance $\text{var}(r_t) = \sigma^2$ is the discussed quantity.

2.3 ACF

In a stationary time series $\{r_t\}$ the autocorrelation is defined as

$$\rho_k = \frac{\text{cov}(r_t, r_{t-k})}{\sqrt{\text{var}(r_t)\,\text{var}(r_{t-k})}} = \frac{\text{cov}(r_t, r_{t-k})}{\text{var}(r_t)} = \frac{\gamma_k}{\gamma_0} \tag{1}$$

where r_{t-k} is the returns k^{th} lag, and γ_k is the autocovariance.

2.4 Ljung-Box Test

In a Ljung-Box test we test the null that data is not serially correlated, against the alternative that data is serially correlated. The test statistic is

$$Q = n\,(n+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{n-k} \sim^{H_0} \chi^2(K).$$

where n is the sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag k, and K is the number of lags being tested. (Ljung and Box, 1978:297-303)

We will use a significance level of 5 percent. We will use $K = 1$ and $K = \ln(T)$ where T is the sample size, for example $K = \ln(T)$ is rounded to $K = 9$ when we use Ljung-Box test in the estimation window as will be discussed later on.

2.5 White Noise

Let $\{z_t\}$ be a time series. If $\gamma_0 = \sigma_z^2 \in \mathbb{R}$ and $\gamma_k = 0$ for lags $k > 0$ then we say that $z_t \sim \mathit{WhiteNoise}(0,\ \sigma_z^2)$. In other words, a process with a zero mean and zero covariance for lags $k > 0$ is said to be a white noise process. Many different distributions for z_t can be used, two exampes are the normal distribution and the Normal Inverse Gaussian (NIG) distribution (Brooks, 2002).

2.6 NIG Distribution

The probability density function (PDF) of Normal distribution is

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

The Parameters are μ, σ^2. The PDF of Normal Inverse Gaussian (NIG) distribution is

$$\frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}\exp\left[\delta\gamma+\beta(x-\mu)\right]$$

where K_1 denotes a modified Bessel function of kind three. The parameters are μ location α tail heaviness, δ scale parameter, β asymmetry parameter, and $\gamma = \sqrt{\alpha^2-\beta^2}$. The NIG has mean $\mu + \delta\beta/\gamma$ variance $\delta\alpha^2/\gamma^3$ skewness $3\beta/(\alpha\sqrt{\delta\gamma})$ and kurtosis $3+3(1+4\beta^2/\alpha^2)/(\delta\gamma)$. As seen from the formulas the normal distribution is a special case of the NIG distribution, where we have let $\beta = 0, \delta = \sigma^2\alpha$, and $\alpha \to \infty$.

A NIG distribution will, in this thesis, be used for the white noise term. Thus a model can be written `ARMA(1,1)-GARCH(1,1)-NIG` or `ARMA(1,1)-ACD(1,1)-NIG` although we will refer to them as GARCH and ACD.

3 Models

In this section we define three statistical models from time series analysis: ARMA, GARCH and ACD. We also define MLE and some information criterions.

3.1 ARMA model

The $AR(k)$ model is written

$$r_t = c + \sum_{i=1}^{k} \varphi_i r_{t-i} + \varepsilon_t.$$

where $\varphi_1, \ldots, \varphi_i$ are the parameters, c is a constant, and the random variable ε_t is white noise. The $MA(q)$ model is written

$$r_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where the θ_i for $i = 1, ..., p$ are the parameters of the model, μ is the expectation of r_t and the ε_{t-i} are white noise error terms. The $ARMA(k, q)$ model contains the $AR(k)$ and $MA(q)$ models, and is written

$$r_t = c + \varepsilon_t + \sum_{i=1}^{k} \varphi_i r_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

By letting $k = 1, q = 1$ we get the ARMA(1,1) model

$$r_t = c + \varepsilon_t + \varphi_1 r_{t-1} + \theta_1 \varepsilon_{t-1}.$$

In this thesis we will use an ARMA(1,1) model for the mean equation in a GARCH or ACD model. We do this to remove any potential movements in the mean, and it is often done with financial data before applying a GARCH model.

3.2 GARCH model

In the GARCH(1,1) model, we have that

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{2}$$

where μ_t is modeled by the *mean equation* e.g. ARMA(1,1) and the innovation is

$$\varepsilon_t = \sigma_t z_t \tag{3}$$

where z_t is white noise and assumed to follow some distribution, in this thesis a NIG distribution. Note that z_t is a random variable (in the easiest models z_t is assumed to be a standard normal) and σ_t is depending on time. The *variance equation* (sometimes called *volatility equation*) is

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4}$$

and the constraints are $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, $\alpha_1 + \beta_1 < 1$. For details on GARCH models and their properties see Tsay (2010:109-173).

We will use an `ARMA(1,1)GARCH(1,1)-NIG` in this thesis, or in other words we have the model $r_t = \mu_t + \varepsilon_t$ where μ_t is modeled by ARMA(1,1) and ε_t is modeled by GARCH(1,1)-NIG.

3.3 ACD model

One extension of the GARCH model is the ACD model. In the GARCH model we let (μ_t, σ_t) vary over time and all other parameters (see the PDF formula for the NIG distribution). In the ACD model, on the other hand, we let $(\mu_t, \sigma_t, \rho_t, \xi_t)$ vary over time where ρ_t is the shape and ξ_t is the skew. The unconstrained motion dynamics of the parameters ρ_t and ξ_t are denoted by $\bar{\rho}_t$ and $\bar{\xi}_t$. Let f and g be the motion dynamic functions for those parameters, and let $\Psi(.)$ be a suitable transformation function. Then we have the ACD model:

$$\begin{aligned}
&r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \\
&z_t \sim NIG\left(0, 1, \rho_t, \xi_t\right) \\
&\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&\Psi\left(\rho_t\right) = f\left(\bar{\rho}_t\right) \\
&\Psi\left(\xi_t\right) = g\left(\bar{\xi}_t\right)
\end{aligned} \tag{5}$$

For a detailed discussion on the motion dynamics and transformation functions see Ghalanos (2015:5-7).

In the GARCH model ρ_t and ξ_t are time invariant whereas in the ACD model we allow the skew and shape to vary over time. In other words we do not only let the first and second moment vary over time (μ_t, σ_t) but also let the third and fourth moment $(\rho_t,\ \xi_t)$ to vary over time. We will use an `ARMA(1,1)-ACD(1,1)-NIG` in this thesis. In practice there is a problem with estimating the ACD model because higher moments require so many data points.

