LUNDS
UNIVERSITET

A simple model of volatility in financial data

An alternative to GARCH models

Alexandra Milton
Marcus Svensson

Master's Thesis in Statistics (15 ECTS)
STAN40
Department of Statistics
Lund University
Spring 2019

Supervisor: Krzysztof Podgórski

Acknowledgements

We would like to thank Krzysztof Podgórski for providing great insights and advice throughout our process. Krzysztof's supervision has been of great value during the planning and development of this thesis. We would also like to thank Maria Juhlin for her comments and suggestions.

Abstract

Financial return series are often characterized by volatility clusters and a leptokurtic distribution. Many models that account for these properties exist, with the GARCH model proposed by Bollerslev (1986) being the most popular. This thesis explores an alternative model to capture the stochastic volatility in financial time series. The considered model is denoted the autoregressive gamma variance Gaussian mixture model and was proposed by Johannesson et al. (2016b). The model consists of the product of two independent time series, namely an autoregressive Gaussian process and an autoregressive gamma process, where the gamma process modulates the variance of the Gaussian white noise.

This thesis proposes and evaluates an alternative method to estimate the correlation coefficient of the gamma process. The proposed method outperforms the original method when the true correlation coefficient is exceedingly large, which is the case for almost all financial return series. In addition, this thesis develops basic unbiased methods to interpolate and predict the gamma process given the observed daily financial return. These partial results require further research to fully develop more advanced prediction methods.

Keywords: Volatility, financial time series, autoregressive gamma process, generalized Laplace distribution, autoregressive gamma variance Gaussian mixture model

Contents

1 Introduction 4

2 Background 6

2.1 Fundamentals of financial data 6

2.2 Dependency in returns 8

2.3 Volatility clustering 9

2.4 Distribution of returns 11

2.5 The ARCH and GARCH models 13

3 Autoregressive gamma model for volatility 15

3.1 Autoregressive gamma process 15

3.2 Exploratory simulations of the autoregressive gamma process 16

4 Autoregressive gamma variance Gaussian mixture model 18

4.1 Model definition 18

4.2 Exploratory simulations of the ARGVG process 19

5 Model Fitting 21

5.1 Estimation of the mean and scale 21

5.2 Estimation of ν 21

5.3 Estimation of the correlation in the volatility process 23

5.3.1 Estimating ρ_r using a weighted geometric mean 23

5.3.2 Estimating ρ_r by fitting the exponential decay 24

5.3.3 Comparison between the two methods 25

6 Interpolation and prediction 26

6.1 The conditional distribution of $R|Y^2$ 26

6.2 Interpolation 30

6.3 Prediction 32

6.4 Closing remarks on the interpolation and prediction method 34

7 Conclusion 36

8 References 37

Appendix 1 39

1 Introduction

The efficient market hypothesis states that the price of a financial instrument fully reflects all available information. This implies that one cannot profit by purchasing or selling assets based on the available information (Malkiel, 1989). From a statistical standpoint, this phenomenon translates to an expected return of zero at any point in a return series, consisting solely of white noise.

While this implies that the conditional mean of a financial return series cannot be meaningfully modeled, modeling financial data remains a widely popular area. The focus lies not on modeling the conditional mean but modeling the measure of risk. Financial time series are known to be volatile and the risk associated with an asset depends on how volatile the asset is. The degree of variation over time is referred to as volatility and is measured using the standard deviation, or variance of the logarithmic returns of a financial time series. In financial data, clusters of less volatile periods and clusters of more volatile periods tend to be present. This was first shown by Mandelbrot (1963) and is referred to as volatility clustering. While it may not be possible to predict profitable trades, ways of predicting the future volatility of an asset exist and are useful tools for making financial decisions.

Other than volatility clustering, also known as conditional heteroscedasticity, financial return series are often characterized by a leptokurtic (heavy-tailed) distribution. A range of different models have been proposed to capture these properties. Well-known examples of these are the autoregressive conditional heteroscedasticity (ARCH) model by Engle (1982) and the generalized ARCH (GARCH) model, an extension of the former, proposed by Bollerslev (1986). The GARCH model is probably the most popular choice when modeling financial return series which, in short, predicts the volatility of the returns using past values of the conditional variance and returns.

An alternative model that could have potential to capture the characteristics often observed in financial return series is the autoregressive gamma variance Gaussian mixture model. The model has been used within the engineering field and was proposed by Johannesson et al. (2016b) to model road topography data. The structure of the road topography data closely resembles the typical volatility clustering observed in financial data. The model consists of the product of two independent time series, namely an autoregressive Gaussian process and an autoregressive gamma process. This product follows a generalized Laplace marginal distribution regardless of autoregressive order in its components. As previously mentioned,

under the assumption of the efficient market hypothesis, the mean process of financial returns should consist solely of white noise. If the Gaussian component is white noise, only the gamma component needs to be fitted, which would further simplify the structure of the model. It will be shown that the gamma component allows for the model to both express excess kurtosis and varying variance, which, as previously stated, are two key properties of financial return series.

One distinct difference between the engineering and financial fields is the degree of importance in interpolating and making future predictions. Johannesson et al. (2016b) focus on estimating parameters to simulate data that mimics the behavior and structure of the original data. No methods of interpolation and predictions are therefore developed in the article. However, in finance interpolation and prediction are of high interest for making financial decisions. Therefore, methods to interpolate and predict volatility would need to be developed for the model to be useful in a financial context.

The purpose of this thesis is to explore the applicability of the autoregressive gamma variance Gaussian mixture model on financial return series. Furthermore, this thesis aims to develop some basic methods for interpolation and prediction. Moreover, because volatility clustering indicates a very slow change in the variance, an alternative method of estimating the correlation coefficient ρ_r that may perform better under the condition that ρ_r is very large is explored. Throughout the thesis two financial stock indices, the OMXS30 and the NASDAQ Composite, are utilized to exemplify various properties of financial data. All computations and illustrations are performed using R.

2 Background

In this section a common transformation which allows for financial time series data to be stationary is proposed. Moreover, the key characteristics and properties of financial time series data are discussed to give an understanding of how financial time series tend to differ from the most basic Gaussian processes. Finally, the ARCH and GARCH models, which account for these characteristics, are briefly introduced.

2.1 Fundamentals of financial data

Two well-known stock indices, the OMXS30 and the NASDAQ Composite, will be used to exemplify the key properties of financial time series. The OMXS30 is an index of the 30 most traded stocks on the Stockholm Stock Exchange and the NASDAQ Composite is one of the largest indices in the US and weighs heavily towards IT companies (Nasdaq Nordic, 2019). Both time series are illustrated in Figure 2.1 and consist of the daily closing price of each index between 2007-01-02 and 2019-05-31.

Figure 2.1 The daily closing price of the OMXS30 (A) and NASDAQ Composite (B). Both series exhibit a clear upward trend over time.

One of the key assumptions in time series modeling is that the process is stationary. A stochastic process Z_t is said to be weakly stationary if the mean of Z_t and the covariance between Z_t and Z_{t-k} do not depend on time t. The mean should be constant, and the covariance should only depend on the time difference k (Tsay, 2010, p.30).

In almost all cases, raw financial index data does not meet the assumption of stationarity. This should be obvious, since if the mean was constant, it would suggest that the price (value) of an index does not change over long periods of time. Figure 2.1 displays clear upward trends in both the OMXS30 and the NASDAQ Composite, which suggests that neither time series has constant mean. A common way to combat this is to transform the data into the daily log-returns, Y_t, through

$$Y_t = \log\left(\frac{p_t}{p_{t-1}}\right) = \log(p_t) - \log(p_{t-1}) \tag{2.1}$$

where p_t is the daily closing price on day t (Cryer & Chan, 2008, p.278).

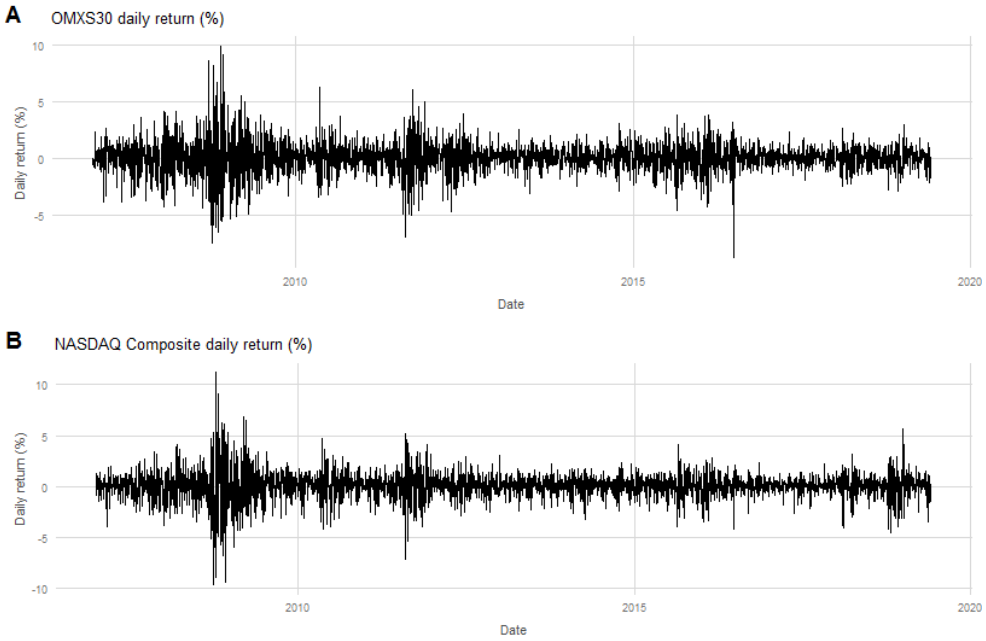

Figure 2.2 The daily return (%) of the OMXS30 (A) and NASDAQ Composite (B). These transformed series suggest that the data has been successfully detrended, with the mean now being constant in both cases.

Moreover, the values in Y_t are multiplied by 100, which allow its values to be interpreted as the percentage increase or decrease of the closing price of a given day. The transformed data is depicted in Figure 2.2 and clearly suggests that the data has been successfully detrended: The mean is constant around zero throughout the entire series. The behavior of the conditional variance will be discussed in section 2.3. Throughout the rest of the thesis, this transformed data will be referred to as returns.

2.2 Dependency in returns

Empirical studies have shown that no or very little dependency exists in daily return series (Fama, 1970; Pagan, 1996). The return series is then said to consist of white noise, serially uncorrelated data with finite mean and variance. If dependency in returns were to exist, one could make profitable predictions on future returns. Therefore, in an efficient market, the conditional mean should be zero (Cryer & Chan, 2008, p.277). This behavior is often described by the efficient market hypothesis, which states that the price of a financial instrument fully reflects all available information, with the implication that one cannot profit by purchasing or selling assets based on the available information (Malkiel, 1989).

