Norwegian School of Economics
Bergen, spring, 2015

NHH

Detection, modelling and implications of non-normality in financial economics

Normal inverse Gaussian modelling of Norwegian stock market returns and consumption growth

Mikael M. Bilet and Stig Roar H. Lundeby

Supervisor: Knut Kristian Aase

Master thesis, Economic Analysis (ECO)

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This thesis shows that the Norwegian stock market deviates significantly from what one might think of as a baseline model with identically and independently normally distributed returns. Firstly, the stock market return does not seem to be normally distributed over any observation frequency (daily, monthly and quarterly) we have investigated in this thesis. More specifically, the return distribution is both leptokurtic and negatively skewed. Secondly, the empirical return distribution is time-varying; we find both autocorrelation in returns and volatility clustering. Both of these deviations from the baseline model can potentially have important implications for theoretical models and practical applications.

In this paper, we will model the return distribution with a normal inverse Gaussian (NIG) distribution, which we indeed find to outperform Gaussian distributions both in- and out of sample. Our NIG modelling approach allows us to deviate from the normality assumption, but it is not able to capture the dependencies across time. This model of returns turns out to be useful in risk measurement, where the baseline model grossly underestimate well-known metrics such as value at risk and expected shortfall the NIG model fits these measures nicely.

This thesis also applies a bivariate NIG distribution to a theoretical model of equilibrium risk-free interest rates and the equity premium, suggested by Aase and Lillestøl (2015), in order to explain the equity premium puzzle. The NIG model allows for fatter tails and negative skewness in the joint return and consumption distribution, thereby reducing the implied risk aversion parameter and increasing the impatience rate of the representative consumer. Although the model takes us in the right direction in terms of both implied parameters, the improvement is only slightly more than negligible and it happens at the cost of a great increase in complexity.

Preface

This thesis completes our Master of Science degree with major in Economic Analysis (ECO) from the Norwegian School of Economics (NHH).

The chosen topic for this thesis reflects our interests in both Finance and Macroeconomics. This topic also allows us to use the knowledge we have acquired through the different courses taken as a part of our major. The work with this thesis has been challenging but also a rewarding experience.

We wish to thank our supervisor Knut Kristian Aase for useful feedback and guidance throughout the process. We also wish to thank Jostein Lillestøl for providing helpful guidance with respect to the modelling in R. Lastly we wish to thank Grethe Nielsen at Statistics Norway for inputs regarding our data collection.

Bergen, June 19, 2015

Mikael M. Bilet

Stig Roar H. Lundeby

Table of contents

1. Introduction

1.1 Background

Many theoretical results and applied models in finance and economics are based upon "mean-variance" analysis, i.e., the drivers of the results are the mean and variance of a certain process, e.g. stock market returns or a consumption process. In utility optimization problems, this type of analysis can only be justified if either the utility function is quadratic or if the process in consideration is normally distributed. In practical applications in finance, e.g. portfolio risk or performance measurement, "mean-variance" analysis is only justified if the return distribution is approximately normal.

In the case of quadratic utility, the marginal utility is a linear function, which implies that all we need to determine expected utility is the mean and variance of the underlying process. However, quadratic utility exhibits increasing absolute risk aversion, which has the unfortunate implication that the dollar amount invested in risky assets is decreasing in wealth.[1] This is in contrast to empirical observations (see e.g. Friend and Blume (1975)) and is therefore a problematic assumption whenever portfolio optimization is an integral part of the analysis.

Whenever the process in consideration follows a Gaussian law, the mean and variance is enough to describe the entire process. Indeed, there is some theoretical backing for assuming Gaussian processes from the central limit theorem (CLT); the sum of a large number of identically and independently distributed (i.i.d.) random variables with a well-defined mean and variance, is normally distributed, regardless of the underlying distribution. There are however, at least two reasons to give pause at such an argument. Firstly, in many situations the i.i.d. assumption does not hold, in particular; there might be significant dependence across time. Secondly, in some situations it might be the case that neither the mean nor the variance is well defined, e.g. Mandelbrot (1963) suggested that infinite variance might explain the non-normality in financial

[1] To see this, consider the certainty equivalent (CE) in the case of constant absolute risk aversion (CARA) of a lottery $\tilde{h}$ that returns $h > 0$ and 0 with equal probability. The CARA utility function is given by $u(w) = 1 - e^{-\alpha w}$, where the coefficient of absolute risk aversion $A(w) = \alpha > 0$. The CE of this lottery is given implicitly by $E[u(w + \tilde{h})] = \frac{1}{2}[u(w + h) + u(w)] = u(w + CE)$ and explicitly as $CE = \frac{1}{\alpha} ln\left[\frac{2}{1+e^{-\alpha h}}\right] > 0$. Notice that the CE is decreasing in *A(w)*, which (anecdotally) means that a utility function that exhibits increasing absolute risk aversion in wealth, implies a decreasing CE in wealth.

data. If either of these underlying assumptions of the CLT should fail, we have less theoretical backing for assuming normality.

This paper consists of four chapters that are looking at what might seem as different issues. There is however, a main theme throughout the entire thesis – deviations from normality. Each chapter attempts to provide some valuable insight as to when a normality assumption is justified and when it is not. The thesis as a whole is meant to provide an overview of how non-normality might be modelled, in both a univariate and bivariate setting, and its implications. Furthermore, even though some topics, e.g. practical risk measurement and theoretical models for stock market returns and risk-free rate might not seem immediately connected, we expect on intuitive grounds that a ceteris paribus (i.e. keep the first two moments constant) increase in tail-risk on the stock market should give rise to a higher risk premium.

1.2 Outline of this thesis

Chapter 2 provides an empirical investigation of Norwegian stock market returns. Here we will explain how the return distribution differs from what one might think of as a baseline model – identically and independently normally distributed returns. In particular, we find that the return distribution deviates in two major ways: there is significant non-normality and it is time-varying. Tests for skewness and kurtosis are also applied to investigate what kind of non-normality we have in our data. All our results provide motivation for leaving the baseline model and this paper will do so by abandoning the normality assumption. However, we will not allow for a time-varying return distribution, with the exception of a brief discussion in chapter 3.

Chapter 3 presents an alternative model for the return distribution, namely the normal inverse Gaussian (NIG) distribution. This distribution has many attractive features when modeling financial data, one of which is that it has a (relatively) simple moment generating function, which also implies that moments of all orders exist. Another important feature is that it allows for skewness, heavy tails and peakedness in the distribution of returns. In this chapter, we will also use the maximum likelihood approach to estimate the parameters of the NIG distribution in two different ways: a direct and an indirect approach. The direct approach simply fits the NIG distribution to the Norwegian stock market returns. The indirect approach first fits an inverse Gaussian (IG) distribution to a variance series and then mixes this distribution with a normal distribution in order to obtain a NIG model of returns. The final section of the chapter provides a

way of assessing the parameter uncertainty in the model using a parametric bootstrap approach, and the parameter uncertainty for monthly returns is shown to be quite extensive.

Chapter 4 applies a NIG model of daily ex-dividend returns to risk measurement. More specifically, we test how well the model fits two simple risk metrics: value at risk (VaR) and expected shortfall (ES). It is important to note that the chapter is not aiming to provide a thorough analysis/description of risk measurement as a whole or the specific risk measures. Instead, we simply aim to illustrate what is to gain by abandoning a simple Gaussian framework in favor of more complex distributional assumptions. This chapter provides overwhelming support for the NIG model when tested against Gaussian models, both in sample and out of sample.

Chapter 5 encompasses three (more or less) distinct topics: an empirical analysis of the Norwegian consumption growth, an empirical analysis of the joint consumption growth and stock market returns and a theoretical model for the equity premium and risk free rate in a NIG framework suggested by Aase and Lillestøl (2015). We find that the consumption growth does not deviate significantly from an i.i.d. Gaussian process. We do however find that the joint return and consumption process significantly deviates from bivariate normality. More specifically, the joint distribution exhibits both coskewness and cokurtosis. In the final sections of chapter 4, we test whether the observed non-normal properties of the joint distribution is enough to explain the equity premium puzzle posed by Mehra and Prescott (1985). It turns out that non-normality is only able to explain a small fraction of the equity premium at the cost of rather uninformative and complex expressions for the equity premium and risk-free rate. This is however in itself a useful insight – the normality assumption might be justified in (some) equilibrium models.

1.3 Data description

The first four datasets are daily, monthly and quarterly Norwegian stock market indices and a monthly US stock market index. Additionally, we also have datasets on population, consumption, inflation and three-month Norwegian government bill rates.

Our monthly index data set, which is applied in most of the analysis of chapter 3, consists of the cum dividend MSCI index for mid and large-cap stocks in the Norwegian stock market (MSCI, 2015a; MSCI, 2015b). For our quarterly analysis in chapter 4, we have aggregated monthly returns into quarterly returns. Daily returns are calculated from an ex dividend version of the

same MSCI index, due to unavailability of a daily cum dividend index[2]. We will also present some results from the US stock market based on a similar mid and large-cap index for the US generated by MSCI (2015a). All these data sets are covering the period from the beginning of 1970 until the end of 2014, except daily Norwegian returns, which covers January 1972 to February 2015.

Data for seasonally and inflation adjusted Norwegian quarterly consumption from 1978 to 2014, population data[3] and inflation data are collected from Statistics Norway (2015), while data for Norwegian 3 month treasury rates are from Eitrheim, Klovland, and Qvigstad (2007)[4].