3.4 MLE

To estimate the model parameters we use maximum likelihood estimation (MLE). Let $\ell(\theta; r)$ be the log-likelihood function, let θ be the parameters to be estimated and let $r = (r_1, \ldots, r_T)$ be the return data. The maximum likelihood estimate is

$$\hat{\theta}_{ML} = \arg\ \max_{\theta} \ell(\theta; r) \tag{6}$$

where the likelihood function is the product of the conditional densities

$$\mathcal{L}(\theta; r) = f(r_1 \mid \theta) \cdot f(r_1, r_2 \mid \theta) \cdot \ldots \cdot f(r_1, \ldots, r_T \mid \theta)$$

(Liero & Zwanzig, 2012). Information criterions are measure of the goodness of fit and the complexity of the model. We want a good fit but a parsimonious model. When comparing similar models to each other, information criteria can help guide us in choosing the best one. There are four information criterias presented in the output when you fit a GARCH or ACD model to a dataset, these are Akaike (AIC), Bayes (BIC), Shibata, Hannan-Quinn. They are defined as

$$\begin{aligned}
Akaike &= \frac{-2\ell}{N} + \frac{2m}{N} \\
Bayes &= \frac{-2\ell}{N} + \frac{m\ln(N)}{N} \\
Shibata &= \frac{-2\ell}{N} + \frac{(2m\ln(\ln(N)))}{N} \\
Hannan - Quinn &= \frac{-2\ell}{N} + \ln\left(\frac{(N+2m)}{N}\right)
\end{aligned} \tag{7}$$

4 Test of Forecasting Accuracy

In this section we present the theory needed to interpret and understand the forecasting comparison in the data analysis. After defining Value at Risk we explain VaR-test and VaRdur-test Then we introduce ways of testing the density forecasts via Berkowitz-test and HL-test. We begin, however, with describing how the forecasts are produced.

4.1 Rolling Window Procedure

Forecasts are produced using a rolling window procedure. When referring to the datasets we will use the notation from Table 1. Thus we can have convenient notation for both in-sample discussion and out-of-sample discussion. From the table we note that $w_E + w_F = T$. In this thesis $T = 5066$ and $w_E = 2000$ so $w_F = 3066$. We have data from $t^* = 1$ to $t^* = T$ or equivalently, from $t = 1 - w_E$ to $t = w_F$.

Table 1: Forecasting window and estimation window.

t*	t	comment
1	$1 - w_E$	in-sample
...	...	in-sample
w_E	0	in-sample
$w_E + 1$	1	out-of-sample
...	...	out-of-sample
$w_E + w_F$	w_F	out-of-sample

We can illustrate the roll by this bullet list:

- first estimation window: $t^* = (1, ..., w_E)$
- second estimation window: $t^* = (1 + 1, ..., w_E + 1)$
- third estimation window: $t^* = (1 + 2, ..., w_E + 2)$
- and so on, until,
- last estimation window: $t^* = (w_E, ..., w_E + w_F - 1)$

Let's outline how a GARCH model uses the rolling window procedure to create its volatility forecasts. The first estimation window is used to forecast $\hat{\sigma}_{t=1}$. The second estimation window is used to forecast $\hat{\sigma}_{t=2}$. The last estimation window is used to forecast $\hat{\sigma}_{t=w_F}$. Thus we use a "rolling window" so that the size of each estimation window is the same. From the rolling window procedure we end up with w_F number of estimations ($\hat{\sigma}_1$, $\hat{\sigma}_2$, . . . , $\hat{\sigma}_{w_F}$).

When doing a GARCH or ACD model it is up to the researcher to decide how often the parameters should be updated.[1] If they are updated every day then we will have w_F number of estimated parameters. For economic reasons the true parameters do not shift a lot from day to day, so updating the parameters every day incurs more cost than benefits. Every 10 or 50 days is sufficient for most applications.

4.2 VaR Definition

VaR increased in popularity when J.P. Morgan published RiskMetrics in 1994 (Longerstaey and Spencer, 1996).

Let r_t be the return of an asset, let x be a real number and let a be a probability. Define

$$VaR_{t,a} = \inf\{x : \Pr(r_t < x) \leq a\}. \tag{8}$$

In other words, VaR is nothing but a quantile of the returns distribution. An informal definition could be that $VaR_{t,a} = x$ means the probability that the return r_t is less than the number x is at most a. Often we will drop the index t and just write VaR_a. The formula to empirically estimate VaR is

$$VaR_{t,a} = \hat{\mu}_t - q_{1-a}\hat{\sigma}_t. \tag{9}$$

where q_{1-a} is the $(1-a)$ quantile of the distribution, $\hat{\mu}$ and $\hat{\sigma}$ are estimates from a given econometric model e.g. the GARCH. [2] In this thesis we will use $a = 0.01$ and $a = 0.05$.

Our rolling window procedure combined with GARCH or ACD model produce a forecast of the distribution one day ahead, but how do we test the accuracy? We will use three tests, presented in the subsections that follow. But first, a definition of violations.

4.3 Violation

Define the indicator variable I_t for $VaR_{t,a}$ as

$$I_t = \begin{cases} 0, & \text{if } r_t \geq \text{VaR}_{t,a} \\ 1, & \text{if } r_t < \text{VaR}_{t,a} \end{cases}. \tag{10}$$

Hence if $I_t = 1$ a *violation* has occurred. An informal definition could be that our realized return r_t was lower than our estimated $VaR_{t,a}$.

[1] In R this is done by adjusting $refit.every$ varibale in the rugarch package.

[2] As an example $q_{1-0.05} = 1.65$ for a standard normal. Since $q_{1-0.05} > 0$ and $\hat{\mu} < \hat{\sigma}$ VaR is a negative number.

We wish for the number of violations to be close to the expected number of violations. We also wish for the violations to be independent or un-clustered. (Jorion, 2007:143-152.) Define

$$expected = a \cdot w_F$$

$$actual = \sum_{t=1}^{w_F} I_t$$

so if $expected = 0.01 \cdot 3000 = 300$ but $actual = 311$ we have too many violations and our stated VaR is underestimating the risk.

4.4 Conditional Coverage Test (VaR-test)

A conditional coverage test (CC-test) evaluates if actual number of violations are the same as the stated number, *and* at the same time it checks if violations are clustered.

Let $\pi_{ij} = \Pr(I_t = j \mid I_{t-1} = i)$. The violations I_t can be modeled with a Markov chain having transition probabilities

$$\begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}. \tag{11}$$

The null hypothesis is

$$H_{0,cc} : \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} 1-a & a \\ 1-a & a \end{bmatrix}. \tag{12}$$

The test statistic is

$$LR_{cc} = -2 \ln \frac{(1-a)^{T_{00}} a^{T_{01}} (1-a)^{T_{10}} a^{T_{11}}}{(1-\pi_{00})^{T_{00}} \pi_{01}^{T_{01}} (1-\pi_{10})^{T_{10}} \pi_{11}^{T_{11}}}. \tag{13}$$

where T_{ij} are the actual number of times $I_{t-1} = i$ is followed by $I_t = j$. Under the null hypothesis LR_{cc} is distributed as $\chi^2(2)$. (Christoffersen, 1998)

From the definition of π_{ij} it is obvious that the test only concerns two consecutive days. That is a short memory. For financial reasons it might be interesting to look two days back. Using the CC-test, we cannot. This is a major critique and one of the reasons for creating the test in the next subsection.