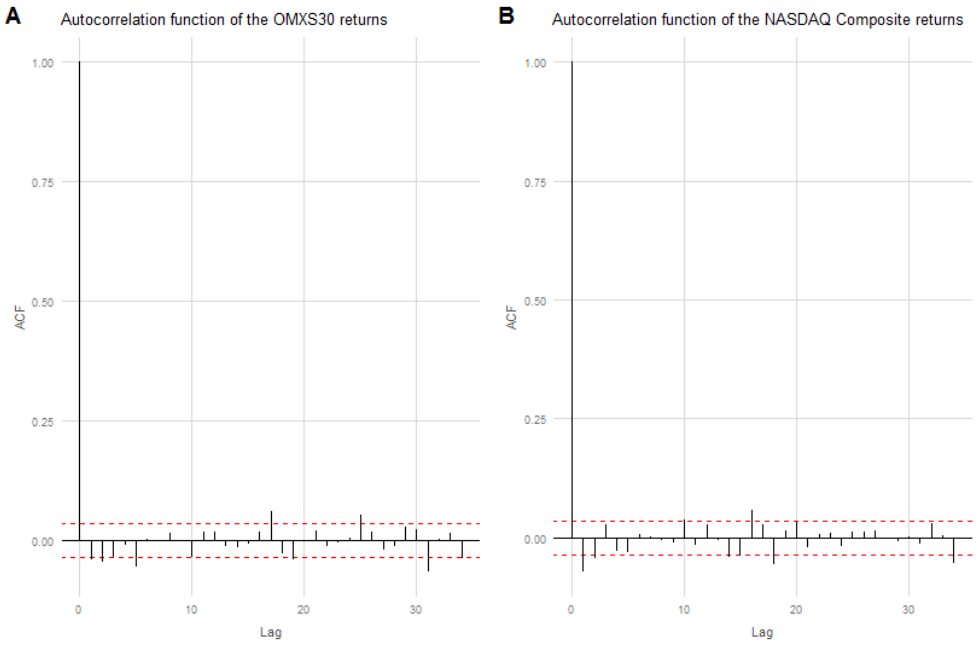

Figure 2.3 Autocorrelation functions of the OMXS30 (A) and NASDAQ Composite (B). In both cases there are no signs of autocorrelation, which is in agreement with the efficient market hypothesis of financial returns being white noise.

The dependency in a time series is commonly measured by its autocorrelation. The sample autocorrelation function measures the correlation between a time series and a delayed version of itself at various lags k and is defined as

$$r_Y(k) = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}. \quad (2.2)$$

The autocorrelation functions of the OMXS30 returns and the NASDAQ Composite returns are portrayed in Figure 2.3. In both cases there are no signs of autocorrelation. Throughout the rest of the thesis, the assumption is made that the daily returns of both the OMXS30 and the NASDAQ Composite are white noise.

2.3 Volatility clustering

Financial time series are characterized as heteroscedastic time series because the conditional variance in not constant over time. As described by Mandelbrot (1963), in financial time series, the conditional variance appears in clusters. He stated that "large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes" (Mandelbrot, 1963). This slow process of varying conditional variance is often referred to as volatility clustering due to the clusters of both high and low conditional variance that naturally appear. What this means is that an unstable period with violent price changes tends to be followed by a larger conditional variance compared to a stable period which tends to be followed by a smaller conditional variance (Cryer & Chan, 2008, p.277).

Naturally, this common characteristic in financial return series is crucial for any model to capture. Simple models that do not account for heteroscedasticity cannot capture the dependency that is present in the conditional variance. In Figure 2.4 the returns of the OMXS30 and the NASDAQ Composite are illustrated again, this time with the volatility clusters roughly outlined. Plots of the squared returns are also included, which is a popular (albeit primitive) proxy of the volatility clusters – a period with high conditional variance will be represented by a higher ceiling of the squared values. Moreover, the clear dependency in the conditional variance can be further highlighted by calculating the autocorrelation of the squared returns. The autocorrelation functions of the squared returns are depicted in Figure 2.5, which reveals a very long dependence that concurs with the observed volatility clusters and with Mandelbrot's description above.

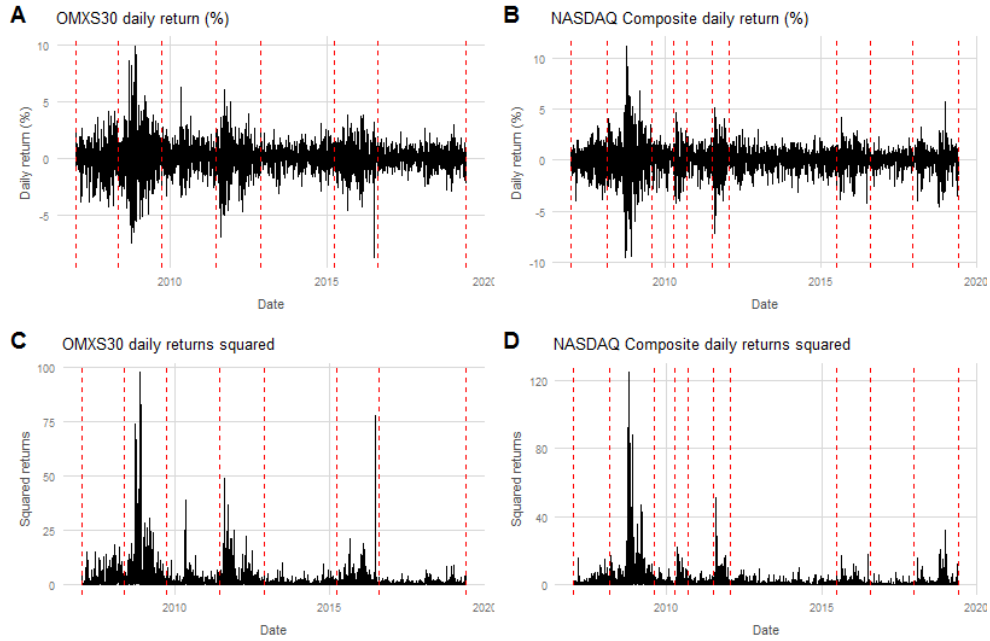

Figure 2.4 These figures highlight the volatility clustering that is prevalent in financial time series. Figure A illustrates rough outlines of the volatility clusters in the OMXS30 returns. Figure B depicts the same process for the NASDAQ Composite. Figures C and D illustrate the same clusters on the squared return series, which can be seen as a primitive volatility proxy. Both cases display typical behavior in financial time series where the clusters last for several months at a time.

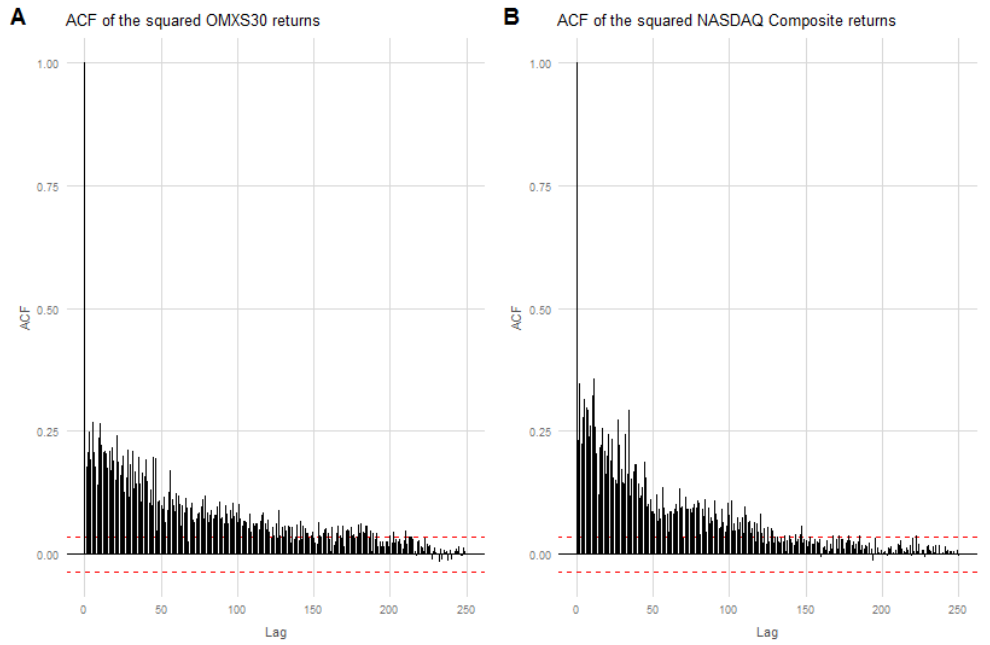

Figure 2.5 Autocorrelation functions of the squared returns of the OMXS30 (A) and NASDAQ Composite (B). Both figures exhibit long term dependency, which concurs with the volatility clusters of the returns that span several months.

2.4 Distribution of returns

Financial returns are often observed to follow distributions with leptokurtic properties. A distribution is said to be leptokurtic when its tails are heavier than a normal distribution with equivalent mean and standard deviation. A concept closely related to this is the kurtosis of a distribution. The kurtosis is a measurement of the heaviness of a distribution's tails and is calculated through

$$\kappa = \frac{m_4}{m_2^2} \tag{2.3}$$

where m_4 is the fourth moment about the mean and m_2 is the second moment about the mean. Commonly, the kurtosis is compared to the kurtosis of the normal distribution, which is exactly 3. Thus, a distribution is considered leptokurtic when its kurtosis is greater than 3 (often referred to as excess kurtosis). The leptokurtic features in distributions of financial returns was early described by Mandelbrot (1963) and Fama (1965) and suggestions for suitable distributions have been explored ever since.

In section 2.3 it was discussed that any successful model must be able to account for the conditional heteroscedasticity that is commonly present in financial return series. The same applies for the distributional properties of the returns: It is crucial that any model accounts for the leptokurtic distribution. One way to accomplish this is by assigning random variance to a normal distribution, which increases the kurtosis. This can be shown using a simple example. Consider two normal distributions $f_{x_1}(x_1)$ and $f_{x_2}(x_2)$ with zero mean and standard deviations $\sigma_1 = 1$ and $\sigma_2 = 3$ respectively. The densities of these distributions are portrayed in Figure 2.6A. Furthermore, define $\delta \sim Bernoulli(p)$, where p is any probability between 0 and 1, and the outcome $\delta \in \{0,1\}$. The density of

$$Y = \delta X_1 + (1 - \delta)X_2 \tag{2.4}$$

is illustrated in Figure 2.6B and clearly exhibits excess kurtosis. Meanwhile, the underlying function can be simplified as a standard normal random variable X multiplied by the random standard deviation $\sqrt{R}$, defined as:

$$R = \begin{cases} \sigma_1^2, & with\ probability\ p \\ \sigma_2^2, & with\ probability\ 1-p \end{cases} \tag{2.5}$$

Furthermore, this can be extended to non-trivial cases where R is a random variable that follows any given non-negative distribution and the distribution of the product $\sqrt{R}X$ follows a leptokurtic distribution.

To summarize, one way to model data that follows a leptokurtic distribution is by allowing for random variance. One such method is to define the model as the product of a normal distributed variable and a random variable of a non-negative distribution, e.g. the gamma distribution.