2. Empirical analysis of Norwegian stock market returns

2.1 Stylized facts

2.1.1 Return data

In this sub-section, we analyze return data. For daily, monthly and quarterly indices, we have calculated a continuously compounded return series the following way:

$$r_t = \log\left(\frac{\text{Level of index}_{\text{t}}}{\text{Level of index}_{\text{t}-1}}\right), \quad (1)$$

[2] To get a feeling for whether (and how) our results might be influenced by using ex dividend returns as opposed to cum dividend returns, we have calculated several descriptive statistics for a time interval where daily returns from both indices are available (2001-2015). Firstly, the correlation between total returns and price returns is 0.99. Secondly, the average return on the price index has been 0.000164 while it has been 0.0003095 on the total return index, the difference representing dividend payments. Thirdly, the volatility of the price return and total return are 0.01564 and 0.01562 respectively. An almost equal volatility, almost perfect correlation and about twice as high return for total returns imply that dividend payments have been a near constant share of the total return index from 2001. This implied low volatility of dividend yields compared to total returns is consistent with empirical findings for the US (Shiller, 1981; Campbell & Shiller, 1988). Shiller (1981) shows, conventionally assuming stock prices to be expected future dividends discounted, that the volatility of dividends is way too low in order to explain the observed volatility in stock returns (the so-called "volatility puzzle", for recent explanations of this puzzle see for instance van Binsbergen, Brandt, and Koijen (2012)). Minimum, maximum, sample skewness and sample kurtosis are also calculated and found to be similar in the two indices. The main relevance for us is that the ex-dividend returns will deliver results similar to what we would obtain by using a cum-dividend return data set with a constant shift-parameter on the mean.

[3] Population data from 1978-1998 is collected on a yearly basis while 1998-2014 is on a quarterly basis.

[4] Up until 2003 the 3 month treasury rates are collected from Eitrheim, Klovland and Qvigstad (2007). From 2003 and onwards the treasury rates are calculated as an average of the yields for the treasuries each day, collected from Macrobond (2015).

where t corresponds to the current day, month or quarter and log represents the natural logarithm. Henceforth we will refer to the continuously compounded returns simply as returns.

In Table 1 we have calculated some simple descriptive statistics for the Norwegian daily, monthly and quarterly returns, in addition to monthly US returns for comparison. The monthly returns in the Norwegian stock market have been between negative 35 percent in October 1987 and positive 23 percent in April 1973. The average monthly return has been 0.87 percent, which corresponds to an annual return of 10.39 percent. For comparison the US market generated a monthly return of 0.81% from 1970 to 2014.

We also note that the Norwegian market is more volatile than the US market. The coefficient of variation, defined as the sample standard deviation divided by the mean, is 8.32 in Norway compared to 5.49 in the US. The coefficient of variation decreases when we aggregate from daily to monthly to quarterly returns.

The skewness and kurtosis measure presented in Table 1 are standardized, which means that they can be used to compare the empirical distributions of returns for all time intervals, regardless of scale. The sample skewness and sample kurtosis are calculated by using the R-package *e1071* developed by Meyer, Dimitriadou, Hornik, Weingessel, and Leisch (2014), which uses the following expressions from Joanes and Gill (1997):

$$\text{Sample skewness} = \frac{m_3}{{m_2}^{\frac{3}{2}}}, \tag{2}$$

$$\text{Sample kurtosis} = \frac{m_4}{{m_2}^2}, \tag{3}$$

where m_q is the sample moment of order q calculated as

$$m_q = \sum_{i=1}^{n} \frac{(r_i - \bar{r})^q}{n}. \tag{4}$$

Here n is the total number of observations and $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$. Joanes and Gill (1997) point out that these measures are not unbiased estimates of the population moments. For our purpose however

these measures works well enough, and we only get negligible differences by using the alternative measures suggested by Joanes and Gill (1997).

The Gaussian distribution has skewness always equal to zero and kurtosis (by our measure) always equal to three[5]. The calculated sample skewness and kurtosis differ from what is expected by a Gaussian distribution in all our data sets, which is an indicator of stock returns being non-normal. We will address the issue of non-normality further in section 2.3.

	Daily	Monthly	Quarterly	US Monthly
Min	-0.23697	-0.35229	-0.56850	-0.23855
Max	0.11440	0.23189	0.42932	0.16374
Mean	0.00027	0.00866	0.02576	0.00810
Variance	0.00022	0.00518	0.01906	0.00198
Coefficient of variation (CV)	54.53	8.31	5.36	5.49
Skewness	-0.5752	-0.7410	-0.7253	-0.6727
Kurtosis	11.870	5.445	5.206	5.525
N	10838	541	180	542
P-value in normality test	NA	6.732e-10	8.315e-05	5.323e-09

Table 1: Descriptive statistics and sample moments for stock returns.

In Figure 1, we have plotted the monthly return data from January 1970 until December 2014. We notice that there are no linear long-term trends in the returns[6]. Figure 1 shows no obvious signs of autocorrelations of returns, but we will address this more formally in section 2.2. Another feature worth mentioning is that extreme negative outcomes are more severe than extreme positive outcomes. This is a common feature of stock returns also noted by for instance Duffie and Pan (1997). The larger absolute size of negative returns compared to positive returns is connected with the negative skewness in returns. Figure 1 also seems to point to non-constant variation in stock returns over time, which is another common feature in stock returns (Engle,

[5] Conventionally, when we use the term excess kurtosis in this paper we refer to our calculated kurtosis minus three, i.e., the kurtosis relative to the Gaussian distribution.
[6] We have run an OLS regression that confirms that there is no statistically significant trend in stock returns.

2002). It seems for instance that the years 1976-1978 were years with low variation in the stock returns, while 2007-2009 were a period of high variation. We will address this issue of volatility clustering further by analyzing volatility of returns by itself in sub-section 2.1.2.

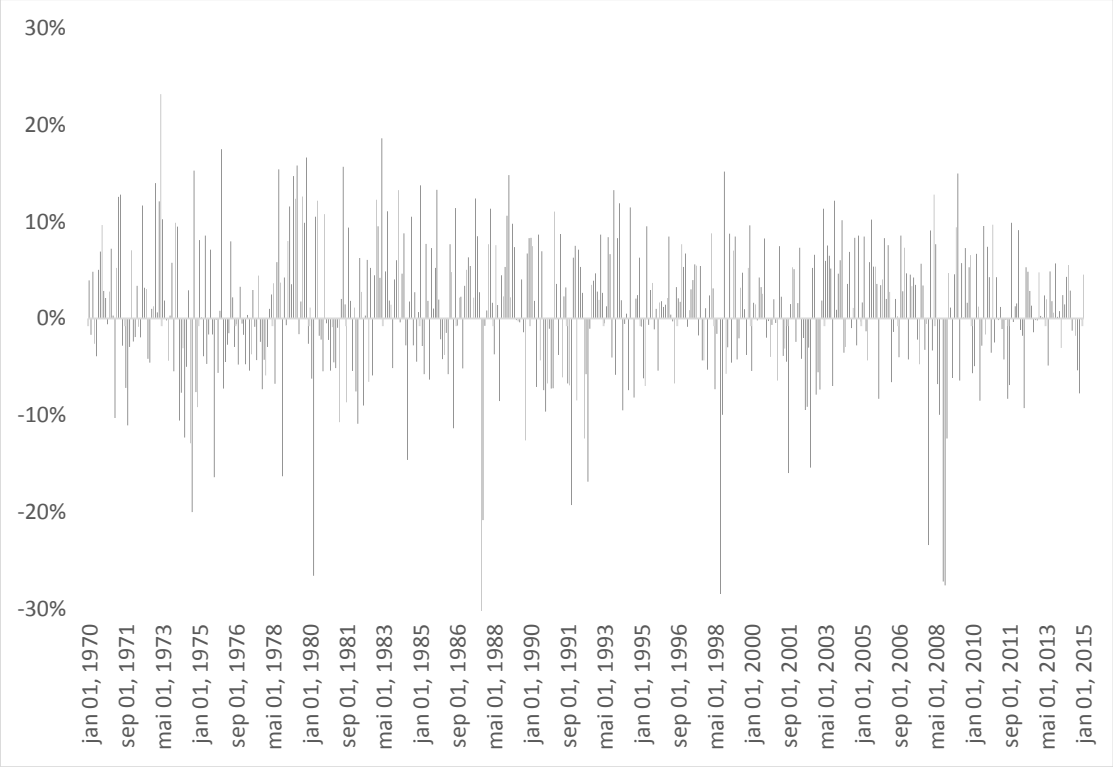

Figure 1: Monthly continuously compounded returns (1970-2014).

In Figure 2, we have made a density plot (histogram) of our return data, together with a Gaussian distribution with the same mean and variance. As previously stated, the left tail seem more heavy than the right tail (negative skewness in the data). There also seems to be more weight at the center (high peakedness) of the distribution compared to the Gaussian distribution. The combination of peakedness and heavy tails gives us a sample kurtosis greater than 3 (Balanda & MacGillivray, 1988). The return data seems to deviate from a Gaussian distribution, and we will test whether this is the case in section 2.3.