4.5 VaR Duration Test (VaRdur-test)

The VaR Duration Test (VaRdur-test) is a likelihood ratio test. Consider the process $\{I_t\}$. If the violations have no memory, it follows an exponential distribution as this is the only memoryless distribution. We test

$$H_0 : I_t \sim Weibull(b = 1)$$

since $Weibull(b = 1)$ is the exponential distribution. If the likelihood ratio test statistic is too big, the null is rejected and violations are not independent—so we wish to accept the null. A LR test will be explained in the next subsection so we don't go into details here. For details see Christoffersen and Pelletier (2004).

4.6 Berkowitz Tail Test (Berkowitz-test)

Let r_t be a stochastic process that is forecasted using $\mathcal{F}_{t-1}$ with PDF $f(r_t)$ and CDF $F(r_t)$. Let $\hat{f}(.)$ and $\hat{F}(.)$ be the forecasted PDF and CDF respectively. Rosenblatt (1952) showed that $\hat{F}(r_t)$ are iid and distributed as $U(0, 1)$ where r_t is the ex post realized profit or loss and $\hat{F}(.)$ is the ex ante forecasted density.

Let

$$z_t = \Phi^{-1}[\hat{F}(r_t)]$$

so that we have transformed a uniform random variable to $N(0, 1)$ and we have that z_t is a standard normal. The null hypothesis of the full Berkowitz test is that z_t are iid and distributed as $N(0, 1)$ or more compactly written

$$H_0 : z_t \sim \text{iid } N(0, 1)$$

against the alternative that z_t follows and AR(1) process $H_1 : z_t - \mu = \phi(z_{t-1} - \mu) + \varepsilon_t$. Of course we can use AR(k) but in this thesis we will use AR(1). The test statistic is

$$LR_{full} = -2\ell(0, 1, 0) + 2\ell(\hat{\mu}, \hat{\sigma}^2, \hat{\phi})$$

which under the null is distributed as a $\chi^2(3)$.

Having said that, the above description is for the full Berkowitz test but in this thesis we use Berkowitz tail test. One drawback with the full test is that risk managers care about large losses (one tail of the distribution) not the entire distribution. Having a bad forecast for the center of the density is not of concern to them. For this reason we consider instead a test based on the censored likelihood. Truncate z_t into z_t^* so

$$z_t^* = \begin{cases} \text{VaR}_t, & \text{when } z_t \geq \text{VaR}_t \\ z_t, & \text{when } z_t < \text{VaR}_t \end{cases}.$$

In this thesis the cutoff point VaR will be for $VaR_{0.05}$ and $VaR_{0.01}$. The null is

$$H_0 : z_t^* \sim \text{iid } N(0,1). \tag{14}$$

The LR test statistic is is

$$LR_{tail} = -2\ell(0,1) + 2\ell(\hat{\mu}, \hat{\sigma}^2) = -2uLL + 2rLL \tag{15}$$

where uLL and rLL stand for unconditional log-likelihood value and restricted log-likelihood value, respectively. Under the null our test statistic LR is distributed as $\chi^2(2)$. In this thesis we use a Berkowitz tail test i.e. with the truncated z_t^* because the tail is often more interesting in applications.

4.7 Summary of Tests

The CC-test and the VaRdur-test are for the violations. The former tests if violations follow a Markov chain with consistent transition probabilities, whereas the latter is more sophisticated and tests if violations follow a memoryless process. The Berkowitz-test is a test for the density of the tail.

In all these test we do not wish to reject the null hypothesis. In the data analysis part of this thesis we will therefore look for high p-values as an indication of a good model.

Part II

Analysis & Conclusion

5 Data

In this section we present some summary statistics for our return series.

The data-set is prices on an ETF that follows the Canadian stock market. In the output below, which present a summary of returns, we see that our data:

- have 2000 observations in the estimation window (and 3066 in the forecasting window) but no missing values (seen by w_E and *NA's*),
- is close to zero (seen by *min max median mean SE.mean*)
- is not normal (seen by the values for *kurtosis* and *skew*),
- have an average volatility of 1.49/100 (seen by *std.dev*),
- returns are correlated according to a Ljung-Box test of r_t indicating a need to model the mean equation, [3]
- have ARCH effects according to a Ljung-Box test of r_t^2 indicating a need to model the volatility equation.

Statistic	Value	Statistic	Value
w_E	2000	var	2.22e-04
NA's	0	std.dev	1.49e-02
min	-1.165	kurtosis	5.21
max	1.167	skewness	-0.44
mean	2.96-04	LB r_t p-value	0.03
SE.mean	2.09e-04	LB r_t^2 p-value	0

[3] Although using a constant would also work since the mean is so close to zero.

6 Model comparison

In this section we compare the two models. In subsection 6.1 we compare the fit (in-sample). In subsection 6.1 and forward, we compare their forecasts (out-of-sample).

6.1 Fit Comparison

We use the word *fit* when discussing in-sample properties and the word *forecast* when discussing out-of-sample forecasts. This fit comparison is not a major part of the thesis but should be seen as a first step. We fit two models to the estimation window, conclude they are different and then move on to compare their out-of-sample forecasts. Therefore this subsection should be seen partly as a way of explaining the differences between the two models, and partly as an introduction to the forecast comparison because that is what we are most interested in.

Output from the *fit* are on page 30 in Appendix. In the paragraphs below we will comment this output.

ARCH-type models are appropriate for our data set. This can be seen from the high p-values of Ljung-Box Test on Standardized Residuals $\tilde{\varepsilon}_t$ and Ljung-Box Test on Standardized Squared Residuals $\tilde{\varepsilon}_t^2$.

For the information criteria (Akaike, Bayes, Shibata, and Hannan-Quinn) the GARCH have slightly higher values. However, a formal LR test need to be done if you want to see if one model is statistically significantly better.

From the estimated parameter values and their standard errors we conclude that most parameters are significant for both models.

By comparing the estimated values for the classical coefficients *ar1, ma1, alpha1, beta1, omega* we see that both models have similar values as it should be because the models themselves are similar.

Given the result from this fit output, we believe the models fit the data well enough to be used in predicting. Thus we can move on with the more interesting part of our analysis, namely comparing the predicting performance.

6.2 Forecast Plots

In section 6.3 we perform a VaR-test and VaRdur-test in the first subsection as well as a Berkowitz-test in the second subsection. As a way to visually illustrate what we will statistically test, this section 6.2 will create two Figures: one figure with the VaR plots as well as one figure with plots of the density forecast.

6.2.1 VaR plot

Since we perform many test for the VaR forecasts it is instructive to plot VaR_t. In Figure 1 we have two VaR Plots (one for ACD 0.01 and one for ACD 0.05) as well as the difference between them. A VaR plot shows the realized return in gray whereas the forecasts $(VaR_1, VaR_2, ..., VaR_{w_F})$ are plotted in the black line that closely follows the realized returns. Whenever the realized return (gray) is less than the VaR line (black) a violation has occurred.