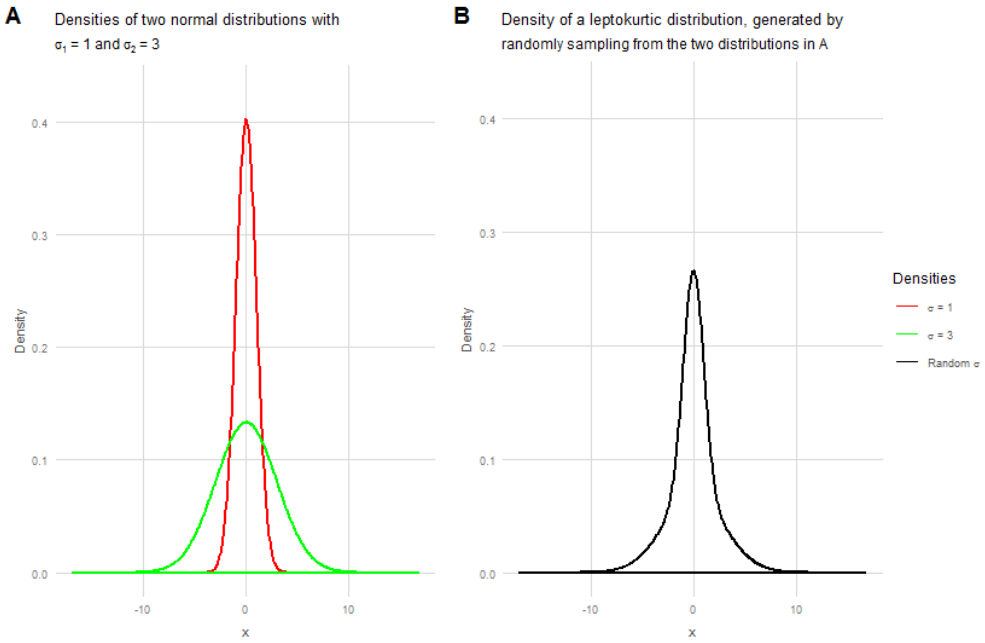

Figure 2.6 Density plots of two normal distribution with different standard deviations (A) and the density of the leptokurtic distribution generated by randomly sampling from the two normal distributions

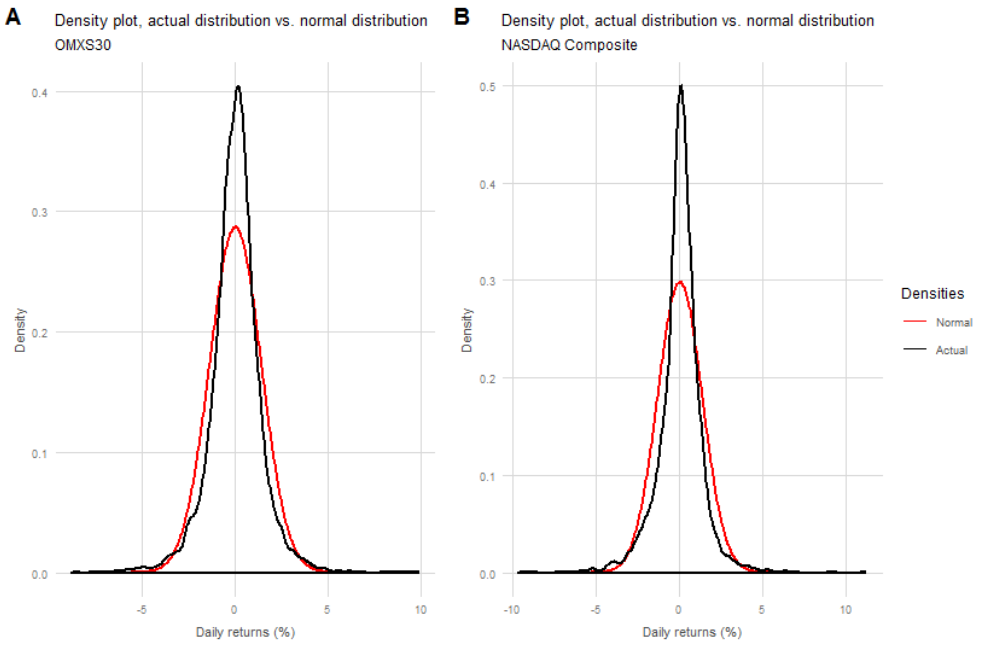

Figure 2.7 Density plots of the daily returns of the OMXS30 (A) and NASDAQ Composite (B). In both cases the distribution of the observed density (black line) is compared to a normal distribution with equivalent mean and variance (red line). In both cases there are clear signs of excess kurtosis with the peak and tails having more density than the normal distributed equivalent.

Density plots of the two return series along with a normal distribution of equivalent mean and standard deviation are depicted in Figure 2.7. In addition, this can be confirmed by directly calculating the sample kurtosis. The sample kurtosis of the OMXS30 returns is 8.06, and the sample kurtosis of the NASDAQ Composite is 10.15. Thus, both distributions show clear signs of being leptokurtic, both through the computed kurtosis and through visual means in Figure 2.7.

2.5 The ARCH and GARCH models

Different models have been proposed to account for some of the characteristics of financial return series. As discussed, the financial return series in general consist of a serially uncorrelated sequence with zero mean but with variance changing over time, observed as volatility clustering. This implies that the conditional variance, or more commonly known as the conditional volatility, is not constant. The autoregressive conditional heteroscedasticity (ARCH) model was introduced by Engle (1982) to model this conditional volatility. Let ϵ_t denote the error terms with respect to a mean process. The general form of the ARCH(q) model can then be expressed as follows:

$$\epsilon_t = \sigma_t \varepsilon_t \tag{2.6}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{2.7}$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. Here ε_t are independent and identically distributed standard normal random variables independent of σ_t. The ARCH parameter q is the number of lagged error terms to be included in the model.

An extension of the ARCH model was proposed by Bollerslev (1986) as the generalized ARCH (GARCH) model. This model includes p lags of the conditional variance. Like in the ARCH model, ϵ_t denotes the error terms with respect to a mean process. The GARCH(p, q) model is defined as:

$$\epsilon_t = \sigma_t \varepsilon_t \tag{2.8}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{2.9}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and ε_t are once again independent and identically distributed standard normal random variables independent of σ_t. The constraint $\sum_{i=1}^{\max(p,q)} (\beta_i + \alpha_i) < 1$, where $\beta_j = 0$ for $j > p$ and $\alpha_i = 0$ for $i > q$, is necessary for the GARCH model to be weakly stationary. For details on model order selection and parameter estimation, see for example Cryer and Chan (2008) and Tsay (2010). Additional models have been developed to further capture the volatility and other characteristics that might appear in heteroscedastic time series data, see for example Tsay (2010).

In most cases, only the simplest ARCH(1) or GARCH(1,1) model is necessary when modeling financial return series, which may suggest that the simple and slow-changing volatility structure can be modeled in an even simpler fashion. Moreover, a disadvantage of the ARCH and GARCH model is that there is not an explicit marginal distribution.

3 Autoregressive gamma model for volatility

In this section the autoregressive gamma model is defined. Furthermore, simulations are utilized to explore which general range of parameter values may be suitable in order to generate time series that resemble the frequently observed volatility clustering in financial time series.

3.1 Autoregressive gamma process

It was argued in section 2.3 that any model suitable for financial data must be able to account for conditional heteroscedasticity. Furthermore, it was argued in section 2.4 that by modulating a normal distribution with random variance through $\sqrt{R}X$, the distributional features of financial return series can be captured. The gamma distribution is a non-negative distribution that is a suitable candidate for R. More specifically, an autoregressive gamma process R_t is a reasonable option to modulate the conditional variance of X in a manner that resembles the autoregressive and slow-changing nature of the volatility.

A random gamma variable R standardized to $E[R] = 1$ follows the distribution

$$f(r) = \frac{1}{\Gamma\left(\frac{1}{\nu}\right)\nu^{1/\nu}} r^{(1/\nu)-1} e^{-r/\nu} \tag{3.1}$$

where $\nu > 0$. The stationary autoregressive gamma process R_t was presented by Sim (1990). With a given ν and ρ_r the process is defined as

$$R_t = \rho_r K_t R_{t-1} + (1 - \rho_r)\epsilon_t. \tag{3.2}$$

The process has an exponentially decaying autocorrelation function

$$Corr(R_t, R_{t-k}) = \rho_r^k$$

for lag $k = 1,2,3, ...$, which has been shown by Johannesson et al. (2016a). The random innovations ϵ_t are independent gamma distributed variables with $E[\epsilon_t] = 1$ and $V[\epsilon_t] = \nu$. The random innovations are also independent from K_t and R_{t-1}. The random factors K_t are dependent on R_{t-1} and defined as

$$K_t = \frac{1}{m_t} \sum_{i=0}^{N_t} E_i \tag{3.3}$$

where $E_0 = 0$ and for $i > 0$ $E_i \sim Exp(1)$, and thus $E[E_i] = 1$. Furthermore, $N_t \sim Po(m_t)$, where m_t is given by

$$m_t = \frac{\rho_r / \nu}{1 - \rho_r} R_{t-1}. \tag{3.4}$$

3.2 Exploratory simulations of the autoregressive gamma process

With the autoregressive gamma process defined, a brief demonstration of how the behavior of R_t is affected by changes in ν and ρ_r is provided using simulated data. To maintain the focus on the model's possible applicability in specifically financial time series, the examples will only contain extremely large values of ρ_r. Recall from section 2.3 that the volatility clusters in financial time series are usually several months long, which translates into a very slowly changing volatility process R_t. Because the parameter ρ_r represents the *daily* correlation, the correlation in the process on monthly distance (assuming an average of 21 trading days in a month) is then

$$Corr(R_t, R_{t-21}) = \rho_r^{21} \tag{3.5}$$

Two sets of values of both parameters are examined. The correlation coefficient ρ_r is set to 0.99 and 0.95 (which gives correlations between months of order $0.99^{21} \approx 0.81$ and $0.95^{21} \approx 0.34$). The shape parameter ν is set to 0.5 and 2. This results in a total of four combinations of parameter values. The four simulated series are depicted in Figure 3.1.

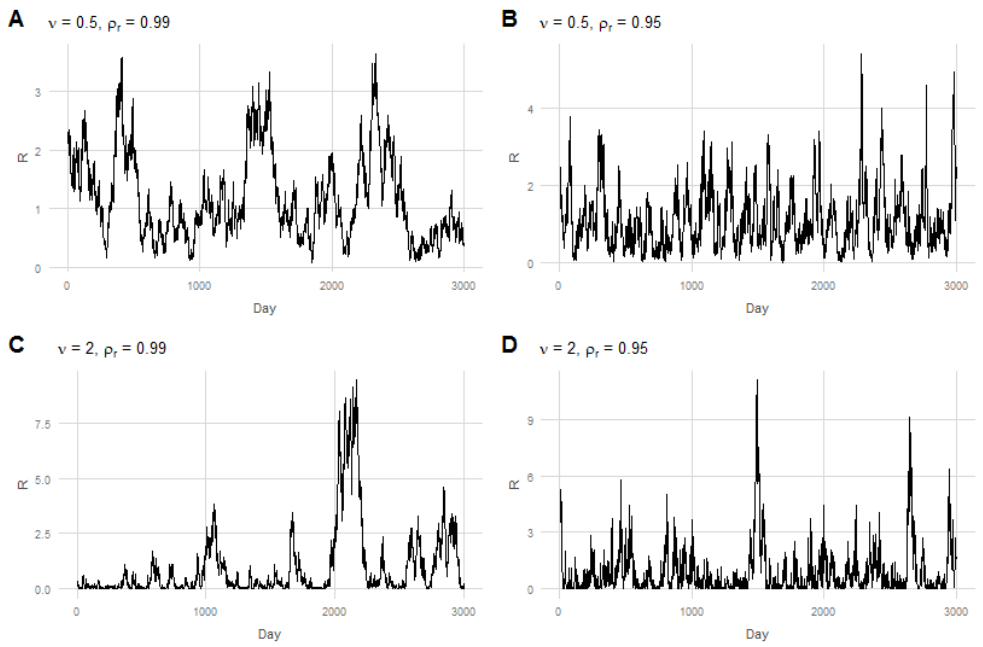

Figure 3.1 The four simulated autoregressive gamma series. The figures show how an increase in ρ_r leads to a slower change in R_t, which is desirable in order to mimic the behavior of the volatility clusters in financial return series. An increase in ν yields more defined clusters of large and small values.