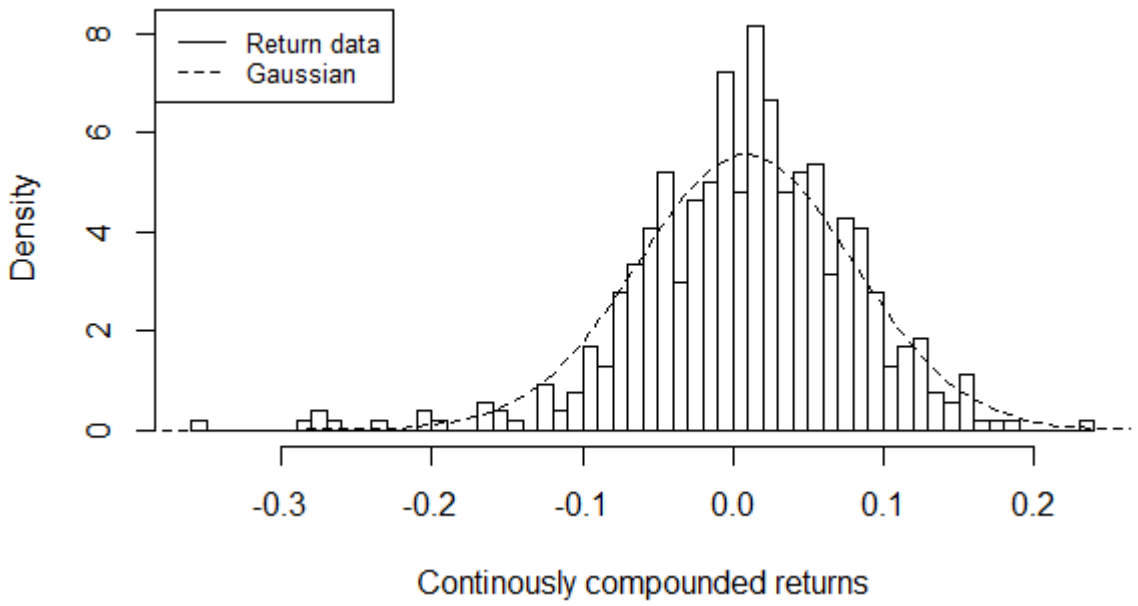

Figure 2: Distribution of monthly returns (1970-2014).

2.1.2 Volatility data

In this sub-section, we analyze the volatility of monthly returns, calculated using the MSCI (2015a) index for monthly data. Volatility at time t is calculated as

$$\widehat{\sigma_t} = \sqrt{\frac{1}{n-1}\sum_{i=1-\frac{n}{2}}^{\frac{n}{2}}(r_{t+i} - \overline{r_t})^2}, \tag{5}$$

where $\bar{r}$ is defined as

$$\overline{r_t} = \frac{1}{n}\sum_{i=1-\frac{n}{2}}^{\frac{n}{2}} r_{t+i}. \tag{6}$$

When we are calculating monthly[7] volatility, we take $n = 12$, which we have plotted in Figure 3. This picture is in line with our earlier observation that the volatility of stock returns are changing over time. We will also formally test whether this is the case in section 2.2. There are no obvious trends in the monthly volatility data and it seems to be mean reverting, with a mean around seven percent.

[7] For daily data these measures are defined the same way but with $n = 30$.

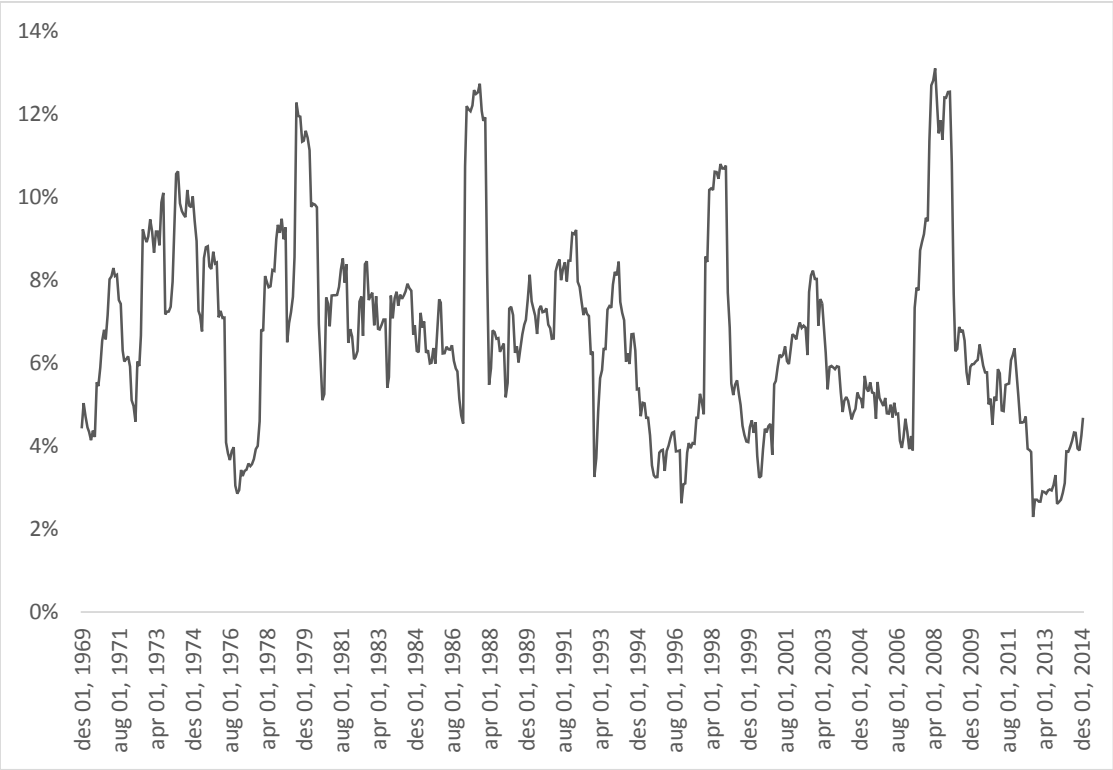

Figure 3: Monthly volatility of returns (1970-2014).

Figure 4 is a density plot of return variances similar to the one for returns in Figure 2. This variance is defined as

$$\text{Variance at time t} := \hat{\sigma}_t^{\,2}, \tag{7}$$

where $\hat{\sigma}$ is the volatility calculated by equation (5). One obvious observation is that variance, like volatility, is non-negative. This has implications for model specifications – a good model of variance needs to be restricted away from negative territory. Another observation from Figure 4 is that it exhibits most of its weight in the interval [0.000, 0.010]. There are however, also some observations with substantially higher variance, i.e. the variance has a heavy right tail.

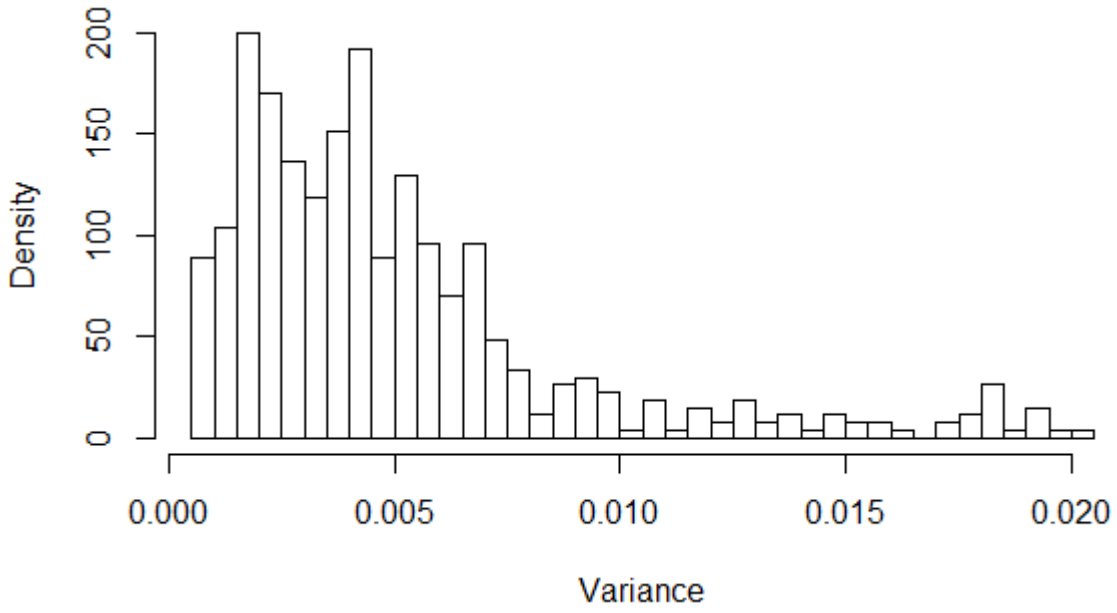

Figure 4: Distribution of variance of monthly returns.

2.1.3 Return and volatility data

In Figure 5, we have plotted monthly volatility together with $\bar{r}$ from equation (6), which has the interpretation of a monthly smoothed average return. The reason why this is an interesting plot is that it describes how stock returns are related to volatility; more specifically, volatility and return seems to be negatively correlated. Schwert (2011) finds similar patterns in the US stock market. We will investigate this correlation further using regression analyses in section 2.2. Note however, that the relationship between volatility and returns before the early 1980s seems to be a positive one[8]. The shift in the 1980s is something we will touch upon again in this thesis, as it has an effect on some of our results in chapter 4.

[8] The correlation between $\hat{\sigma}$ and $\bar{r}$ is 0.0647 before 1983 while -0.5984 after 1983.