The 0.05 VaR line is of course tighter than the 0.01 VaR line by definition. The VaR line is accompanied by the realized return plotted in gray. When returns are more volatile, such as from $t = 1000$ to $t = 1500$ the VaR line is much lower than usual.

The third graph is $\textit{Difference}_t = VaR_{0.01,t} - VaR_{0.05,t}$ and it illustrates that in periods of high volatility the 1 percent VaR differs the most from 5 percent VaR.

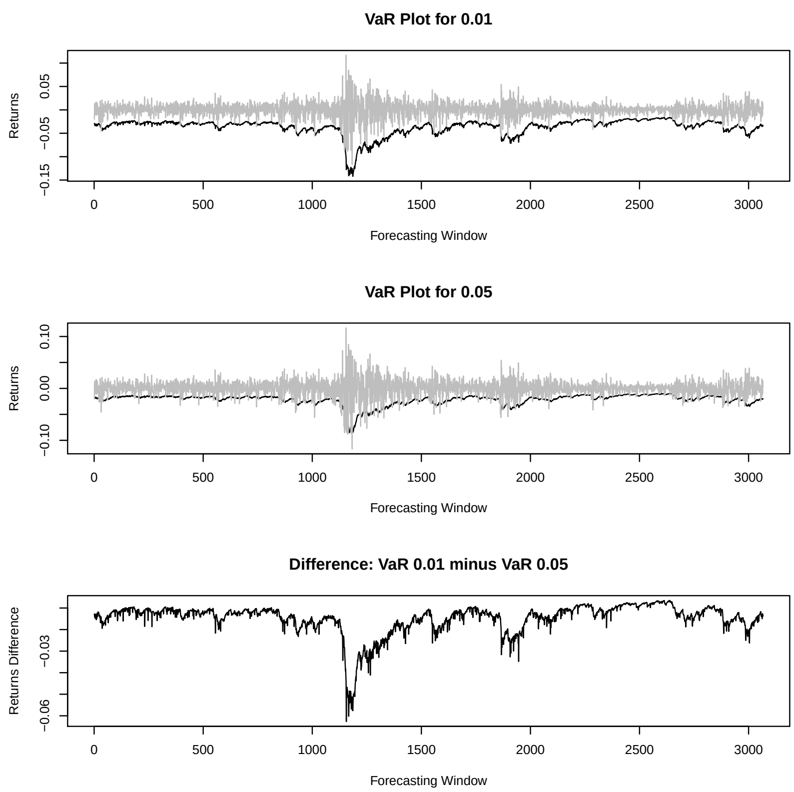

Figure 1: ACD model VaR Plot for 1 percent VaR (first graph) and 5 percent VaR (second graph). The third graph is the first graph minus the second graph i.e. VaR 0.01 minus VaR 0.05.

6.2.2 Plots of density forecast

In Figure 2 we visualize the distinguishing factor between the GARCH model and the ACD model, namely the constant skew & kurtosis versus the time varying skew & kurtosis. In the ACD model we forecast the NIG parameters ρ_t & ξ_t based on $\mathcal{F}_{t-1}$ for all times t in the predicting window, whereas in the GARCH model we simply estimate ρ & ξ using MLE. Be that as it may, the differences between the two models are small. For example the mean are close to each other: $-0.16 \approx -0.17$ and $1.81 \approx 1.96$ as seen in the four plots. By studying the y-axis we see that skew and shape have approximately the same values on the y-axis indicating their range is similar. Hence the models are similar.

The GARCH model let skew and shape be constant over time, so by looking at the right part of Figure 2 you might ask: why does it change? And why does it look like a staircase? Because we update parameters. Recall from section 4.1 that we do update the GARCH model parameters to not use old data when predicting volatility. The variable $refit.every$ is the number of days until we do a refit, and in this thesis $refit.every = 200$. This is the rationale for the staircase look in Figure 2. After 200 days all the GARCH model parameters $(\omega, \alpha_1, \beta_1, skew, shape)$ are updated. The staircase has 16 steps. This is because $w_F/refit.every = 3066/200 = 15.33$ which is rounded up to 16.

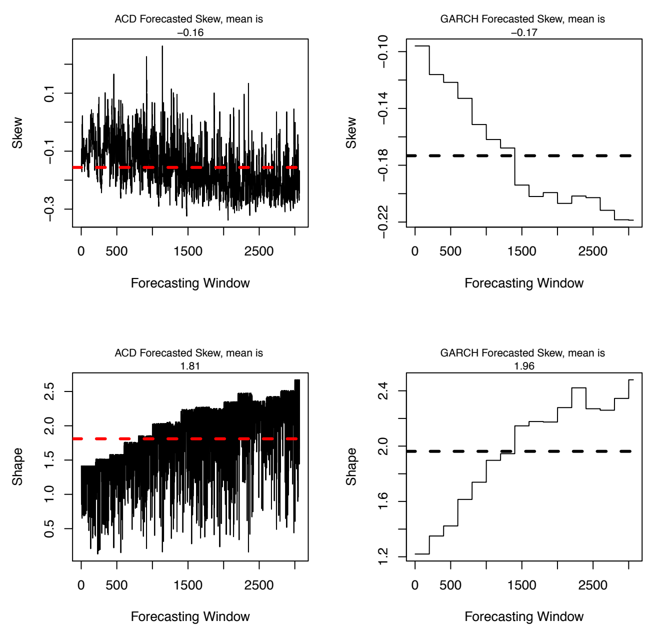

Figure 2: Density forecast plot for ACD (left) and GARCH (right) for skew (upper) and shape (lower). In ACD skew and shape vary over time, whereas in GARCH they are constant—although they are updated when parameters are.

6.3 Forecast Comparison

We use begin with a VaR-test and VaRdur-test, then we move on and do a Berkowitz-test.

6.3.1 Volatility Forecast Comparison via Value at Risk Tests

Both the VaR-test and VaRdur-test are used heavily in risk management. Value at Risk is a measure summarizing how much money an investor can loose over a certain time horizon (e.g. one day) with a certain level of confidence (e.g. five percent) under normal market conditions. For the mathematical definition see equation (8). In the calculation of VaR the hardest problem is to forecast the volatility, where ARCH-type models are used, and that is the rationale for using VaR backtesting procedures in order to differentiate between our two models.

When evaluating how good the model is for VaR application, we use tests based on violations. We want the number of violations to be correct and not to cluster. Clustering is tested in a different way in the VaR-test and in the VaRdur-test. The VaR-test is simpler because only yesterday's violation I_{t-1} is considered whereas the violation from day $t-2$ and earlier are ignored. This can be understood by studying the Markov chain matrix in H_0 on page 14. The VaRdur-test, on the other hand, does not have this flaw as this test takes all previous violations into account by testing if the sequence of violations $\{I_t\}$ follows an exponential distribution. If the null is accepted then we conclude violations are not clustered. By this reasoning we should give more weight to the result of VaR-test than VaRdur-test. Why bother with the CC-test if we consider the VaRdur-test to be better? Because the CC-test is used in practice by financial institutions.