It is immediately clear why the values of ρ_r were chosen as large as 0.95 and 0.99. It is evident that already when $\rho_r = 0.95$, ρ_r is too small in the context of financial time series. R_t is varying too frequently to mimic the behavior of volatility clusters observed in the OMXS30 and NASDAQ Composite examples. Moreover, when $\nu = 0.5$ the volatility clusters mimicked by R_t are not as clearly defined as in the example data. However, when both ν and ρ_r are large, the simulated time series has a general structure that is very reminiscent of the outlines of the squared return plots of the OMXS30 and NASDAQ Composite that were illustrated in Figure 2.4C-D.

4 Autoregressive gamma variance Gaussian mixture model

This section proposes an alternative approach to model conditional heteroscedasticity and excess kurtosis. A brief introduction of the model, its composition and parameters are provided. Furthermore, the simulations of R_t are extended to simulate Y_t, which confirms the conclusions that were drawn in section 3.2 about which parameter values may be reasonable in the context of financial return series.

4.1 Model definition

The model is termed the autoregressive gamma variance Gaussian mixture model, in short ARGVG. The model has its roots in the engineering field where it was originally proposed by Johannesson et al. (2016b) to model road topography data. The model is defined as

$$Y_t = \sigma\sqrt{R_t}X_t + \mu \tag{4.1}$$

where R_t and X_t are two independent time series. The gamma distributed R_t was introduced in section 3 and allows for Y_t to vary in its conditional variance and express excess kurtosis, while X_t is a normally distributed ARMA process of arbitrary order. Thus, R_t can be considered a process that models the volatility of the time series Y_t. Unlike in the ARCH-model and its extensions, which have untraceable distributional structure, the ARGVG-model has an explicit marginal distribution that follows a generalized Laplace distribution (Johannesson et al., 2016b).

R_t and X_t are generally standardized, which allows the mean and scale of the process to be isolated in the outer parameters μ and σ. X_t is standardized to $N(0,1)$ in traditional fashion, while the non-negative R_t is standardized to $E[R_t] = 1$.

Because the focus of this thesis lies on financial time series, the theory presented is restricted to the case when X_t consists of white noise (See section 2.2). For a more general case, where X_t cannot be assumed to be white noise, see Johannesson et al.

(2016b) where X_t is modeled as an AR(1) process. For general theory on ARMA processes see for example Cryer and Chan (2008).

4.2 Exploratory simulations of the ARGVG process

The behavior of R_t was explored through simulations in section 3.2. With the full model defined, these simulated series can be extended to explore the behavior and structure of the product $Y_t = \sqrt{R}_t X_t$, which models a return series, and not just R_t which models the volatility of a return series. Note that changes in μ and σ result in obvious behavior in Y_t and will not be examined. The parameters μ and σ are set to 0 and 1 respectively in all simulations.

In Figure 4.1, the four sets of Y_t are presented. In Figure 4.2, Y_t^2 is visualized to highlight the volatility clusters and in Figure 4.3, histograms of Y_t are presented.

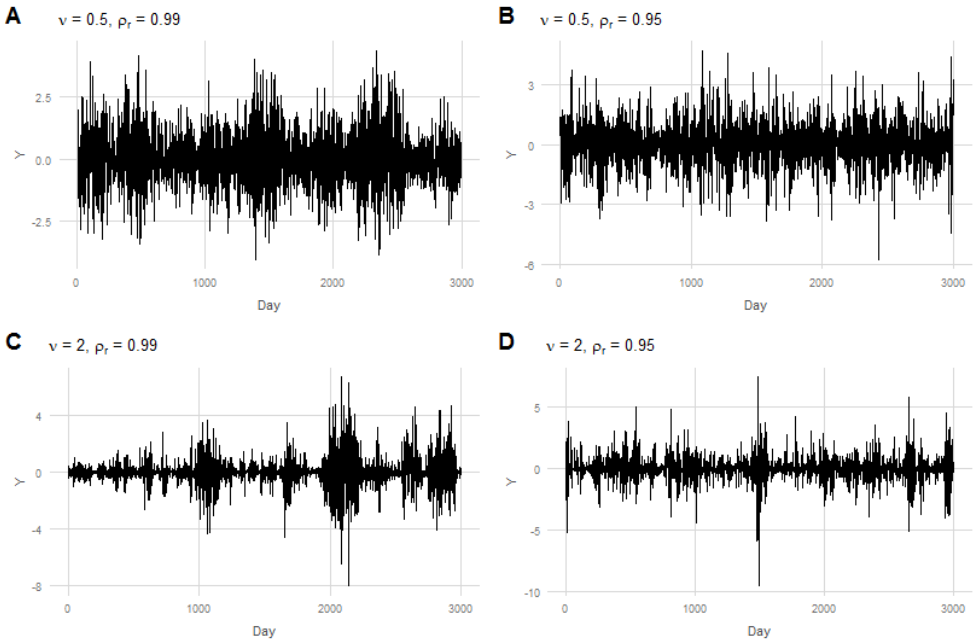

Figure 4.1 Four simulated ARGVG-time series with different sets of parameters. The figures suggest that an increase in ρ_r leads to a slower change in volatility. An increase in ν yields more defined volatility clusters.

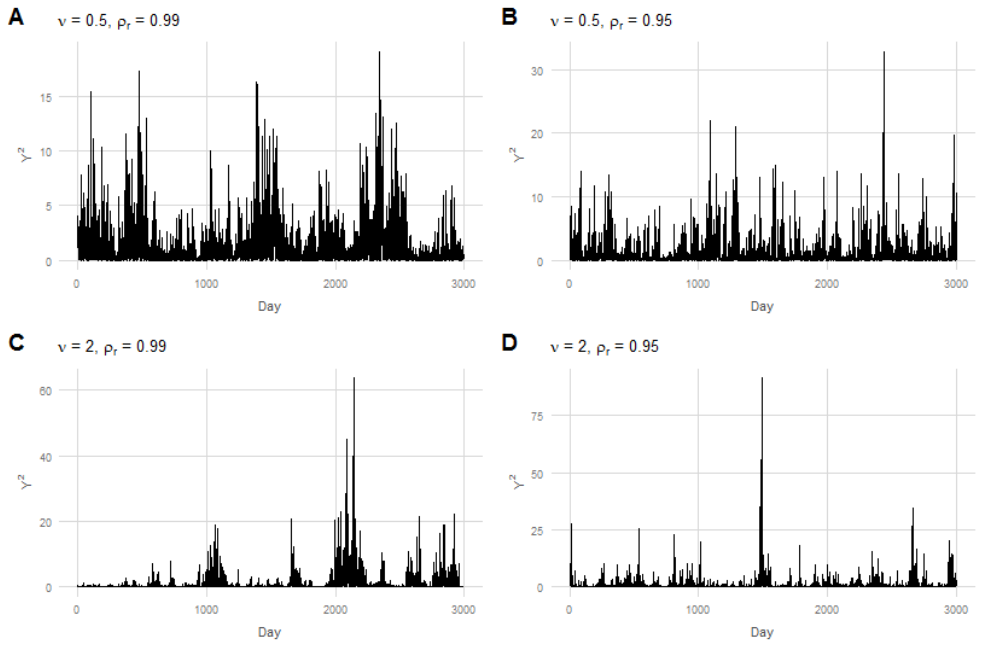

Figure 4.2 An illustration of the squared simulated series. Figure C has remarkably similar structure to the squared return series in Figure 2.4C-D.

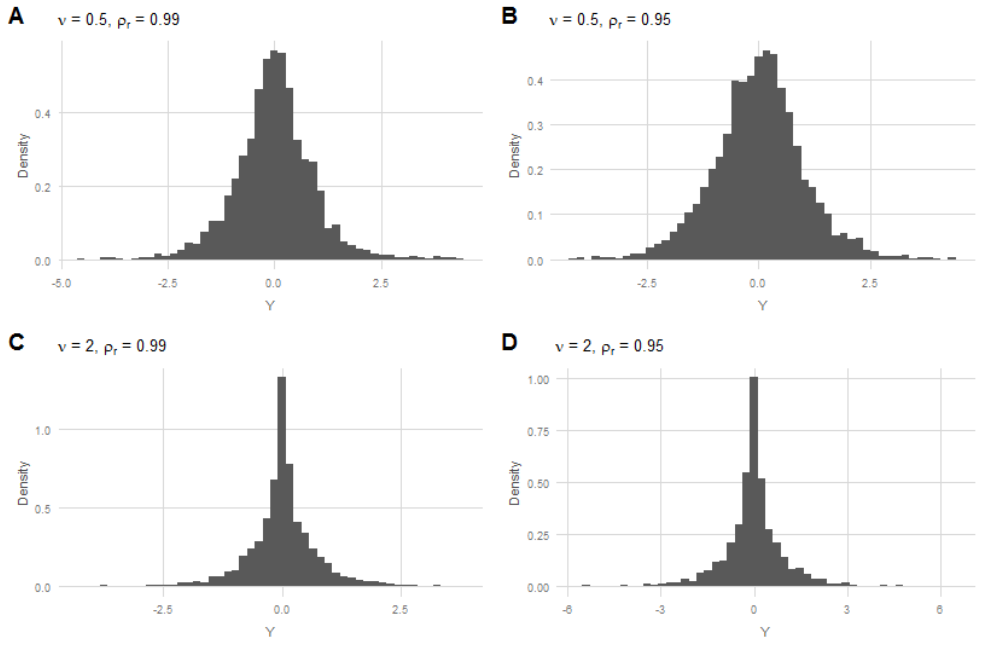

Figure 4.3 Histograms of the four simulated series. The main take away from these figures is that the distribution of the series with a large value in ν provides a closer resemblance to the distributions of the observed data with its high peak and fat tails, previously illustrated in Figure 2.7.

A few conclusions about the behavior and features of Y_t can be drawn, unsurprisingly much of this is in agreement with the observations made in section 3.2. An increase in ρ_r results in a more slowly changing volatility. The leptokurtic properties in Y_t increases as ν increases, which from a time series-perspective (with a large ρ_r) allows for the clusters of high and low conditional variance to be more clearly defined, which is noteworthy as this is usually observed in financial data. The histograms further highlight this. The histograms where $\nu = 2$ display clear excess kurtosis and bear more resemblance to the densities of the OMXS30 and NASDAQ Composite in Figure 2.7 than the histograms where $\nu = 0.5$.