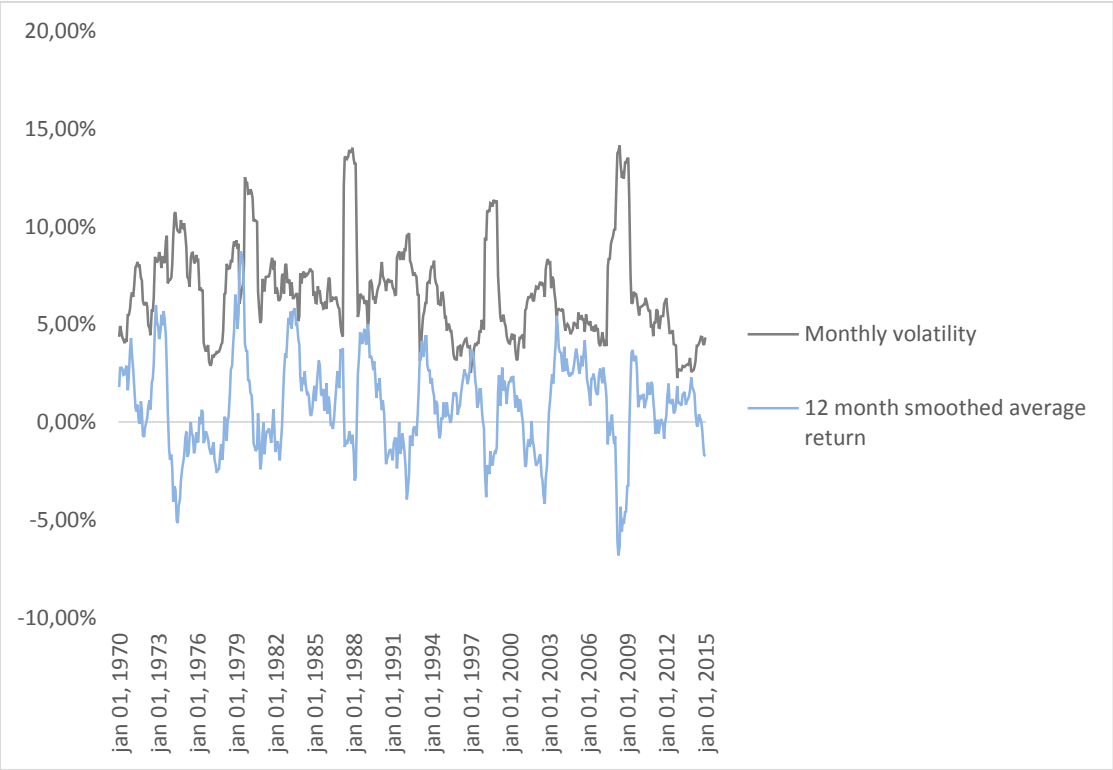

Figure 5: Monthly volatility of returns and smoothed average returns.

2.2 Regressions

In Table 2, we present some results of regressions addressing some of the stylized facts mentioned in the previous sections. The dependent variables (Y) in the six regressions are given in the top row, while independent variables (X) are given in the left column. All regressions are performed on the dataset of monthly returns on the Norwegian stock market, which consists of about 540 observations. Each regression has only one independent variable and takes the form

$$Y = \alpha + \beta X.$$

	Residuals squared	|Residuals|	Returns	Returns	ReturnBar	ReturnBar
Intercept	0.0042***	0.0475***	0.0073*	0.0206***	0.0311***	0.0204***
Lagged residuals squared	0.1987***					
Lagged |residuals|		0.1227**				
Lagged returns			0.1491***			
Residuals squared				-2.3183***		
Volatility					-0.3386***	
Variance						-2.3479***
N	540	540	540	541	541	541

*Table 2: Regression analyses. Significance levels: '***' 0.000, '**' 0.001, '*' 0.01, '.' 0.05.*

The first regression is estimated by regressing squared residuals on its lagged value. The residuals squared is calculated as follows:

$$\text{residuals}_t^2 = \left[r_t - \frac{1}{n} \sum_{i=1}^{n} r_i \right]^2, \tag{8}$$

where n is the total number of observations. The regression results give a significantly positive effect of lagged squared residuals on squared residuals, which is evidence for autocorrelation in residuals squared. This result suggests that we have volatility clustering in our data. The second regression model, where the square root of residuals_t^2 (the absolute residuals) is regressed on its lagged value, confirms this. The interpretation of a significantly positive coefficient on the lag is that the volatility we observed in the previous month can help predict the volatility seen this month[9]. Figure A1 and Figure A2 in Appendix A plot the autocorrelation functions (ACF)[10] connected to these two regressions, which allows us to evaluate lags further back.

The third regression in Table 2 tests whether there is a relationship between return and lagged return. The result is a statistically significant autocorrelation, which implies that return this period in fact can help predict return next period. The significant autocorrelation in stock returns (and

[9] We interpret the residuals here as volatility even though it is not calculated the same way as volatility in equation (5).
[10] Regressions like the first two in Table 2 but with more lags.

volatility) is surprising if one believes the efficient market hypothesis, which states that all prices capture all available information at any time. We note however that the size of the autocorrelation is 0.15, which means that it might be economically insignificant when accounting for transactions costs, etc. Significant autocorrelation in returns for both individual stocks and indices is, however, a stylized fact in empirical finance according to Anderson, Eom, Hahn, and Park (2012)[11]. Anderson et al. (2012) make a review of the literature on this topic and address the possible explanations to this phenomenon in fixed interval stock returns. In our analysis in chapters 3-5, we will sometimes make the assumption of identically and independently distributed returns. We will therefore discuss this issue further in these chapters.

Regressions four to six address the correlation between stock returns and the volatility of the returns. The three regressions provide the same qualitative results. Regression number four suggest a significant negative relationship between monthly stock returns and its residuals squared as defined in equation (8). Regression number five gives the same negative relationship between our constructed volatility measure in equation (5) and its corresponding $\bar{r}$ from equation (6). Finally, we get a significantly negative relationship between $\bar{r}$ and the variance as defined in (7). These regressions confirm our observation in sub-section 2.1.3 where we pointed out that a period of poor performance of the index is typically connected with high volatility in the market[12].

The regression analyses confirm that some of our findings in section 2.1 – non-constant volatility and its negative correlation with the stock market returns – are statistically significant. As we will point out in chapter 3, these properties of the volatility have implications for the distributional assumptions underlying a model of stock returns.

2.3 Normality

This final section of chapter 2 tests whether monthly stock returns are normally distributed. Stock returns are often (implicitly or explicitly) assumed to be normally distributed in the finance and economics literature, making many theoretical and empirical results dependent on this underlying assumption. For this reason, the assumption of Gaussian returns has received extensive attention

[11] This autocorrelation is common for stock returns at fixed intervals (daily, monthly, quarterly), as in our case.

[12] We also ran a regression analysis of $\bar{r}$ on $\hat{\sigma}$ and a dummy variable equal to 1 after 1983 and 0 before. This regression showed that the correlation between $\bar{r}$ and $\hat{\sigma}$ was positive before 1983 and negative after, and the difference is statistically significant.

in research (Mandelbrot, 1963; Asparicio & Estrada, 1997). This section evaluates whether there is non-normality in the Norwegian stock market returns and provides possible explanations for this finding.

2.3.1 Normality test

In Table 1, we have provided P-values for the Shapiro and Wilk (1965) test (R Core Team, 2014) for normality on our four datasets of stock returns. The null hypothesis of this test is that the data are drawn from a Gaussian distribution. In all our four cases, this null hypothesis can be rejected with P-values close to zero[13]. Field (2009) however points out that this test is biased by sample size in such a way that for large samples, the P-value could be low even though the deviations from normality are negligible. It is therefore important to supplement the test results with Q-Q plots in order to fully conclude whether data is normal (Field, 2009).

In Figure 6, we provide a normal Q-Q plot with sample quantiles on the vertical axis and theoretical (normal) quantiles on the horizontal axis. If the data is drawn from a Gaussian distribution, all data points should to be located on a straight line. We can clearly see that there are deviations from a straight line in Figure 6; we therefore conclude that the data are non-normal. Figure A3, A4 and A5 in Appendix A, provide similar normal Q-Q plots for daily and quarterly returns on the Norwegian stock market and monthly returns on the US market, respectively. These plots corroborate the conclusion from Table 1. The normality test for US monthly returns has a somewhat higher P-value than for the Norwegian monthly returns. Additionally, the data points lie closer to a straight line in the US Q-Q plot. This may be an indication of the US market being closer to normal than the Norwegian one for monthly return data, even though the conclusion of non-normality is (qualitatively) the same in both markets. Another observation is that monthly and quarterly returns seem to be closer to normality than daily returns.

The central limit theorem (CLT) states that the sum of identically and independently distributed (i.i.d.) random variables with well-defined expected value and variance will approach normality when the number of observations gets sufficiently large, regardless of the underlying distribution

[13] For the daily return data a P-value is not provided due to the fact that the test does not handle that large amount of data. The Q-Q plot for daily return data in Figure A3 in Appendix A however makes it clear that we draw the right conclusion by rejecting the null hypothesis.

(Rice, 2007). One can view stock market returns at any frequency as a sum of returns of higher frequency, thus CLT provides a hypothesis that even daily returns should be normal. There are however, at least two possible explanations for why CLT might break down for stock returns: firstly, the returns might not have well-defined expectation or variance, and secondly returns might not be i.i.d. Mandelbrot (1963) advocated the first explanation, more specifically that returns in the financial markets have infinite variance. The regression analyses made in section 2.2 provide some support for the second explanation – we find that returns are dependent on its own lagged value and negatively correlated with volatility, which in turn is dependent on its own lagged value.