Recall that a is the probability in $VaR_{t,a}$ from definition (8). We use $a = 0.05$ and $a = 0.01$. Even though the statistical power is higher for 0.05 than for 0.01 since there are more violations to calculate the data on, it is more important in practice to consider $VaR_{0.01}$ than to consider $VaR_{0.05}$ because risk managers and regulators care more about the 1% VaR. To us this means that for practical applications the 0.01 models are more interesting to compare than the 0.05 models.

	expected	actual	cc.LRp	cc.Decision
ACD(0.01)	30	30		
GARCH(0.01)	30	30		
ACD(0.05)	153	189	0.0029	Reject H0
GARCH(0.05)	153	190	0.0015	Reject H0

Table 3: CC-test null hypothesis: VaR violations are correct and independent. High p-value in column *cc.LRp* is good.

In Table 3 we use 0.01 and 0.05 as a on row 1-2 and 3-4. For 0.05 models the decision rule from the CC-test is to reject as seen by the column *cc.Decision*. This rejection of the null is based on comparing the p-value *cc.LRp* with our chosen significance level 5 percent. The null is rejected when $cc.LRp < 5\%$. The observed LR statistic is not displayed in the table but of course calculated before being able to calculate the p-value. The equation for calculating the LR statistic is found in equation (13).

Comparing row 1 with row 2 we see that the 0.05 models perform similarly since the *Decision* is the same for both models. When comparing ACD(0.01) against GARCH(0.01) we see that p-values are missing because there is not enough data. Too few violations have occured. This is a problem with the CC-test for low levels of p—violations occur too rarely to be tested. It is good, though, to see that $actual = expected$ for both models indicating they at least have the correct number of exceedances—but we cannot test if they are independent using CC-test, for that we need a better test that utilizes more information.

	b	uLL	rLL	LRp	Decision
ACD(0.01)	1.07	-164.1	-164.2	0.6839	Fail to Reject H0
GARCH(0.01)	1.15	-163.8	-164.2	0.3835	Fail to Reject H0
ACD(0.05)	0.93	-711.7	-712.8	0.1515	Fail to Reject H0
GARCH(0.05)	0.93	-714.6	-715.6	0.1718	Fail to Reject H0

Table 4: VaR Duration Test null hypothesis: VaR violations lack memory. High p-value is good. Column b is the Weibull parameter.

Table 4 is read in a similar way to Table 3. Again we have coverage ratios both 0.01 and 0.05 for our two models, creating four rows. The decision is found by comparing *LRp* to our significance level of 5 percent. In the *Decision* column we see that the null is failed to be rejected for all models, which is good.

The b column display the estimated value for the shape parameter b in the Weibull distribution. When $b = 1$ it is equivalent to an exponential distribution which is memoryless. We want violations to lack memory.

The uLL and rLL columns displayed are not of major interest. They are used to calculate LR in a similar way as equation (15). The *LRp* column is the most important one. The value LRp is calculated from the test statistic LR. According to the higher p-values all models fail to reject, which is good. The major take away from Table 4 all models are good to VaRdur-test.

6.3.2 Density forecast Goodness of Fit Comparison via Berkovitz-test

Berkowitz-test is for the tail density. The result from the Berkowitz-test is seen in Table 5. In the Berkowitz-test the values 0.01 and 0.05 stands for the tail cutoff point. This table is read in a similar way that the VaR-table. A high p-value is good since the the Berkowitz tail test (2001) evaluates the goodness of fit regarding the tail, and the null is that the fit is good. For 0.05 both models are bad. For 0.01 both models are good.

	uLL	rLL	LRp	Decision
ACD(0.01)	-154.2	-156.7	0.08403	fail to reject H0
GARCH(0.01)	-153.5	-156.2	0.06325	fail to reject H0
ACD(0.05)	-702.4	-709.6	0.00074	reject H0
GARCH(0.05)	-702.3	-710.2	0.00037	reject H0

Table 5: Berkowitz tail test null hypothesis: Data is N(0,1). High p-value is good.

6.4 Model Comparison Conclusion

The three statistical tests had one table each. Their theory was presented under subsections 4.4, 4.5 and 4.6. To summarize these forecasting test, and draw conclusions from them, we present Table 6 which have the following columns.

- Column *Test Performed* and *Test Null Hypothesis* states the name of the test and the null hypothesis. For an exact formulation of H_0 see their respective subsections.
- Column *Table* refers to the table number in this document.
- Column a stands for the tail cutoff point a and is either 0.01 or 0.05 as usual. This value a refers to the VaR_a so we can use $VaR_{0.01}$ or $VaR_{0.05}$.
- Column *Good Models* the most important column, consisting of a qualitative statement of which models are good according to the test. We arrive at the statement by comparing p-values to our significance level of 5%. In all tests we wish to fail to reject the null so whenever both models have a *higher* p-value than 5% we denote it as "both", and whenever both models have a *lower* p-value than 5% we denote it as "none". [4]

In Table 6 we see the extra costs of the ACD model are not worth it since we always draw the same conclusion for ACD as for GARCH in our statistical tests—it is either "both" or "none" so there is no difference. [5]

The result in Table 6 is logical because the GARCH model residuals and the ACD model residuals are very similar, suggesting that in practice the two models are not so different from each other.

[4] If for example the ACD model had a p-value lower than 5% and the GARCH model p-value higher than that, then we would write "ACD" (and vice versa) but this never happened.

[5] We have missing data NA in the first row due to the fact that there were too few violations to backtest. This was discussed earlier and it is a known flaw with the CC-test: for low values of a there are so few violations that testing is not possible.

Test Performed	Test Null Hypothesis	Table	a	Good Models
VaR-test	Viol. are correct and independent	3	0.01	NA
VaR-test	Viol. are correct and independent	3	0.05	None
VaRdur-test	Viol. lack memory i.e. Weibull(1)	4	0.01	Both
VaRdur-test	Viol. lack memory i.e. Weibull(1)	4	0.05	Both
Berkowits-test	Data is N(0,1) i.e. Tail is correct.	5	0.01	Both
Berkowits-test	Data is N(0,1) i.e. Tail is correct.	5	0.05	None

Table 6: According to column *Model Winner* which is a qualitative statement based on comparing the p-value from the test, ACD model is better than GARCH model for most tests.

7 Monte Carlo Bootstrap

So far we have compared ACD vs GARCH, but we have ignored model uncertainty. Here we illustrate that uncertainty. This section should merely be seen as a comment to the previous main result, rather than new conclusions.