5 Model Fitting

To fit the ARGVG model, a total of four parameters must be estimated. Methods of how to estimate these are presented in the following section. Furthermore, an alternative method of estimating ρ_r is proposed along with simulations to investigate its performance. In addition, the estimation process is applied to the previously introduced OMXS30 and NASDAQ Composite time series.

5.1 Estimation of the mean and scale

Estimating μ and σ in the ARGVG model can be easily accomplished by calculating the sample mean and the sample standard deviation of Y_t and will not be discussed further. The estimated mean and scale of the OMXS30 and the NASDAQ Composite time series are reported in Table 5.1.

5.2 Estimation of ν

Johannesson et al. (2016b) propose using the sample kurtosis κ of Y_t to estimate ν through

$$\nu = \frac{\kappa - 3}{3} \tag{5.1}$$

where κ can be estimated using the method of moments described in section 2.4. Table 5.2 presents the result from a set of simulations estimating ν using this method.

Each set of simulations consist of 10000 iterations of a time series with 3000 observations and varying combinations of ρ_r and ν. The results suggest that the method of moments may not be the optimal method of estimating the ν parameter when ρ_r is very large and the length of the time series is not unreasonably long. Other methods such as Maximum Likelihood estimation may prove superior, however no investigation into this matter will be covered in this thesis. Furthermore, the estimates of ν using the method of moments for the OMXS30 and NASDAQ Composite are presented in Table 5.3.

Table 5.1 The sample mean and standard deviation of the OMXS30 and NASDAQ Composite.

	$\hat{\mu}$	$\hat{\sigma}$
OMXS30	0.008	1.390
NASDAQ Composite	0.036	1.337

Table 5.2 Estimation of ν based on a set of simulations, each consisting of 10000 iterations of a time series with 3000 observations. The results suggest that the method of moments may not be the optimal method of estimating the ν parameter when ρ_r is very large. The method performs well when $\rho_r \leq 0.95$.

	$\nu = 1,$ $\rho_r = 0.90$	$\nu = 1,$ $\rho_r = 0.95$	$\nu = 1,$ $\rho_r = 0.99$	$\nu = 2,$ $\rho_r = 0.90$	$\nu = 2,$ $\rho_r = 0.95$	$\nu = 2,$ $\rho_r = 0.99$
Mean $\hat{\nu}$	0.978	0.966	0.882	1.938	1.909	1.701
(sd) $\hat{\nu}$	(0.22)	(0.24)	(0.30)	(0.48)	(0.52)	(0.61)

Table 5.3 The estimates of ν for the OMXS30 and NASDAQ Composite return series.

	$\hat{\nu}$
OMXS30	1.69
NASDAQ Composite	2.38

5.3 Estimation of the correlation in the volatility process

The correlation coefficient ρ_r's impact on R_t has previously been reviewed in section 3.2. It was concluded that very large values of ρ_r are required to generate a process that resembles the volatility clusters in financial time series. In addition, section 3.2 also provides arguments as to why the performance of any method of estimating ρ_r will be highly important. As seen in Figure 3.1, the structure of R_t is drastically altered when ρ_r decreases by just a few hundredths. It should therefore be of high priority to identify the best possible method of estimating ρ_r, even if the accuracy appears to be just marginally more precise.

5.3.1 Estimating ρ_r using a weighted geometric mean

It has been shown by Johannesson et al. (2016b) that in the case of the Gaussian component being white noise, the autocorrelation function of Y^2 can be used to obtain estimates of $Corr(R_t, R_{t-k}) = \rho_r^k$ at each lag k through

$$\rho_r^k = \frac{Corr\left(Y_t^2, Y_{t-k}^2\right)(2+3\nu)}{\nu}\,. \tag{5.2}$$

Johannesson et al. (2016b) propose the following method to estimate ρ_r: Estimates of $Corr\left(Y_t^2, Y_{t-k}^2\right)$ are calculated using the sample autocorrelation of Y^2, $r_{Y^2}(k)$, as defined in Equation (2.2). These values are then put into Equation (5.2) to obtain k estimates of ρ_r^k. Any estimate of ρ_r^k that exceeds one is set to one and any estimate that is smaller than zero is set to zero. Using these k values, estimates of ρ_r are obtained through

$$\hat{\rho}_{r(k)} = [\hat{\rho}_r^k]^{\frac{1}{k}}. \tag{5.3}$$

Finally, a weighted geometric mean – with the weight decaying as k increases – of these estimates is calculated to obtain an estimate of ρ_r through

$$\hat{\rho}_r = \left(\hat{\rho}_{r(1)}^k \hat{\rho}_{r(2)}^{k-1} \dots \hat{\rho}_{r(k)}^1\right)^{\frac{2}{k(k+1)}}. \tag{5.4}$$

No definite value of how many lags to utilize is suggested, however it is suggested that it should be kept low enough to keep some meaningful dependence (Johannesson et al., 2016b).

5.3.2 Estimating ρ_r by fitting the exponential decay

An alternative method of estimating ρ_r is proposed in this section. Recall that ρ_r^k is exponentially decaying for lag $k = 1,2,3$ .... Thus, it should be obvious that $\log(\rho_r^k)$ is linearly decreasing. Therefore, one could utilize Equation (5.2) and fit a simple linear regression of the n estimates of $\log(\hat{\rho}_r^k)$ with the lag $k = 1, \dots, n$ as a covariate through

$$\log(\hat{\rho}_r^k) = \alpha + \beta k + \epsilon_k \tag{5.5}$$

The estimated regression coefficient $\hat{\beta}$ can be viewed as an estimate of the linear decrease in $\log(\hat{\rho}_r^k)$ as the lag increases by one. Thus, $\exp(\hat{\beta})$ acts as an alternative estimate of ρ_r. This is visualized in Figure 5.1 using simulated data with $\rho_r = 0.99$. The red line represents the linear regression fit based on the simulated data.

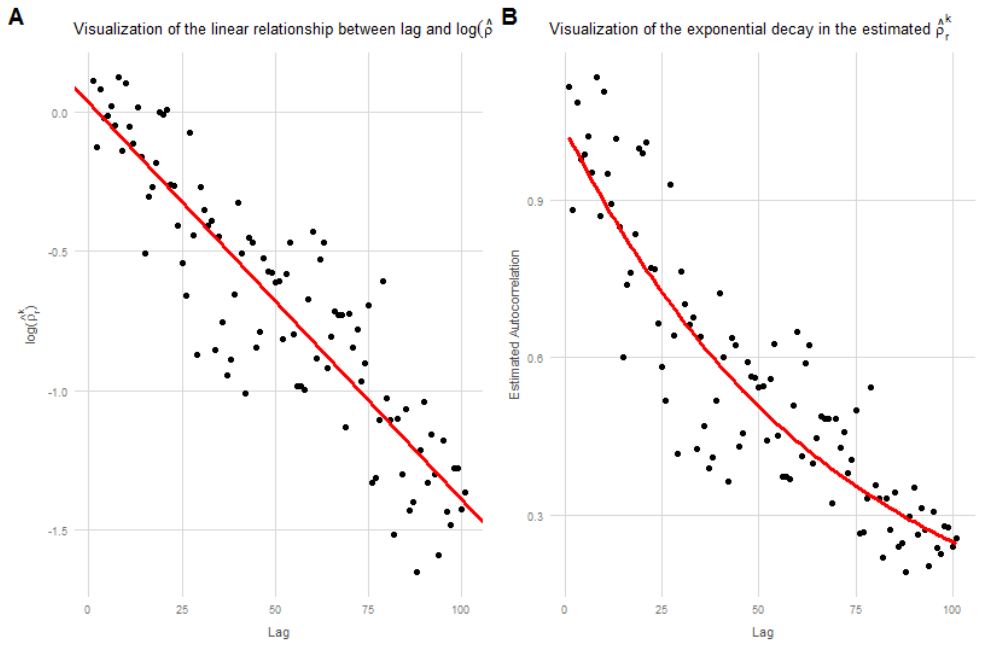

Figure 5.1 An illustration of the estimated $\hat{\rho}_r^k$ using data simulated with $\rho_r = 0.99$. Figure A illustrates the linear relationship between the lag k and the logarithm of $\hat{\rho}_r^k$. Figure B illustrates the exponential decay in $\hat{\rho}_r^k$. Using this relationship, a simple regression model can be fitted to the logarithm of $\hat{\rho}_r^k$.

Furthermore, a cut-off value of the number of $\hat{\rho}_r^k$ to include in this estimation process is proposed to be set to n, where lag $n+1$ is the first lag with non-significant autocorrelation in Y^2. When the real ρ_r is very large, this will provide a relatively large number of observations to be utilized, resulting in a more precise fit.

5.3.3 Comparison between the two methods

To compare the performance of the two methods, three sets of time series are simulated. Because the focus lies on financial time series, the performance is evaluated on data sets generated with parameters mimicking the characteristics of financial time series. The true values of ρ_r are set to $\rho_r = 0.99$, $\rho_r = 0.97$ and $\rho_r = 0.95$ respectively (or the correlation on a monthly basis, $0.99^{21} \approx 0.81$, $0.97^{21} = 0.53$ and $0.95^{21} \approx 0.34$ respectively). The true value of ν is set to 2 in all three sets. Each set of simulations consists of 10000 iterations of a time series with 3000 observations. A summary of the simulations is presented in Table 5.4.

This method of estimating ρ_r appears superior when X_t is white noise and ρ_r is very large, which is the case for volatile financial time series. However, do note that its performance in the general case when ρ_r could be any value between zero and one has not been thoroughly investigated.

Furthermore, one should be aware that it is theoretically possible for this method to yield invalid estimates. This can happen either when an estimate is larger than one – It is after all technically possible for the estimated slope $\hat{\beta}$ to be positive. The other way it can happen is if the first lag in the autocorrelation of Y^2 is below the level of significance, resulting in the linear regression being fitted to no data. However, these would be extremely uncommon with the values of ρ_r that are relevant in the context of financial time series and should only begin to present themselves when $\rho_r \leq 0.9$. In the case that this method cannot be properly utilized to a data set, the weighted geometric mean method should still provide valid estimates for the given data.

Ultimately, the final parameter ρ_r is estimated for the OMXS30 and NASDAQ Composite data. The estimators based on both the weighted geometric mean method and the proposed linear regression method are reported in Table 5.5. The estimated monthly correlation $\hat{\rho}_r^{21}$ is also reported to highlight the notable practical difference in the two estimates.

Table 5.4 A comparison of the performance in estimating ρ_r using the weighted geometric mean method and the proposed method that utilizes a linear fit to the logarithm of the exponential decay. Each set of simulations consist of 10000 iterations of time series with 3000 observations. The proposed linear regression method preforms better. Note that in all cases a very large value of ρ_r is used, its performance when ρ_r is lower has not been investigated.

	$\rho_r = 0.95$	$\rho_r = 0.97$	$\rho_r = 0.99$
Mean (sd) of $\hat{\rho}_r$ using the geometric mean	0.937 (0.032)	0.958 (0.024)	0.979 (0.015)
Mean (sd) of $\hat{\rho}_r$ using the lin. reg. method	0.944 (0.020)	0.964 (0.014)	0.985 (0.007)

Table 5.5 The estimates of ρ_r for the OMXS30 and NASDAQ Composite return series. In both cases, the estimate using both the weighted geometric mean method and the linear regression method is provided.