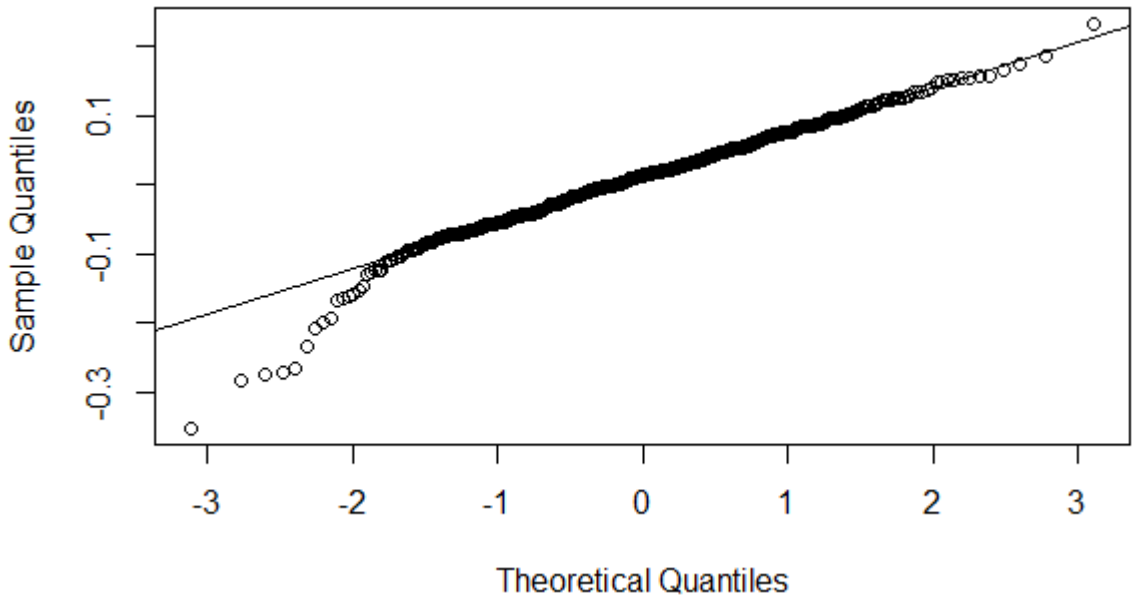

Figure 6: Normal Q-Q plot for monthly returns.

We see from Figure 6 that there are a few outliers in our dataset. There are observations in our sample that have a negative return of more than 20 percent, even though in a theoretical normal distribution, this would occur a lot less frequent than our data suggest[14]. We investigate the origin of the non-normality by removing the extreme negative outcomes (remove returns less than -0.2) and then evaluating whether the data is normal. The dataset is reduced by eight extreme negative outcomes and the Shapiro and Wilk (1965) normality test now yields a P-value of 0.9057. Thus

[14] An observation as extreme as our lowest observation of monthly continously compounded return of -0.35 would occur approximately once every 315,000 years if our monthly return data in fact was normally distributed with mean and standard deviation like our sample.

we are nowhere close to reject the null hypothesis of normality. The Q-Q plot in Figure 7 confirms this conclusion. Our monthly return dataset hence moves from an empirical distribution clearly non-normal to a dataset looking normal by just removing eight extreme negative outcomes.

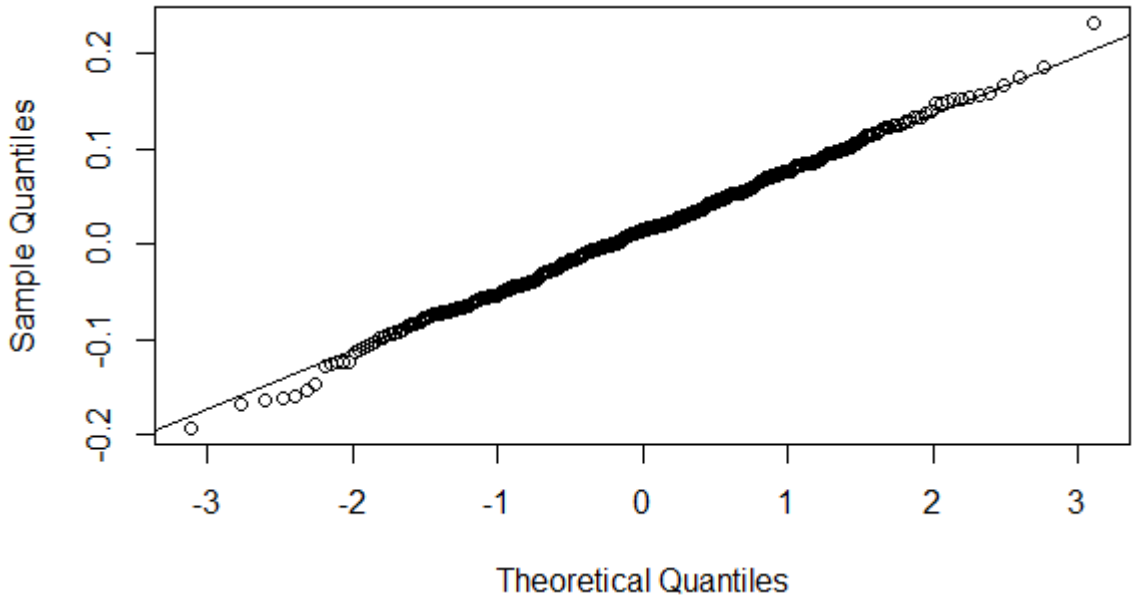

Figure 7: Normal Q-Q plot for monthly returns without outliers.

That we seemingly remove all non-normality from our data by removing the worst crises in the Norwegian stock market, suggests that an assumption of normality of returns on a monthly basis could be acceptable during normal times. The problem of making this assumption in general however is that occasionally extreme negative observations will occur, and these events are almost certainly ruled out if one assumes normality. A model of monthly returns expected to capture the risks involved in investing in the Norwegian stock market should therefore be able to capture these rare and extreme negative outcomes. We performed a similar exercise in order to try to explain non-normality in daily and quarterly return data, but just removing the most extreme outcomes did not change the conclusion of non-normality.

2.3.2 Skewness and kurtosis testing

As DeCarlo (1997) points out, univariate skewness and kurtosis tests can help pinpoint what type of properties that makes a particular set of observations deviate from normality. DeCarlo (1997) suggests that an informative way of testing for normality is to combine our approach in sub-

section 2.3.1, using the Shapiro and Wilk (1965) test and Q-Q plots, with skewness and kurtosis testing. The tests we apply in this thesis for univariate skewness and kurtosis is the one presented by D'Agostino, Belanger, and D'Agostino (1990). The purpose of these tests is to evaluate whether the skewness or kurtosis are significantly different from what one would find in a normally distributed dataset (D'Agostino et al., 1990).

The first test made in this section is a skewness test where the null hypothesis of normality is tested against the alternative hypothesis of non-normality due to skewness[15]. We apply a normal approximation of the test by D'Agostino et al. (1990), valid for datasets with more than eight observations. P-values for the skewness test in all of our four datasets are provided in Table 3. It is evident from the results in Table 3 that all datasets exhibit skewness that are significantly different from zero. The Z values in Table 3 are the test statistics that should be standard normally distributed under the null hypothesis of normality. We notice that the test results reject the null-hypothesis most strongly for daily returns.

The second test made in this section is a kurtosis test where the null hypothesis is that the data are normal, and the alternative hypothesis is that the data are non-normal due to non-normal kurtosis. We also apply a normal approximation to this test, valid for datasets consisting of more than 20 observations, by Anscombe and Glynn (1983). We see from Table 3 that we can reject the null hypothesis for all four datasets. The excess kurtosis is thus statistically significant. In this case too, we are most confident in rejecting the null hypothesis for daily data. The conclusion of this sub-section is thus that returns at all frequencies considered here deviates from normality due to both skewness and kurtosis.

[15] We apply two-sided tests for skewness and kurtosis in this section, so we have H_0: Normality, H_A: Non-normality due to skewness/kurtosis, without saying anything beforehand of which direction the skewness or kurtosis measure deviates from normality.

	Daily returns	Monthly returns	Quarterly returns	US monthly returns
Z in skewness test	-22.8	-6.42	-3.75	-5.93
P-value skewness test	5.413e-115	1.346e-10	0.00018	3.098e-09
Z in kurtosis test	36.8	5.82	3.62	5.92
P-value kurtosis test	0	5.997e-09	0.00029	3.314e-09

Table 3: Z and P-values in skewness and kurtosis tests.

3. Modelling of stock returns – a normal inverse Gaussian approach

We consider a baseline case for stock returns where they are assumed to be i.i.d. Gaussian. From this baseline, there are two possible extensions, as illustrated by the branches in Figure 8. The first is to allow for dependence of returns and volatility across time, by dropping the i.i.d. assumption and using for instance GARCH-models (Engle, 2002). The second possible extension is to drop the assumption of normality, by using more flexible i.i.d. Levy-processes. In this thesis we will consider, and investigate the implications of, the latter and only mention the first branch when discussing possible extensions to our modelling approach.