The main result of this thesis is that the benefits of ACD model do not outweigh the computational costs, at least for our dataset and our backtesting prodecures. In these backtests, however, we used asymptotical results for our statistical tests. And these tests are based on the assumption of no model uncertainty. To illustrate the uncertainty in ARCH-type models we will perform estimation on simulated data.

The perfect thesis would do a Monte Carlo bootstrap for both models. We will only do it for the GARCH model, partly because it's very hard for ACD, and partly because this section is not the main issue of this thesis. One reason why it's difficult is that for the ACD models there are no analytical formulas readily available—instead we need to simulate when doing the forecasts. The model and samplig uncertainty is analyzed using a bootstrap procedure. Thus it would be very time consuming to do a bootstrap on the forecasts.

Even though this thesis is concerned with one-day ahead forecasts, it is useful to consider many step ahead forecasts. For example, banks need to not only present their 1 day VaR but also their 10 day VaR and in doing so they need to forecast the volatility 10 days ahead i.e. forecast $\hat{\sigma}_{t+10}$ based on $\mathcal{F}_t$.

Let *n.bootfit* be how many simulation based re-fits we do to generate the paremeter distribution. Let *n.ahead* be the length of the longest forecast, so that a value of 250 means that we forecast $(\hat{\sigma}_{t+1}, ..., \hat{\sigma}_{t+250})$ at day t. For a given forecast, e.g. the five-day ahead forecast, let *n.bootpred* be the number of bootstrap replications per parameter distribution used to generate the predicitve density. We employ a bootstrap with *n.bootfit = 100*, *n.ahead = 250*, and *n.bootpred = 500.* Results are found in Figure 3 and 4. Figure 4 is less important and the interested reader can look in Appendix on page 35.

In Figure 3 where we plot the bootstrap forecasts $\hat{\sigma}_{t+1}, ..., \hat{\sigma}_{t+250}$ based on $\mathcal{F}_t$ where t is 2015-05-20 i.e. approximately a year ago since this sentence was written. These volatility forecasts are the black line ("bootstrapped"). The bootstrap forecasts are close to the analytical forecasts [6] as seen by the distance between the black line and the red line ("forecast"). When performing a bootstrap we get the bootstrapped standard errors and from these we can calculate the confidence interval. Two confidence intervals are drawn, the most conservative one is 5% to 95% and the other one is 25% to 75%. One obvious conclusion from these bands is that we get more uncertain the farther into the future we go. Another obvious conclusion is that the 5% to 95% confidence interval is broader than the 25% to 75% confidence interval, and this is by definition, but in the plot we can see just how much more uncertain of the volatility forecast we are when higher confidence is required. A more interesting exercise is to compare the black line ("bootstrapped") and the green line ("filtered"). We see that for the first 50 days the forecast is good, since the forecast of sigma is close to the actual sigma, but at around day $t+60$ the actual volatility (in green) spikes. At day t we could not have foreseen this and that is why the actual volatility (in green) is much higher than the forecasted volatility (in black). Although, by being conservative about our forecasts we can see that the confidence interval 5% to 95% actually covers this event.

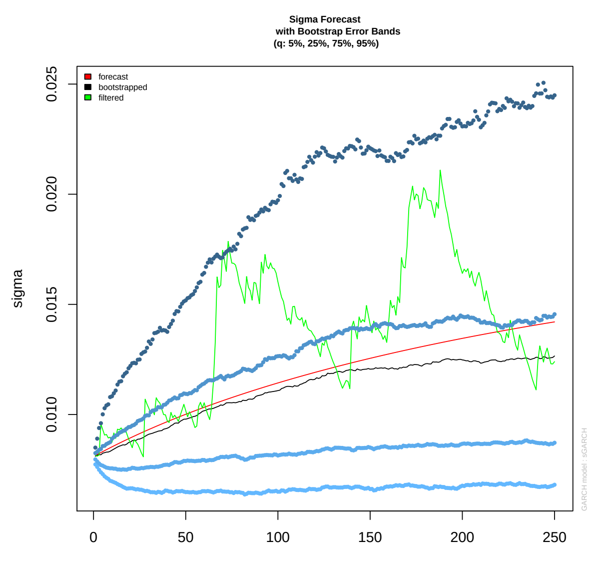

Figure 3: Bootstrap volatility forecast for 250 days ahead. The green line ("filtered") is ε_t^2 as a proxy for the actual volatility. The red line ("forecast") is the forecast based on analytical formulas derived under the assumption that the model is correct. The black line ("bootstrapped") is the forecast based on bootstrapping.

[6]By analytical we refer to using the estimation window to estimate the GARCH parameters, then forecast $\hat{\sigma}_1$ and from this use the GARCH fomula to recursively forecast $\hat{\sigma}_2$ etc.

8 Conclusion

We have evaluated an ARMA(1,1)-ACD(1,1)-NIG against an ARMA(1,1)-GARCH(1,1)-NIG. The focus has been on the one-day ahead forecast with extra attention to the tail events, since these are of interest in financial applications.

Three tests have been employed: VaR-test, VaRdur-test and Berkowitz-test. For our data on Canadian stock market returns, these three test point in the direction that the ACD model and the GARCH model perform similarly. This can be seen in Table 6. In addition, the residuals from these models are very similar suggesting that in practice there is a small difference between the two models.

We must also take the model uncertainty into account, not only sampling uncertainty. Inherit in these forecasts are model uncertainty, because the tests employed use asymptotical results. To shed light on model uncertainty we have performed a simulation with bootstrap. The outcome should be seen as a warning to not have a naive interpretation the forecasting power of the GARCH model.

8.1 Further Research

It requires more research—for example using other data-sets and other tests—to say for sure if the more complex ACD model is worth the extra costs.

Some possible tweaks to build upon the research produced in this thesis is to change some of the points below.

(I) Use other tests, since none of the tests presented in this thesis are perfect and not all of them point in the same direction. Also, the test should be used with a clear purpose in mind such as risk management or model fine tuning but not both. One test for the entire density, rather than the tail, is the test of Hong and Li (2001).

(II) Do 5 and 10 day ahead forecasts, as well as backtesting it, because 10 day VaR is is used for banks. And because we saw in the bootstrap plot that it is difficult to forecast 10 days ahead.

(III) Have other assets or even other asset classes. Perhaps their return series behave differently, for economic reasons, and this will affect the ACD and GARCH model.

(IV) Impose other distributions for the innovations, such as Johnson's SU or the GEV distribution used in extreme value theory.

(V) Perform a multivariate extension of ACD models is an extension of the GO-GARCH model of van der Weide (2002). For a M.Sc. thesis this can be interesting, and dimension reduction techniques would probably be needed.

(VI) Change the model family from standard GARCH to other models such as E-, AV-, GJR-, NAGARCH or apARCH. These models account for leverage effects observed in financial data. The leverage effect arise because negative news impact the volatility more than positive news does, and this is ignored by the standard GARCH model as seen by its symmetric news impact curve.