	$\hat{\rho}_r$ ($\hat{\rho}_r^{21}$), geometric mean	$\hat{\rho}_r$ ($\hat{\rho}_r^{21}$), lin. reg. method
OMXS30	0.975 (0.581)	0.985 (0.735)
NASDAQ Composite	0.984 (0.717)	0.980 (0.653)

6 Interpolation and prediction

The previous section demonstrates how parameters can be estimated, which provides information of the overall structure and general behavior of the volatility process in a given time series. The parameter estimates alone do not provide the ability to interpolate or predict the volatility R_t of a given day, both in the case of a future day $Y_{t+h}, h > 0$ or an already observed day Y_t. This section proposes a method of how to interpolate and predict the underlying volatility R_t of an observed time series Y_t using the volatility proxy Y_t^2. Initially, the conditional distribution, mean and variance of $R|Y^2$ is derived and presented. Subsequently, some basic methods of interpolating and predicting $\hat{R}_t$ are proposed and tested using simulated data.

6.1 The conditional distribution of $R|Y^2$

In real financial data Y_t is directly observed, while the proposed model consists of the two, individually unobserved, components R_t and X_t. As previously discussed,

X_t is assumed to be white noise and therefore poses no issues in terms of prediction. This leaves the task of predicting the volatility process $\hat{R}_{t+h}$, $h > 0$ given the past observation Y_u, $u \leq t$. Furthermore, this task can be rephrased in the simplest formulation into finding interpolation of $\hat{R}_t$, given the currently observed volatility proxy Y_t^2. The predictor that primarily is considered is the expected value of the conditional probability distribution function, i.e. $\hat{R} = E[R|Y^2]$. However, in the early stages, the mode of the probability distribution is tested and compared as a predictor to ensure it is not performing better than the more complex method that utilizes the expected value.

The first step to obtain this is to calculate the conditional probability distribution of $R|Y^2$. The definition of a conditional probability distribution function in the general case is

$$f_{A|B}(a|b) = \frac{f_{B|A}(b|a)f_A(a)}{f_B(b)}, \tag{6.1}$$

or more relevant to this case, the conditional probability distribution function is proportional to

$$f_{A|B}(a|b) \propto f_{B|A}(b|a)f_A(a). \tag{6.2}$$

Setting $A = R$ and $B = Y^2 = RX^2 = W$, the following proportional expression for the distribution of the volatility R_t, conditional on the observed volatility proxy is obtained:

$$f_{R|W}(r|w) \propto f_{W|R}(w|r)f_R(r). \tag{6.3}$$

One may be reminded that R follows a Gamma distribution with density

$$f_R(r) = \frac{1}{\Gamma(1/\nu)\nu^{1/\nu}} r^{1/\nu-1}e^{-r/\nu}. \tag{6.4}$$

With X_t being standard normal, it follows that $X^2 \sim \chi^2(1)$ with density

$$f_X(x) = \frac{1}{\Gamma(1/2)\sqrt{2}\sqrt{x}} e^{-x/2}. \tag{6.5}$$

Therefore, $f_{W|R}(w|r)$ can be considered a $\chi^2(1)$ distribution scaled by r with density

$$f_{W|R}(w|r) = \frac{1}{\Gamma(1/2)\sqrt{2} \cdot r\sqrt{w/r}} \cdot e^{-w/2r}. \tag{6.6}$$

Substituting Equation (6.4) and Equation (6.6) into Equation (6.3) yields the following proportional expression

$$f_{R|W}(r|w) \propto r^{\left(\frac{1}{\nu}-\frac{3}{2}\right)} e^{-\left(\frac{w}{2r}+\frac{r}{\nu}\right)}. \tag{6.7}$$

This expression is proportional to a known distribution, namely the generalized inverse Gaussian distribution, albeit commonly defined with alternative parameterization by Jørgensen (1982) as

$$f_R(x) = \frac{1}{2\eta K_p(\theta)} \left(\frac{x}{\eta}\right)^{p-1} e^{-\frac{\theta\left(\frac{x}{\eta}+\frac{\eta}{x}\right)}{2}} \tag{6.8}$$

where $\theta = \sqrt{2w/\nu}$, $\eta = \sqrt{w \cdot \nu/2}$, $p = 1/\nu - 1/2$ and K_p is a modified Bessel function of the second kind. Because the parameter ν is restricted to positive numbers in the ARGVG model, a simplified definition that applies under the condition that $p > -0.5$ can be used to define $K_p(\theta)$ (see Spanier & Oldham, 1987, p.500) as

$$K_p(\theta) = \frac{\sqrt{\pi}}{(p-1/2)!} \left(\frac{\theta}{2}\right)^p \int_1^\infty e^{-\theta x} (x^2-1)^{p-\frac{1}{2}} \, dx. \tag{6.9}$$

The expected value of the generalized inverse Gaussian distribution is given by

$$E[R] = \eta \frac{K_{p+1}(\theta)}{K_p(\theta)}$$
$$= \sqrt{(w \cdot \nu)/2} \cdot \frac{K_{\frac{1}{\nu}+\frac{1}{2}}\left(\sqrt{2w/\nu}\right)}{K_{\frac{1}{\nu}-\frac{1}{2}}\left(\sqrt{2w/\nu}\right)} \tag{6.10}$$

and

$$V[R] = \eta^2 \left[\frac{K_{p+2}(\theta)}{K_p(\theta)} - \frac{K_{p+1}(\theta)}{K_p(\theta)}\right]^2$$
$$= \frac{w \cdot \nu}{2} \cdot \left[\frac{K_{\frac{1}{\nu}+\frac{3}{2}}\left(\sqrt{2w/\nu}\right)}{K_{\frac{1}{\nu}-\frac{1}{2}}\left(\sqrt{2w/\nu}\right)} - \left(\frac{K_{\frac{1}{\nu}+\frac{1}{2}}\left(\sqrt{2w/\nu}\right)}{K_{\frac{1}{\nu}-\frac{1}{2}}\left(\sqrt{2w/\nu}\right)}\right)^2\right]. \tag{6.11}$$

Moreover, the mode of the distribution is defined as

$$Mode = \frac{1 - \frac{3\nu}{2} + \nu\sqrt{\left(\frac{1}{\nu} - \frac{3}{2}\right)^2 + \frac{2w}{\nu}}}{2}. \tag{6.12}$$

Note that the mode lacks any special functions such as Bessel functions, making it much simpler to calculate and use.

Density, quantiles, CDF and random number generation of the generalized inverse Gaussian distribution can be obtained through the R package *GeneralizedHyperbolic* by calling the functions *dgig*, *qgig*, *pgig* and *rgig* respectively (Scott, 2018). It should be noted to the reader that the *GeneralizedHyperbolic* package by default uses yet another alternative parameterization with the following conversions to what is used throughout this thesis: $chi = w$, $psi = 2/\nu$ and $lambda = 1/\nu - 1/2$.

Consequently, the density, expected value and variance of $R|W$ can easily be calculated. See Appendix 1 for a brief overview of the distributional properties of the generalized inverse Gaussian distribution.

6.2 Interpolation

As an initial test of the performance of the mode and expected value of $f_{R|W}(r|w)$ as predictors of $R|W$, the most basic scenario where Y, R and X are merely i.i.d. random variables are assumed. Based on 1 000 000 simulated observations of $X \sim N(0,1)$ and $R \sim Gamma(\nu = 2)$, generalized Laplace distributed observations of Y is constructed through $Y = \sqrt{R}X$. The values of $W = Y^2$ and $\hat{\nu}$ are then put into Equation (6.10) and Equation (6.12) to obtain $\hat{R}_{mean}$ and $\hat{R}_{mode}$ respectively. Moreover, the residuals res_{mean} and res_{mode} are calculated through $\hat{R} - R$. Histograms of the residuals based on both methods are displayed in Figure 6.1. Table 6.1 presents the mean, median and variance of the residuals.

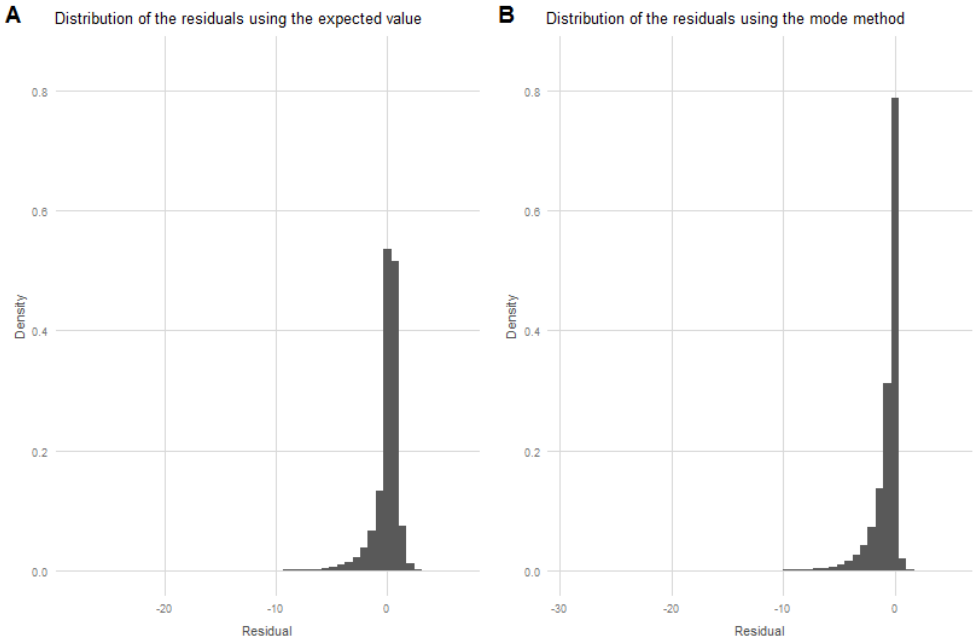

Figure 6.1 Histograms of the residuals obtained when interpolating $\hat{R}$, both using the conditional expected value (A) and using the mode of the conditional distribution (B).

Table 6.1 Mean, variance and median of the residuals based on both interpolation methods. The method utilizing the expected value of $R|W$ outperforms the method that utilizes the mode of the conditional distribution.

	$\hat{\mu}$	$\hat{\sigma}^2$	Median
Expected value method	≈ 0	1.32	0.28
Mode method	-0.7	1.42	-0.24

It can be concluded that the method utilizing the expected value of $R|W$ outperforms the method that utilizes the mode of the conditional distribution. No further attention will be given to the mode-method.

Furthermore, albeit primitive, this method may be applied to interpolate time series data as well: If Y_t^2 is known for a given day t, the interpolated volatility of day t, $\hat{R}_t$ may be calculated as instructed above. Testing this method on simulated time series data with $\rho_r = 0.99$ and $\nu = 2$ unsurprisingly shows that the performance is equal to the i.i.d. case.