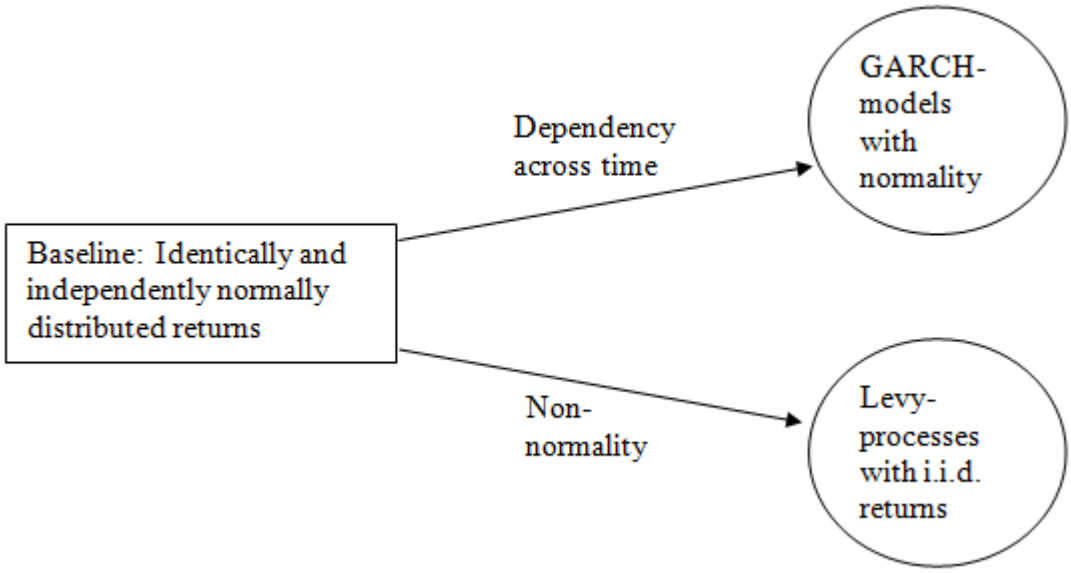

Figure 8: Decision tree for modelling of stock returns.

In the following we model continuously compounded returns as normal inverse Gaussian (NIG) distributed random variables. The NIG distribution was introduced to finance by Barndorff-Nielsen (1997), and it has since gained attention for its ability to fit financial data. In particular, the NIG distribution allows for skewness, heavy tails and peakedness, all of which are common features in most financial markets, including the Norwegian stock market, as seen in chapter 1. Another desirable quality of the NIG distribution, as opposed to certain other fat-tailed distributions like for instance the Cauchy distribution, is that a moment generating function exists and so therefore do all moments. An added advantage of the NIG distribution, along with its close relatives in the generalized hyperbolic family of distributions, is that there are readily available software packages (such as the package *ghyp* for R developed by Luethi and Breymann (2013)) that are able to handle complex calculations.

3.1 Theoretical model

Formally, we are considering the following model of returns

$$\log\left(\frac{S_t}{S_{t-1}}\right) := r_t \sim \text{NIG}(\alpha, \beta, \mu, \delta). \tag{9}$$

Loosely speaking, the parameter alpha is related to tail-heaviness, beta to symmetry, mu to location and delta to scale (Benth, Groth & Kettler, 2005). The normal distribution is obtained in the limit when alpha and delta goes to infinity, regardless of beta (Lillestøl, 1998).

The $\text{NIG}(\alpha, \beta, \mu, \delta)$ distribution has the following probability density function (Eriksson, Ghysels & Wang, 2009)

$$f_{\text{NIG}}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \mathrm{e}^{\delta\sqrt{\alpha^2-\beta^2}-\beta\mu} \; \frac{K_1\left(\alpha\delta\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}\right)}{\sqrt{1+\left(\frac{x-\mu}{\delta}\right)^2}} e^{\beta x}, \tag{10}$$

where $K_1(\cdot)$ denotes the Bessel function of the third kind with index 1 (see Abramovich and Stegun (1974) for descriptions of these types of functions).

The NIG distribution can also be written as a mean-variance mixture of the normal distribution and an inverse Gaussian (IG) distribution in the following way

$$r_t = \mu + \beta Z_t + \sqrt{Z_t} U_t \quad \text{where} \quad U \sim N(0,1) \text{ and } Z \sim IG\left(\delta, \sqrt{\alpha^2 - \beta^2}\right), \tag{11}$$

where Z and U are independently drawn from their respective distributions (Lillestøl, 1998).

It is important to note that we are not allowing for the possibility of time varying parameters of the NIG distribution in our specification of the model. This implies that we are in fact assuming both Z and U to be i.i.d., which is a questionable assumption given our results of volatility clustering and autocorrelations in returns from chapter 1. We can view our model of returns as a one-period model without prior knowledge of last period's return and volatility, i.e. an unconditional model of returns.

From specification (11), we see that Z has a close resemblance to the variance of returns. In fact, the marginal distribution of returns conditional on Z equal some z, is normal with variance z,

$$r_t | Z = z \sim N(\mu + \beta z, z). \tag{12}$$

This implies that controlling for Z should give us normally distributed returns, which is something we will explore further with our return data in section 3.3. We will henceforth refer to any realization of Z as the instantaneous variance. One may loosely think of (11) as a model of normally distributed returns with stochastic expectation and volatility.

From (11) and (12), it can easily be seen that a negative (positive) beta will give rise to negative (positive) skewness in the return series. To see this, consider a negative beta and two given values of Z – a high and a low value. In the case when Z is high, expected returns will be low and the conditional distribution will be symmetric with a high variance, i.e. it is fairly likely to end up with extremely low returns. When Z is low, expected return is high, but the conditional symmetric distribution will have a low variance, i.e. extremely high returns are quite unlikely. Similarly, it is easy to see that a beta equal to zero gives rise to a symmetric return distribution.

The first four moments of the theoretical return distribution of NIG type have the following expressions (Lillestøl, 1998)

$$E[r] = \mu + \beta \frac{\delta}{\gamma}, \qquad \text{Var}[r] = \delta \frac{\alpha^2}{\gamma^3}$$

$$\text{Skewness}[r] = 3\frac{\beta}{\alpha}\frac{1}{(\delta\gamma)^{1/2}}, \qquad \text{Kurtosis}[r] = 3\left(1 + 4\left(\frac{\beta}{\alpha}\right)^2\right)\frac{1}{\delta\gamma},$$

where $\gamma := \sqrt{\alpha^2 - \beta^2}$.

From the expression for kurtosis it is easily seen that even with a beta equal to zero the NIG distribution can match important aspects of the stock market, e.g. peakedness and heavy tails.

The first two moments of the IG distribution are (Barndorff-Nielsen, 1997)

$$E[Z] = \frac{\delta}{\gamma} \quad and \quad Var[Z] = \frac{\delta}{\gamma^3}.$$

We note that if β is small relative to α, then $\gamma \approx \alpha$. In this case, the mean of Z would be approximately equal to the variance of returns. We also note that whenever β is different from zero, the variance of returns is higher than the expected value of Z – the extra variance of returns is coming from the effect of βZ on the expected return in (12). Note also that the expectation of Z and the variance of returns are both proportional to the scale parameter, δ.

As mentioned already, when returns are modelled as a mean-variance mixture of the normal and the IG distribution, it is in fact a stochastic volatility model of returns. This specification is able to handle the non-normality we see in returns, but it is not able to model volatility clustering. The reason is that our return distribution is time-invariant. One way to deal with this issue is to make use of the proportionality of the variance of returns with respect to delta. Andersson (2001) does this by allowing a time-dependent structure of δ in the model, more specifically

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

Andersson (2001) calls this model the normal inverse Gaussian stochastic volatility (NIGSV(p,q)) model. In addition to capturing the non-normality in the data, this model is also able to capture the volatility clustering.

3.2 Fitted return distribution

In this section, we estimate the parameters of the NIG distribution by fitting it to the sample distribution of monthly returns, using the maximum likelihood estimation (MLE) in the R-package *ghyp* developed by Luethi and Breymann (2013). Maximum likelihood estimations solve the following optimization problem.

$$\max_{\boldsymbol{\theta}} \mathrm{L}(\boldsymbol{\theta}; x_1, x_2, \dots, x_n), \tag{13}$$

where $\mathrm{L}(\theta; x_1, x_2, \dots, x_n)$ is the likelihood function of a model, representing the likelihood of the model generating the data points $\boldsymbol{x} = x_1, x_2, \dots, x_n$ when applying the (vector of) parameters, $\boldsymbol{\theta}$. An equivalent, and often simpler, approach is to rather maximize the log-likelihood:

$$\max_{\boldsymbol{\theta}} \log[\mathrm{L}(\boldsymbol{\theta}; x_1, x_2, \dots, x_n)],$$

which can be written as a sum of the log-likelihood of obtaining each data point $x_i \forall i = 1,2, \dots n$,

$$\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log[f_x(x_i; \boldsymbol{\theta})].$$

Here $f_x(x_i; \boldsymbol{\theta})$ represents the probability density function of a given distribution. In the case of NIG, the probability density function is given in equation (10). Loosely speaking, the MLE finds the particular parameter values of a distribution that maximizes the likelihood of obtaining a given set of observations. In the R-package *ghyp* by Luethi and Breymann (2013), a modified expectation-maximization (EM) approach (the multi-cycle, expectation, conditional estimation (MCECM) algorithm, see McNeil, Frey and Embrechts (2005)) is used with an augmented likelihood function, but the intuition provided here still applies.

One obvious issue when estimating a return distribution is that we do not observe returns in different states of the economy – we only observe a realized time series, which we take as a proxy for returns across states. For this approach to be valid, we need all the observed returns to be independent realizations of the same underlying distribution. As mentioned in section 3.1, this is a problematic assumption given our findings of significant autocorrelation in returns and volatility. For our purposes the perhaps most problematic consequence of the i.i.d. assumption is that we are not using all the information available at time t when forecasting the time t+1 return, e.g. we probably could get a more accurate estimate of the time t+1 return distribution by taking

into account our knowledge of time t volatility. Forsberg and Bollerslev (2002) suggest that a combination of a return distribution of NIG type and a GARCH model for volatility might be a good model for financial data that also takes into account the information embedded in realized volatility[16].