Part III

References & Appendix

9 References

Articles

Berkowitz, J. 2001, Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics*, 19(4), 465-474.

Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31(3), 307-327.

Bollerslev, T., 1987, A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *The Review of Economics and Statistics* 69(3), 542-547.

S.A. Bond and K. Patel. The conditional distribution of real estate returns: Are higher moments time varying? *Journal of Real Estate Finance and Economics*, 26(2):319-339, 2003.

K. Brannas and N. Nordman. Conditional skewness modelling for stock returns. *Applied Economics Letters*, 10(11):725-728, 2003.

C. Brooks, S.P. Burke, S. Heravi, and G. Persand. Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, 3(3):399-421, 2005.

Christoffersen, P. F., 1998, Evaluating Interval Forecasts, *International Economic Review* 39(4), 841-862.

Christoffersen, P. and Pelletier, D. 2004, Backtesting value-at-risk: A duration-based approach, *Journal of Financial Econometrics*, 2(1), 84-108.

Engle, R. F., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50(4), 987-1008.

Giot, P. and Laurent, S., 2004, Modelling daily Value-at-Risk using realized volatility and ARCH type models, *Journal of Empirical Finance* 11(3), 379-398.

B.E. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35(3):705-730, 1994.

C.R. Harvey and A. Siddique. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465-487, 1999.

E. Jondeau and M. Rockinger. Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699-1737, 2003.

Kupiec, P. H., 1995, Techniques for Verifying the Accuracy of Risk Measurement Models, *The Journal of Derivatives* 3(2), 73-84.

Ljung, G. M. and Box, G. E. P., 1978, On a measure of lack of fit in time series models, *Biometrika* 65(2), 297-303.

A. Leon, G. Rubio, and G. Serna. Autoregresive conditional volatility, skewness and kurtosis. *Quarterly Review of Economics and Finance*, 45(4):599-618, 2005.

G. Premaratne and A.K. Bera. Modeling asymmetry and excess kurtosis in stock return data. Illinois Research & Reference Working Paper No. 00-123, 2000.

M. Rockinger and E. Jondeau. Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, 106(1):119-142, 2002.

A. Tolga Ergun and Jongbyung Jun. Time-varying higher-order conditional moments and forecasting intraday var and expected shortfall. *The Quarterly Review of Economics and Finance*, 50(3):264-272, 2010.A.

A. Wilhelmsson. Value at Risk with time varying variance, skewness and kurtosis-the NIG-ACD model. *Econometrics Journal*, 12(1):82-104, 2009.

Books

Brooks, C., 2002, *Introductory Econometrics for Finance*, The Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

Jorion, P., 2007, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed., The McGraw-Hill Companies, Inc., Two Penn Plaza, New York.

Liero H., and Zwanzig S., 2012, *Introduction to the theory of Statistical Inference.* Boca Raton: Taylor & Francis Group. Tsay, R. S., 2010, *Analysis of Financial Time Series*, 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey.

Tsay, R. S., 2013, *An Introduction to Analysis of Financial Data With R*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Electronic resources

Ghalanos, A., 2014, *racd: Autoregressive Conditional Density Models*, R package version 1.0-5., URL: `https://bitbucket.org/alexiosg/` (accessed 2016-05-15).

Longerstaey, J. and Spencer, M., 1996, *RiskMetrics—Technical Document*, Morgan Guaranty Trust Company and Reuters Ltd, New York, URL: `http://pascal.iseg.utl.pt/~aafonso/eif/rm/TD4ePt_2.pdf` (accessed 2016-05-02).

10 Acknowledgement

I would like to thank my supervisor Per Johansson for his support and insightful comments throughout the writing of this thesis.

11 Appendix

11.1 Fit Details

```
*---------------------------------*
*          ACD Model Spec         *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,1)
Distribution : nig
```

	Estimate	Std. Error	t value	Pr(>|t|)
mu	6.708950e-04	2.859701e-04	2.346032e+00	1.897447e-02
ar1	5.472024e-01	1.457535e-01	3.754300e+00	1.738265e-04
ma1	-6.374474e-01	1.315002e-01	-4.847503e+00	1.250251e-06
alpha1	4.503680e-02	3.873637e-04	1.162649e+02	0.000000e+00
beta1	9.512353e-01	3.824085e-04	2.487485e+03	0.000000e+00
skcons	-1.996289e-01	3.176064e-02	-6.285416e+00	3.269767e-10
skalpha1	1.019303e-01	1.207031e-01	8.444710e-01	3.984063e-01
skgamma1	-2.990029e-02	4.633357e-02	-6.453266e-01	5.187156e-01
skbeta1	6.071943e-05	6.110626e-01	9.936695e-05	9.999207e-01
shcons	8.634448e-01	5.261064e-01	1.641198e+00	1.007563e-01
shalpha1	7.266380e-12	4.502546e-08	1.613838e-04	9.998712e-01
shgamma1	5.300148e-01	2.654258e-01	1.996847e+00	4.584181e-02
shbeta1	6.185813e-01	1.794868e-01	3.446390e+00	5.681304e-04
omega	7.123559e-07	NA	NA	NA

```
                   ACD     GARCH
Akaike       -5.934133 -5.940746
Bayes        -5.870332 -5.911300
Shibata      -5.934466 -5.940818
Hannan-Quinn -5.909884 -5.929554

Box-Ljung test
data:  st.resid.acd
X-squared = 0.41766, df = 1, p-value = 0.5181
data:  st.resid.acd
X-squared = 4.7078, df = 7, p-value = 0.6956
data:  st.resid.acd^2
X-squared = 0.00059456, df = 1, p-value = 0.9805
data:  st.resid.acd^2
X-squared = 1.17, df = 7, p-value = 0.9916


*---------------------------------*
*       GARCH Model Spec          *
*---------------------------------*

Conditional Variance Dynamics
------------------------------------
GARCH Model : sGARCH(1,1)

Conditional Mean Dynamics
------------------------------------
Mean Model : ARFIMA(1,0,1)
```

```
Conditional Distribution
------------------------------------
Distribution :  nig
Includes Skew :  TRUE
Includes Shape :  TRUE

            Estimate   Std. Error     t value     Pr(>|t|)
mu      5.831262e-04 3.125506e-04   1.8657014 6.208316e-02
ar1     5.143659e-01 1.509544e-01   3.4074255 6.557882e-04
ma1    -6.107608e-01 1.397969e-01  -4.3689153 1.248652e-05
omega   9.015186e-07 1.437328e-06   0.6272185 5.305160e-01
alpha1  4.377102e-02 8.720536e-03   5.0193035 5.185916e-07
beta1   9.552288e-01 9.154217e-03 104.3485031 0.000000e+00
skew   -1.133888e-01 6.426918e-02  -1.7642796 7.768492e-02
shape   7.741862e-01 1.555518e-01   4.9770328 6.456637e-07

Box-Ljung test
data:  st.resid.garch
X-squared = 0.70884, df = 1, p-value = 0.3998
data:  st.resid.garch
X-squared = 5.0971, df = 7, p-value = 0.6481
data:  st.resid.garch^2
X-squared = 0.002809, df = 1, p-value = 0.9577

data:  st.resid.garch^2
X-squared = 1.1737, df = 7, p-value = 0.9915
```