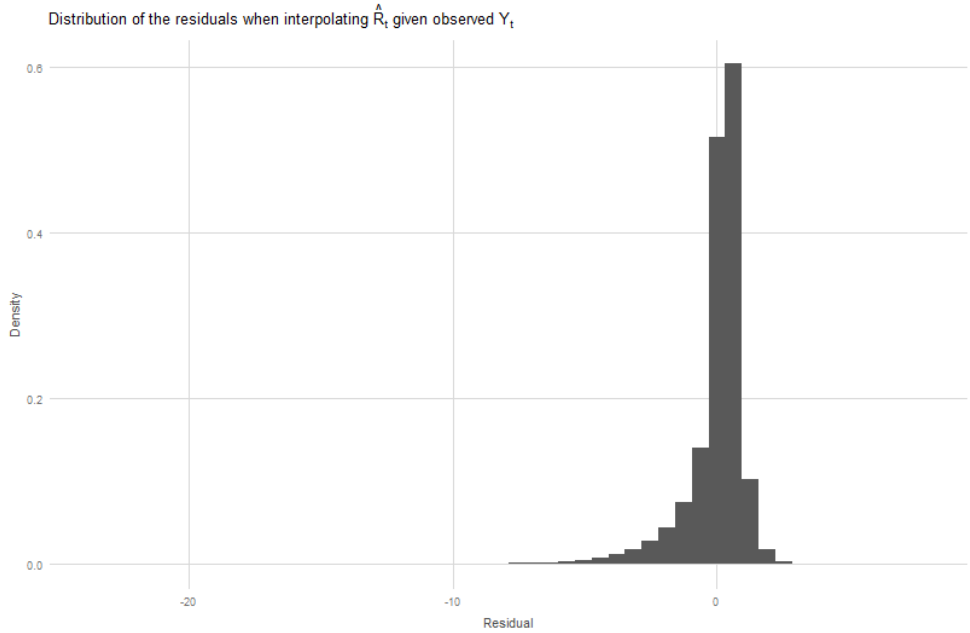

Figure 6.2 Histogram of the residuals obtained when interpolating $\hat{R}_t$ in a time series case. The method appears to perform equally well to that of the i.i.d. case in Figure 6.1.

Table 6.2 Mean, variance and median of the residuals obtained when interpolating $\hat{R}_t$ in a time series case. The results are equivalent to those in the i.i.d. case.

	$\hat{\mu}$	$\hat{\sigma}^2$	Median
res_{int}	≈ 0	1.33	0.28

6.3 Prediction

In many situations, interpolating the volatility is not what is the most important. Instead, what is of interest is to predict the volatility of day $t+1$, $\hat{R}_{t+1}$ using only the information available on day t, namely Y_t. One basic method to do this is presented in this section. One may be reminded that section 3.1 defines the R_t gamma process as

$$R_t = \rho_r K_t R_{t-1} + (1-\rho_r)\epsilon_t \tag{6.13}$$

where the random factors K_t are dependent on R_{t-1} and defined as

$$K_t = \frac{1}{m_t}\sum_{i=0}^{N_t} E_i \tag{6.14}$$

where $E_0 = 0$ and for $i > 0$ $E_i \sim Exp(1)$, and thus $E[E_i] = 1$. Furthermore, $N_t \sim Po(m_t)$, where m_t is given by

$$m_t = \frac{\rho_r/\nu}{1-\rho_r} R_{t-1}. \tag{6.15}$$

While K_t is dependent on R_{t-1} through m_t, one can easily see that $E[K_t] = 1$ regardless of the value of R_{t-1}. If R_{t-1} is large, it will provide a larger m_t, which in return will provide a larger N_t on average. Thus, the expected value of R_t may be defined as

$$E[R_t] = \rho_r E[K_t]E[R_{t-1}] + (1-\rho_r)E[\epsilon_t]. \tag{6.16}$$

Note that section 5.3 provided methods of estimating ρ_r, and the random innovations ϵ_t were defined in section 3.1 to have expected value $E[\epsilon_t] = 1$. With this information, Equation (6.16) can be rewritten as

$$\begin{aligned} E[R_t] &= \hat{\rho}_r \cdot 1 \cdot E[R_{t-1}] + (1 - \hat{\rho}_r) \cdot 1 \\ &= \hat{\rho}_r E[R_{t-1}] + (1 - \hat{\rho}_r). \end{aligned} \tag{6.17}$$

This can be further developed into the conditional expected value

$$E[R_{t+1}|R_t] = \hat{\rho}_r \hat{R}_t + (1 - \hat{\rho}_r), \tag{6.18}$$

of which both $\hat{\rho}_r$ and $\hat{R}_t$ can be estimated and interpolated using the previously presented methods.

In fact, because it is already known that ρ_r is very large in financial return series, one may already suspect that predicting $\hat{R}_{t+1}$ will yield very similar results to merely interpolating $\hat{R}_t$. Afterall, the very large ρ_r contributes to this in two ways. One: The second term of the conditional expectation will be very close to zero, while simultaneously merely scaling R_t in the first term by a percent or two. Two: By the definition of the correlation coefficient ρ_r, a very large value signals that R_t is changing very slowly. Thus, whether one interpolates $\hat{R}_t$ using Y_t^2 or predicts $\hat{R}_{t+1}$ using $\hat{R}_t$ (which in return is interpolated through Y_t^2), the real value of R_{t+1} seldom differs to a great extent from the real value of R_t due to the nature of the very slow-changing process.

To confirm this, another 1 000 000 observations of Y_t are simulated with $\rho_r = 0.99$ and $\nu = 2$. Equation (6.18) is then utilized to predict observation $t + 1$ using observation t for each of observation $t = 1, \dots, 999999$. The result confirms the hypothesis above. The histogram of the residuals is depicted in Figure 6.3. The mean, variance and median of the residuals are presented in Table 6.3. Both of which provide very similar results to the interpolation case in the previous section.

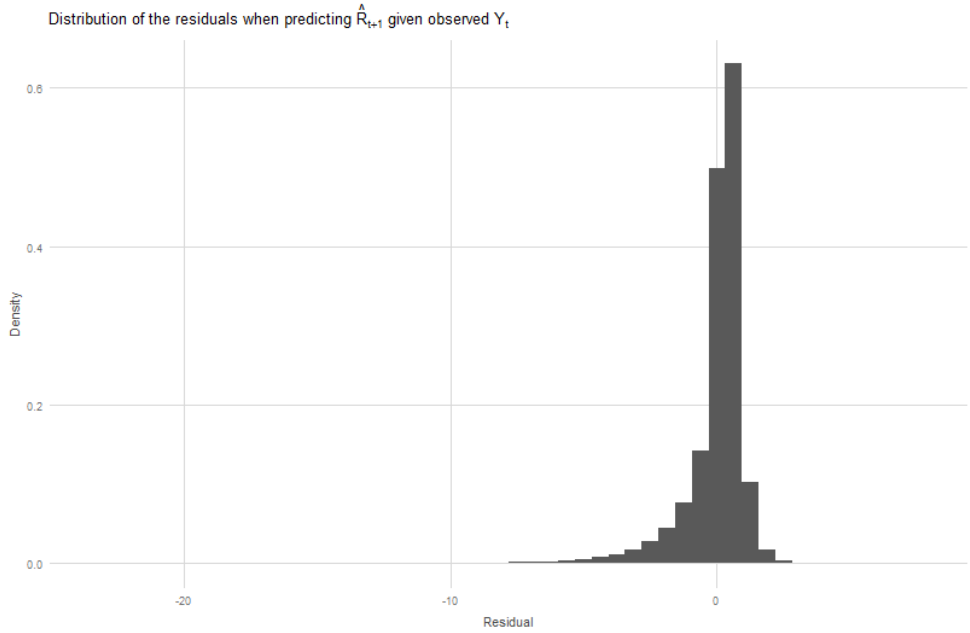

Figure 6.3 Histogram of the residuals obtained when predicting $\hat{R}_{t+1}$ using $\hat{R}_t$. The method appears to perform almost identically to the interpolation illustrated in Figure 6.2.

Table 6.3 Mean, variance and median of the residuals obtained when predicting $\hat{R}_{t+1}$ using $\hat{R}_t$. The results are close to identical to the residuals of the interpolation of $\hat{R}_t$.

	$\hat{\mu}$	$\hat{\sigma}^2$	Median
res_{pred}	≈ 0	1.36	0.29

6.4 Closing remarks on the interpolation and prediction method

It must be emphasized that while both the interpolation and prediction methods are unbiased estimators of the true R_t, and the distribution, mean and variance of the residuals indicate the predictor performs relatively well, the proposed methods are primitive and by no means fully developed.

Note that ν is a fixed estimated value throughout a full process R_t. The predictor $\hat{R}_t$ depends only on Y_t^2, and thus $\hat{R}_t$ will overall mimic the swings of a scaled Y_t^2, which does not resemble the smooth curve with slow changes of the true R_t. This is exemplified in Figure 6.4, where an R_t process with 3000 observations is simulated and subsequently predicted using Equation (6.18). While the residual mean is merely 0.08 in this case, it is still clear that $\hat{R}_t$ does not capture the full behavior of R.

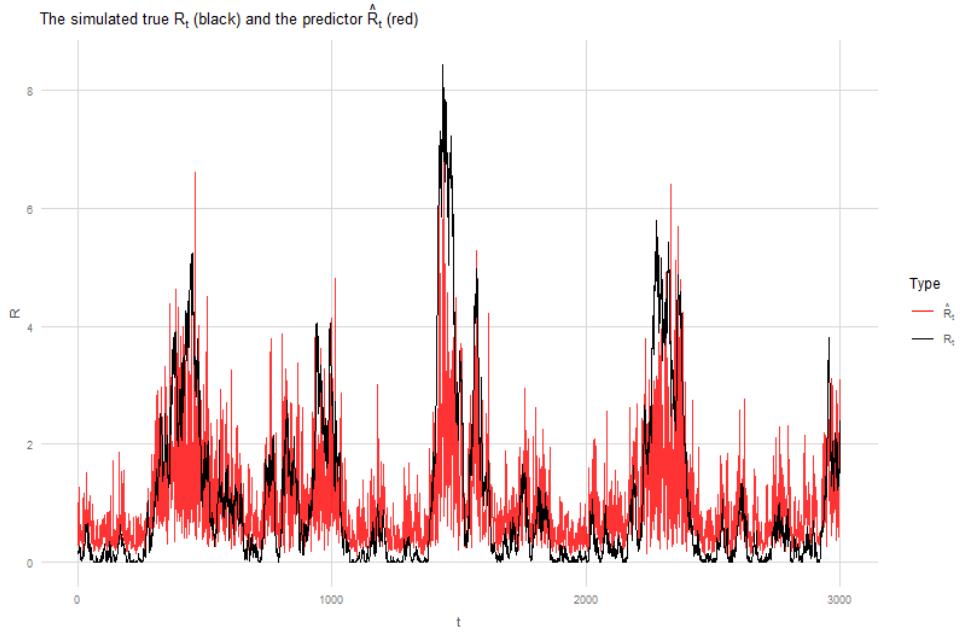

Figure 6.4 A simulated process R_t (black) with the predictor $\hat{R}_t$ (red). While the mean of the residuals is near zero, it is still clear that the predictor does not capture the full behavior of R.