Our estimated NIG model for returns is shown in Figure 9 (for daily, quarterly and US monthly returns see Figure A6, A7 and A8 respectively in Appendix A) as a solid red curve together with a Gaussian distribution as a dotted curve, for comparison. We see that the fitted NIG is able to capture quite a bit of the peakedness and some of the fatness in the left tail. We also note that the fitted NIG distribution is slightly negatively skewed, as we also found to be a feature of the return data in chapter 2.

The Q-Q plot in Figure 10 is interpreted as follows. The triangular points are data points that should be compared to the theoretical Gaussian quantiles given by the dotted line. Similarly, the circular points are data points that should be compared to the theoretical NIG model given by the solid line. Figure 10 tells us that the fitted NIG model seems to match the empirical distribution nicely, perhaps with the exception of the left tail, which is even fatter in reality than what our model is able to predict (we will address this finding further for daily returns in chapter 4).

We have also calculated the Akaike information criterion (AIC) for the fitted NIG and Gaussian distributions (Akaike, 1974). The AIC is defined as

$$\mathrm{AIC} = 2k - 2\log(L), \tag{14}$$

where k is the number of parameters and L is the likelihood value of the model (attained from the optimal solution to problem (13)). A high AIC means that the model performs poorly and a low AIC implies that the model performs well. Note that AIC punishes the inclusion of more parameters by the principle of parsimony. Burnham and Anderson (2004) suggest using AICc[17], which adjusts AIC for the number of observations. For a large number of observations, however, like we have (in this respect) at all frequencies, the AICc converges to AIC, which is why AIC is

[16] This is a similar approach to the NIGSV(p,q) of Andersson (2001) outlined in section 2.1.
[17] $\mathrm{AICc} = \mathrm{AIC} + 2\mathrm{k}(\mathrm{k}+1)/(\mathrm{n}-\mathrm{k}-1)$, where n is the number of observations.

a sufficient alternative to AICc in this thesis[18]. Based on the AIC score, the NIG model[19] clearly outperforms the Gaussian model[20].

A question to be addressed regarding model selection using AIC is whether we can say if a more complicated model significantly outperforms another model. For this purpose we apply the likelihood ratio (LR) test (used as an addition to AIC), which uses the following test statistic (Lütkepohl, 2006),

$$\lambda_{\mathrm{LR}} = 2 * [\log(\mathrm{L}[\boldsymbol{\theta_{UR}}; \boldsymbol{x}]) - \log(\mathrm{L}[\boldsymbol{\theta_R}; \boldsymbol{x}])].$$

Here $\boldsymbol{\theta_{UR}}$ and $\boldsymbol{\theta_R}$ denotes the optimal parameters from two MLEs on an unrestricted (UR) and a restricted ® version of two nested models[21], respectively. Like in problem (13), $\boldsymbol{x}$ is the vector of data points. Under the null hypothesis, which states that the restricted model is the data generating process, $\lambda_{\mathrm{LR}} \sim \chi^2(v)$, with degrees of freedom v equal to the number of parameters in the unrestricted model less the number of parameters in the restricted model (Lütkepohl, 2006). We use the R-package *ghyp* (Luethi & Breymann, 2013) to apply the chi-squared distribution, $\chi^2(v)$ (valid under H_0), in order to get P-values for the likelihood ratio test. A low P-value implies rejection of H_0 and a conclusion that the unrestricted model significantly outperforms the restricted one.

Using the likelihood ratio test on our NIG model (with parameters $\boldsymbol{\theta_{UR}}$) and Gaussian model (with parameters $\boldsymbol{\theta_R}$) estimated in this section we get an extremely low P-value[22], suggesting that our NIG model significantly outperforms the Gaussian model.

[18] Using AICc instead of AIC in all analyses of this thesis does not alter any conclusions.
[19] $\mathrm{AIC(NIG)} = 2 * 4 - 2 * 676.76 = -1345.52.$
[20] $\mathrm{AIC(Gaussian)} = 2 * 2 - 2 * 656.51 = -1309.02.$
[21] Nested models are defined as two models where the first one (the unrestricted, complex one) can be transformed into the second one (the restricted, simple one) by imposing constraints on the parameters of the first one.
[22] The test provides a P-value$= 1.6 * 10^{-9}$, suggesting NIG having a substantially better fit than the Gaussian.

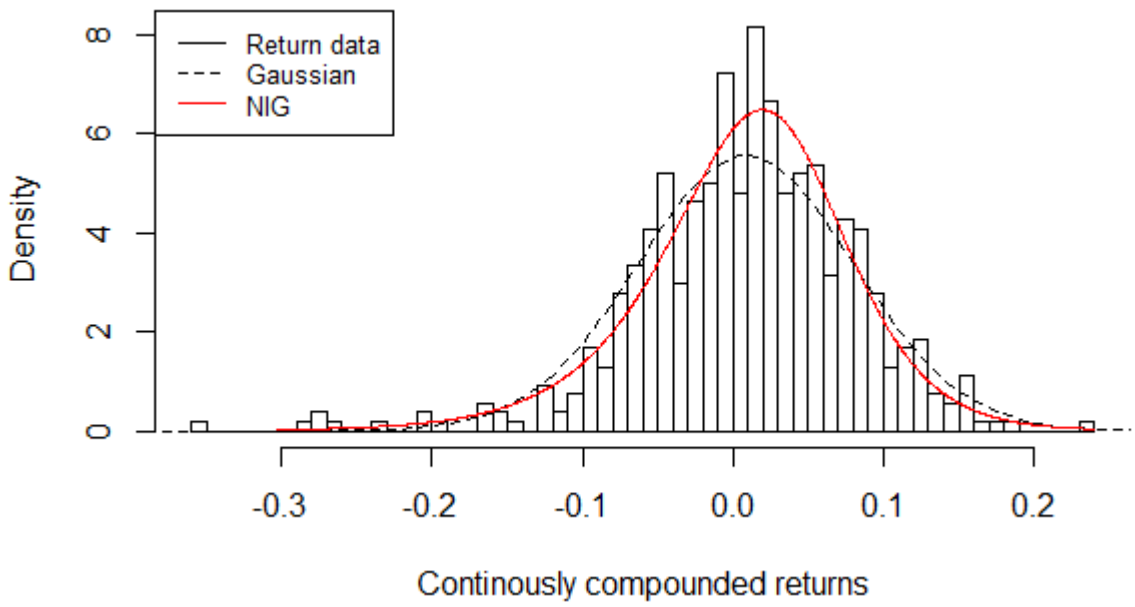

Figure 9: Distribution of monthly returns with NIG-fit.

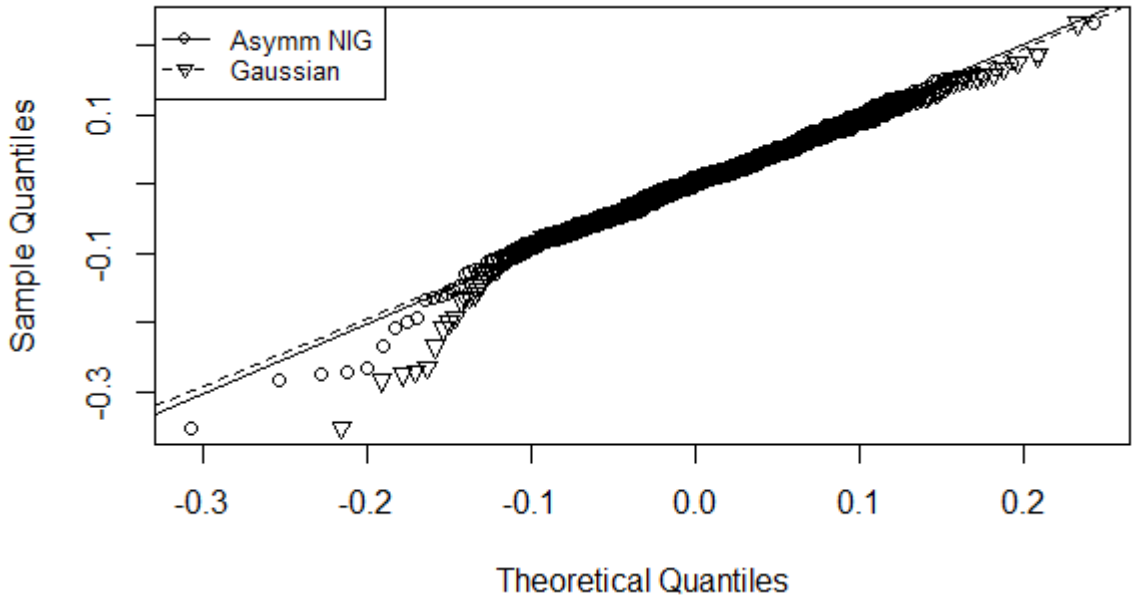

Figure 10: Generalized hyperbolic Q-Q plot for monthly returns.

Table 4 shows that both the expected value and the variance of returns match the empirical counterparts in Table 1 closely, but we are somewhat underestimating the negative skewness, and the sample kurtosis is a bit higher than our model kurtosis. The NIG distribution could of course match all these four moments perfectly, seeing as it has four free parameters. The reason for why

this is not the case here is as mentioned earlier that we are using the maximum likelihood estimation (as opposed to a method of moments approach), which takes into account that our empirical distribution is not fully described by its first four moments.

Table 5 shows our parameter estimates. Of particular interest is the negative beta, which gives us a negative relationship between expected returns and variance (in line with what we observed in our monthly return data in chapter 2).