11.2 Roll details

```
*-------------------------------------*
*              GARCH Roll             *
*-------------------------------------*
No.Refits : 16
Refit Horizon : 200
No.Forecasts : 3066
GARCH Model : sGARCH(1,1)
Distribution : nig

Forecast Density:
                Mu  Sigma    Skew  Shape Shape(GIG) Realized
2004-03-15  0.0024 0.0111 -0.0963 1.3147          0  -0.0199
2004-03-16  0.0035 0.0118 -0.0963 1.3147          0   0.0135
2004-03-17  0.0013 0.0117 -0.0963 1.3147          0   0.0140
2004-03-18 -0.0002 0.0118 -0.0963 1.3147          0   0.0021
2004-03-19  0.0000 0.0116 -0.0963 1.3147          0  -0.0105
2004-03-22  0.0013 0.0115 -0.0963 1.3147          0  -0.0071
..........................
               Mu  Sigma    Skew  Shape Shape(GIG) Realized
2016-05-10 0.0019 0.0126 -0.2159 2.5703          0   0.0183
2016-05-11 0.0009 0.0129 -0.2159 2.5703          0   0.0041
2016-05-12 0.0007 0.0125 -0.2159 2.5703          0   0.0029
2016-05-13 0.0006 0.0122 -0.2159 2.5703          0  -0.0111
2016-05-16 0.0010 0.0122 -0.2159 2.5703          0   0.0148
2016-05-17 0.0004 0.0123 -0.2159 2.5703          0   0.0020

Elapsed: 57.57761 mins
```

```
*------------------------------------*
*              ACD Roll              *
*------------------------------------*
No.Refits : 16
Refit Horizon : 200
No.Forecasts : 3066
GARCH Model : sGARCH(1,1)
Mean Model : ARFIMA(1,0,1)
Distribution : nig

Forecast Density:
                Mu  Sigma    Skew  Shape Shape(GIG) Realized
2004-03-15  0.0024 0.0111 -0.1525 1.4142          0  -0.0199
2004-03-16  0.0035 0.0118 -0.1601 1.4140          0   0.0135
2004-03-17  0.0013 0.0117 -0.1310 1.0404          0   0.0140
2004-03-18 -0.0002 0.0118 -0.0915 0.8537          0   0.0021
2004-03-19  0.0000 0.0116 -0.0977 1.3829          0  -0.0105
2004-03-22  0.0013 0.0115 -0.1239 1.4138          0  -0.0071

..........................
               Mu  Sigma    Skew  Shape Shape(GIG) Realized
2016-05-10 0.0019 0.0126 -0.2696 2.6702          0   0.0183
2016-05-11 0.0009 0.0129 -0.1667 1.4408          0   0.0041
2016-05-12 0.0007 0.0125 -0.1882 2.6005          0   0.0029
2016-05-13 0.0006 0.0122 -0.2046 2.6409          0  -0.0111
2016-05-16 0.0010 0.0122 -0.2445 2.6701          0   0.0148
2016-05-17 0.0004 0.0123 -0.1682 1.6735          0   0.0020

Elapsed: 1.356587 hours
```

11.3 Parameter Uncertainty

In Figure 4 we display one density plot for each GARCH parameter for a total of eight density plots, in order to illustrate the parameter uncertainty.

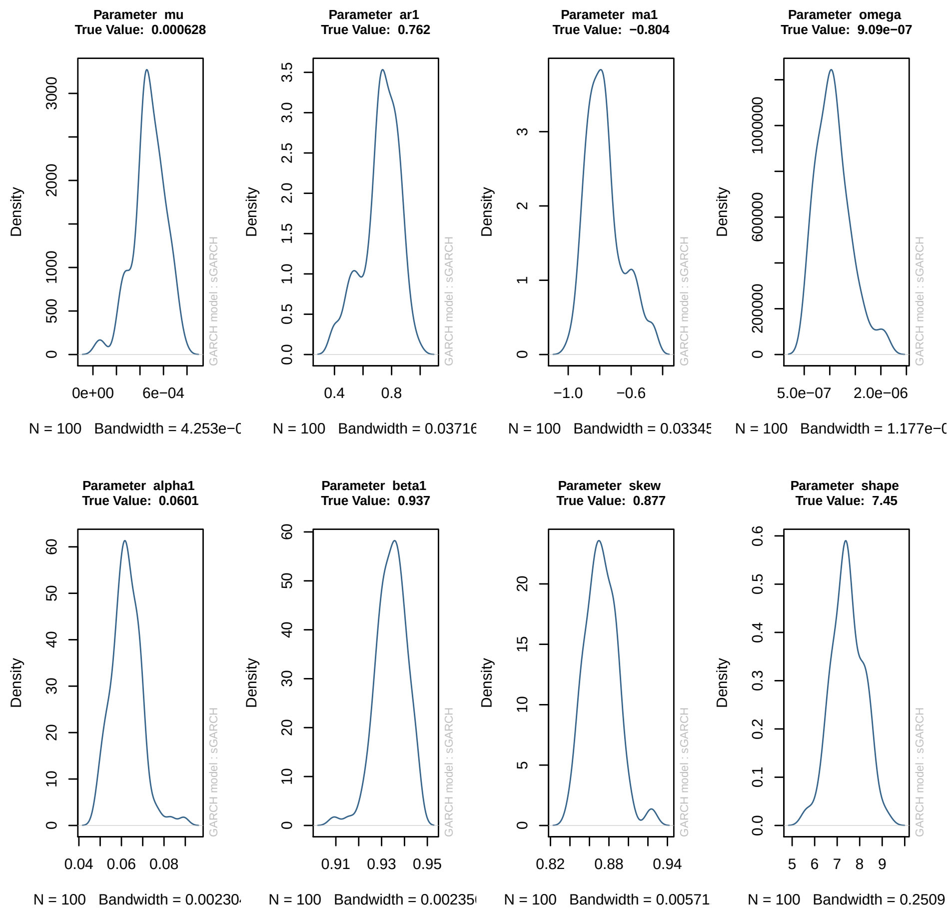

Figure 4: Bootstrap parameter density plots. Whereas we in Figure 3 we see the uncertainty of sigma, here 4 we see the parameter uncertainty. Each GARCH parameter have one density plot. Since n.bootfit=100 we have N=100.

11.4 Code

In the output above some redundant rows were deleted for increased readability. The statistical software and programming language R was used to produce the results in this thesis. The document preparation system LaTeXwas used to write this document. Both the R code and the document is available by googling *jacoblindberg github*.