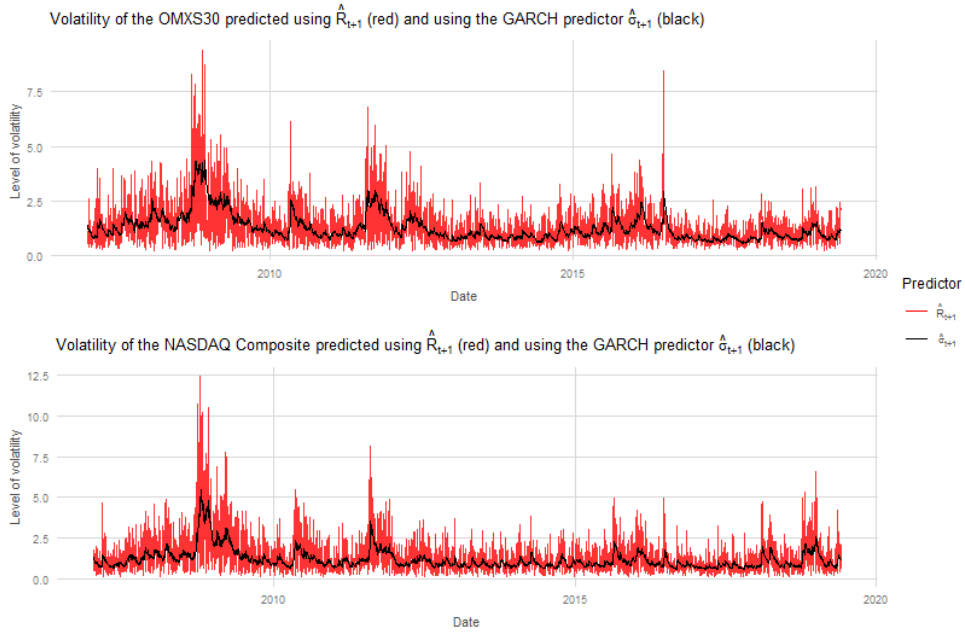

Figure 6.5 The predictor $\hat{R}_{t+1}$ tested on the OMXS30 and NASDAQ Composite return series. Because the true R_{t+1} is not known in real financial data, the predictor is compared to the GARCH predictor $\hat{\sigma}_{t+1}$.

In closing, the developed prediction method is tested on the OMXS30 and NASDAQ Composite. Because the volatility in real financial data is not directly measurable, the predictor cannot be compared to a true value of R_t as in the simulation study. Instead, the GARCH predictor $\hat{\sigma}_{t+1}$ is utilized as a substitute. Figure 6.5 depicts the two volatility predictors. The results are similar to the results of the simulation study.

7 Conclusion

The aim of this thesis has been to explore an alternative approach to model volatile financial time series using the autoregressive gamma variance Gaussian mixture model. Basic unbiased methods to interpolate and predict current and future values of the volatility process given the observed volatility proxy, the squared returns, have been developed and presented. However, as discussed in section 6.4, there is still a lot of room for improvement in the predictor. The developed predictor does not mimic the slow-changing nature of the true volatility process. In addition, an alternative method to estimate the correlation coefficient of the volatility process has been proposed and evaluated. The proposed method outperforms the original method when the true correlation is exceedingly large, which is the case for almost all financial return series.

To determine the performance of the model it needs to be compared to the well-established GARCH model(s). To accomplish this, further research is required to fully develop more advanced prediction methods that allow for the estimated volatility process to account for the fact that the process is extremely slow-changing.

It has also been demonstrated that the method of moments-estimate of the model parameter ν suffers from high variance when using 3000 observations (a reasonable length of financial data) and the correlation of the volatility process is very large. It is possible that better estimates could be obtained by other methods, e.g. maximum likelihood estimation.

Another direction of future research is to expand the model to the multivariate case. This thesis only considered the univariate case, however, the example data suggests that there may be a univariate volatility effect that influences all kinds of financial instruments. Thus, one could attempt to model entire portfolios of financial instruments with a single measure of the volatility process, combined with a multivariate normal ARIMA-process unique to each financial instrument.

8 References

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, vol. 31, no. 3, pp.307–327.

Cryer, J. D. & Chan, K. (2008). Time Series Analysis: With Applications in R, 2nd ed., New York: Springer.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, vol. 50, no. 4, pp.987–1007.

Fama, E. F. (1965). The Behavior of Stock-Market Prices, *The Journal of Business*, vol. 38, no. 1, pp.34–105.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance*, vol. 25, no. 2, pp.383–417.

Johannesson, P., Podgórski, K. & Rychlik, I. (2016a). Modelling Roughness of Road Profiles on Parallel Tracks Using Roughness Indicators, *International Journal of Vehicle Design*, vol. 70, no. 2, pp.183–210.

Johannesson, P., Podgórski, K., Rychlik, I. & Shariati, N. (2016b). AR(1) Time Series with Autoregressive Gamma Variance for Road Topography Modeling, *Probabilistic Engineering Mechanics*, vol. 43, pp.106–116.

Jørgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution, Vol. 9, [e-book] New York, NY: Springer New York, Available Online: http://link.springer.com/10.1007/978-1-4612-5698-4 [Accessed 2 September 2019].

Malkiel, B. G. (1989). Efficient Market Hypothesis, in J. Eatwell, M. Milgate, & P. Newman (eds), *Finance*, [e-book] London: Palgrave Macmillan UK, pp.127–134, Available Online: http://link.springer.com/10.1007/978-1-349-20213-3_13 [Accessed 12 August 2019].

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices, *The Journal of Business*, vol. 36, no. 4, pp.394–419.

Nasdaq Nordic. (2019). Vad Är OMX Stockholm 30 Index? - Nasdaq, *Nasdaq Nordic*, Available Online: http://www.nasdaqomxnordic.com/utbildning/optionerochterminer/vadaromxstockholm30index [Accessed 12 August 2019].

Pagan, A. (1996). The Econometrics of Financial Markets, *Journal of Empirical Finance*, vol. 3, no. 1, pp.15–102.

Scott, D. (2018). GeneralizedHyperbolic: The Generalized Hyperbolic Distribution. R Package Version 0.8-4, Available Online: https://CRAN.R-project.org/package=GeneralizedHyperbolic.

Sim, C. H. (1990). First-Order Autoregressive Models for Gamma and Exponential Processes, *Journal of Applied Probability*, vol. 27, no. 02, pp.325–332.

Spanier, J. & Oldham, K. B. (1987). An Atlas of Functions, Washington: Hemisphere Pub. Corp.

Tsay, R. S. (2010). Analysis of Financial Time Series, 3rd ed., Cambridge, Mass: Wiley.

Appendix 1

The surface of this distribution for various values of W and with a given value of $\nu = 2$ is visualized in Figure A.1. Moreover, surface plots of $E[R|W]$ and $V[R|W]$ for various combinations of ν and w are displayed in Figure A.2 and Figure A.3 respectively.

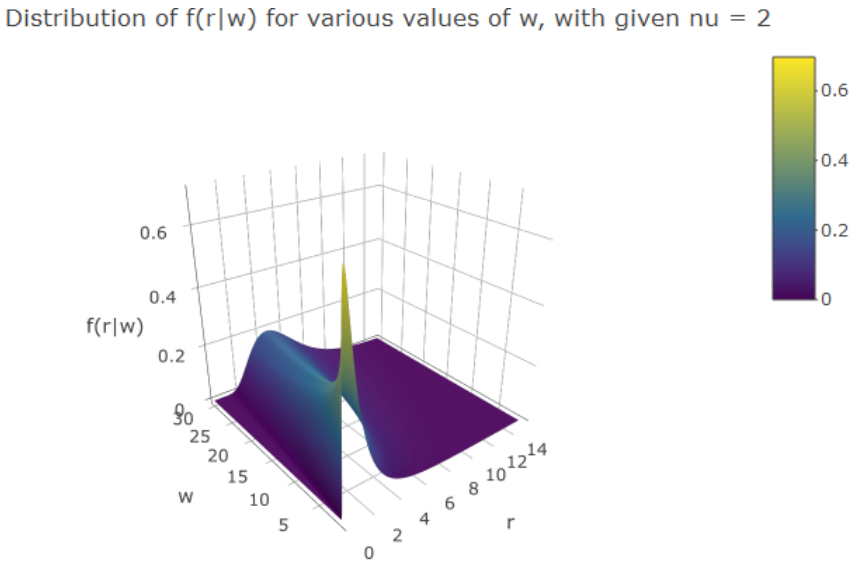

Figure A.1 The conditional distribution of $R|W$ for various values of w. ν is held constant at 2.

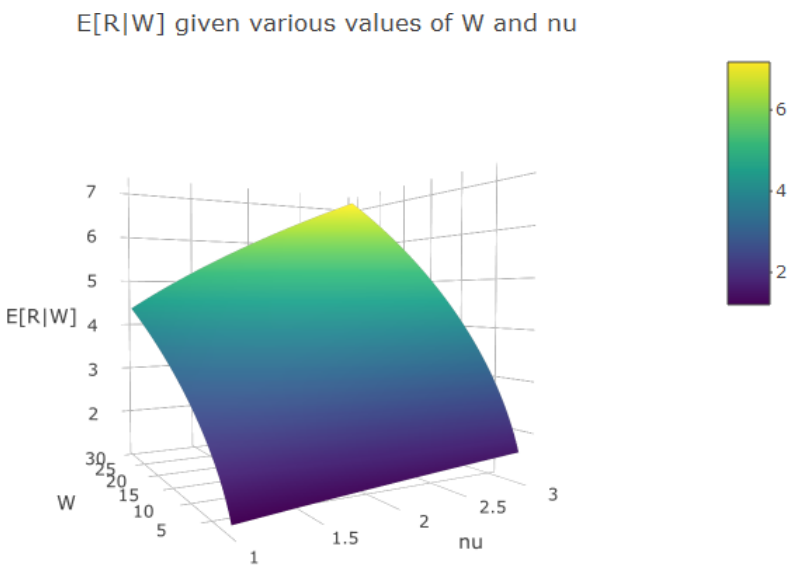

Figure A.2 Expected value of R|W for various combinations of ν and w.

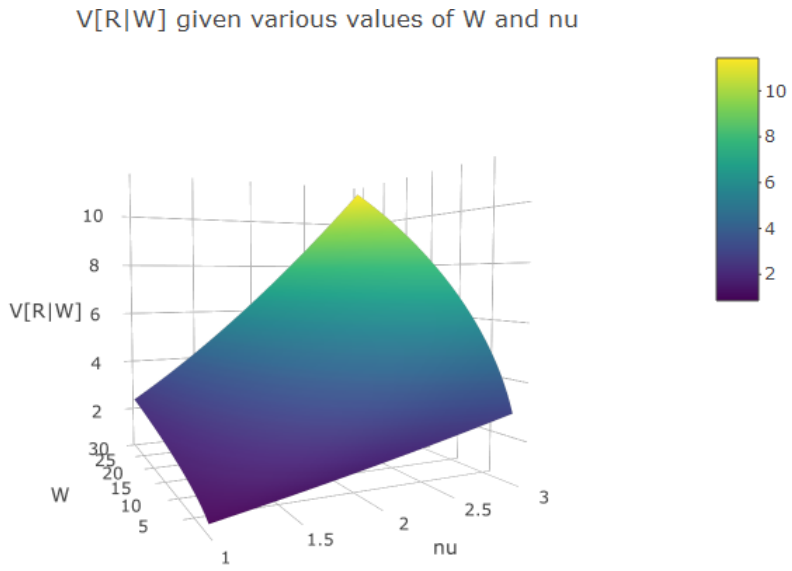

Figure A.3 Variance of R|W for various combinations of ν and w.