	Monthly NIG-fit
Expected value	0.00865
Variance	0.00506
Skewness	-0.4764
Kurtosis	4.5475

Table 4: Moments for NIG-fit of monthly returns.

	Monthly NIG-fit
α	23.239
δ	0.1070
β	-5.7292
μ	0.0359
γ=(α^2-β^2)^0.5	22.521

Table 5: Parameter values for monthly NIG-fit of returns.

3.3 Return distribution conditional on instantaneous variance

From the expression in (12), we know that the distribution of returns conditional on the instantaneous variance should be normal if the true return distribution is indeed NIG. In this part, we will try to control for the instantaneous variance in our return data and then test whether the resulting distribution is normal. Our approach is to split the data set into several sub-datasets, where each sub-datasets consist only of observations with instantaneous variances between

certain levels. We will use the variance from expression (7) in sub-section 2.1.2 as a proxy for the instantaneous variance of our return series. This approach allows us to reduce the variability of the variance, but it is not a perfect control for the instantaneous variance. Even with such a rough approach however, the conclusion is that we are able to get quite far.

As mentioned in chapter 2, monthly returns are normal except for eight observations in the extreme negative territory. This implies that we really do not have enough non-normality[23] for our approach to be of any use for monthly data. In this section, we will therefore use daily data, where the non-normality is much more severe and, importantly, where there are many more observations readily available. The dataset for daily returns and variances consist of 10 823 observations. We split this dataset into 20 subsets – subset 1 consisting of the five percent observations with the lowest variance and subset 20 the five percent observations with the highest variance.

Table 6 shows descriptive statistics and test results for the three least normal subsets and the most normal subset (measured by their corresponding P-values in the Shapiro and Wilk (1965) test). As a control set, we have also included a subset of the same size of randomly chosen observations from the whole dataset. Figure 11, 12 and 13 show Q-Q plots for the randomly drawn subsample, the most normal subsample and the least normal subsample, respectively. In total, we cannot reject normality for 11 of 20 subsets on a five percent significance level. All the 9 remaining subsets have a significantly higher P-value from the normality test than the random control set. It is worth mentioning that the least normal set (subsample 20) has one extreme negative observation and if we remove this, the P-value rises to 0.5%.

Furthermore, the skewness and kurtosis is significantly lower for all subsets (except perhaps subsample 20) compared to the control set. From the tests explained in section 2.3.2, (expectedly) neither kurtosis nor skewness is significantly different from what one would expect from a normal distribution in the 11 subsamples with P-values above five percent in the Shapiro and Wilk (1965) normality test. For the remaining 9 subsamples, kurtosis and/or skewness are

[23] All of these eight observations end up in the same high volatility subsample, making all other subsamples normal, and this sample non-normal. This implies that we are not really testing the effect of controlling for the variance on normality.

significantly different from a normal distribution on a five percent level (and hence the source of non-normality is varying).

We see that our method of controlling for the instantaneous variance has reduced the volatility of volatility[24] significantly for all subsets, except subsample 20, compared to the random subsample. Herein lies also some of the problems with our approach – the dispersion in volatility is much higher in the highest variance subset compared to the medium and low variance subsets, the reason being that there are relatively few high-variance observations in our dataset. Another issue with our approach is that the variance series is too smooth[25] to capture the "real" volatility in the most extreme observations. Overall we conclude that our calculated variance series using expression (7) is a proxy, albeit not a perfect one, for Z.

[24] Defined as the standard deviation of the square root of our variance series.

[25] The variance series is calculated as the variance of returns of the 30 days surrounding any given day (see sub-section 2.1.2). This implies that the calculated variance series will never jump abruptly. When using this variance series we are therefore not controlling for the instantaneous variance, but rather the volatility environment, i.e. whether volatility (here defined as $\sqrt{Z}$) is high or low in the time period surrounding the observation. This approach does not make any sense if our i.i.d. assumption holds, as this would imply that high volatility in one period would not in general be surrounded by other high volatility observations. Our smoothed volatility series would in this case be random noise around a constant. As noted in chapter 2 however, we know that there is significant autocorrelation in volatility (particularly in daily data), thus controlling for the volatility environment could be a useful approach. The failure of the i.i.d. assumption is however, as pointed out in section 3.2, problematic in itself – we need the observations in the time series to be independent if we are to use it as a representation for the return distribution at any particular point in time, i.e., as a representation of the distribution across states as opposed to across time. The only justification we have is that the time series is quite long compared to the half-life of the autocorrelation of volatility, which should give us a time series that is approximately independent as a whole.

	Random sample	Subsample 7	Subsample 10	Subsample 15	Subsample 20
Average volatility	0.0131	0.0095	0.0111	0.0146	0.0332
Volatility of volatility	0.00659	0.00014	0.00017	0.00030	0.01147
Volatility interval	[0.0052, 0.0631]	[0.0093, 0.0098]	[0.0107, 0.0113]	[0.0140, 0.0150]	[0.0238, 0.0645]
P-value in normality test	2.821e-16	0.9786	0.0004	0.0043	3.365e-09
Skewness	-0.6991	0.0514	-0.1536	-0.2174	-0.4648
Kurtosis	9.0586	2.9763	4.2326	3.8483	6.4514
Z in skewness test	-6.12	0.49	-1.47	-2.07	-4.27
Z in kurtosis test	8.54	0.04	3.94	3.08	6.85
N	541	541	541	541	541

Table 6: Properties of subsamples with tests for normality, skewness and excess kurtosis.

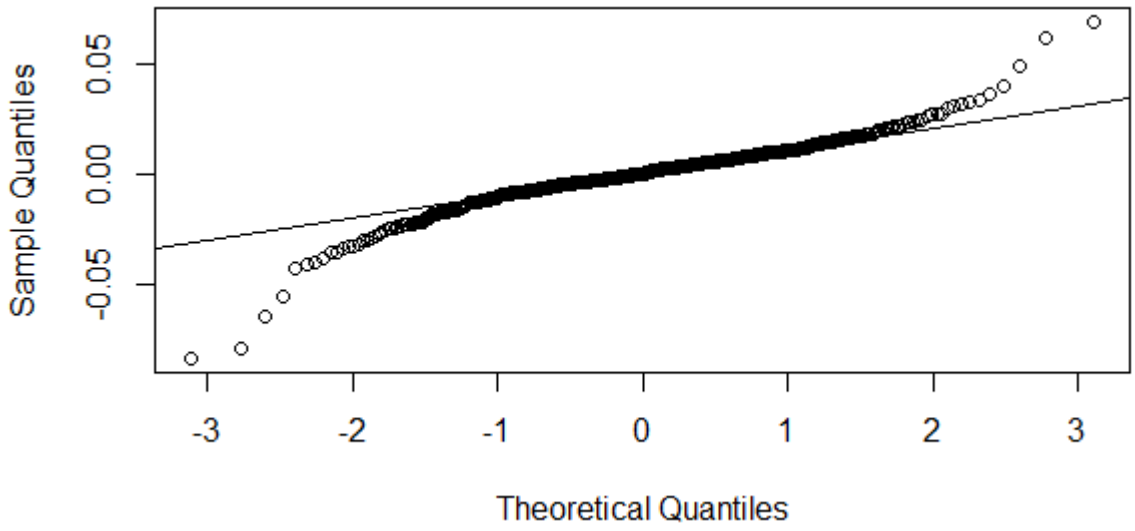

Figure 11: Normal Q-Q plot of random sample.

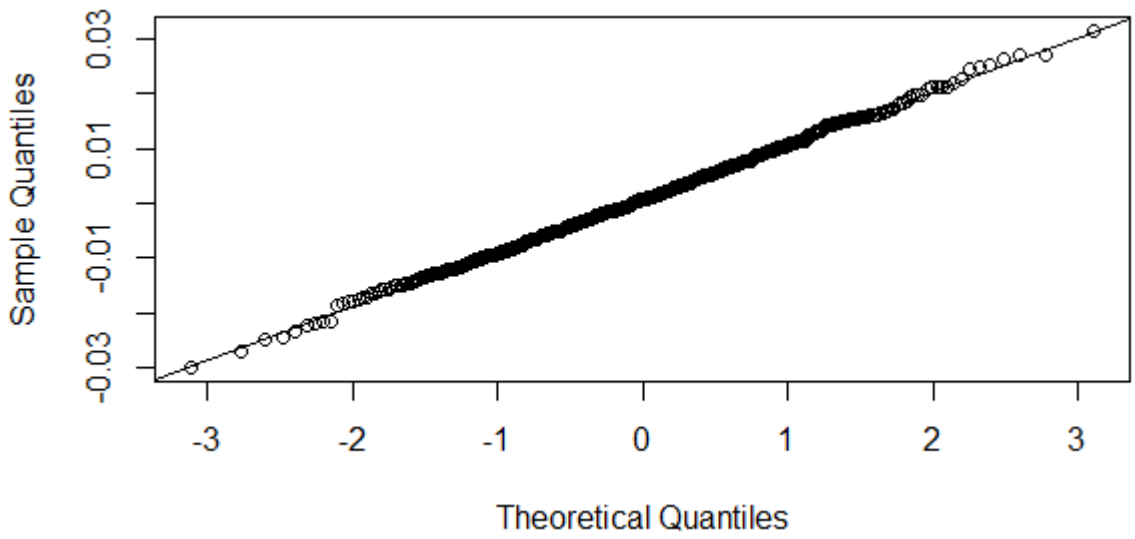

Figure 12: Normal Q-Q plot of subsample 7.

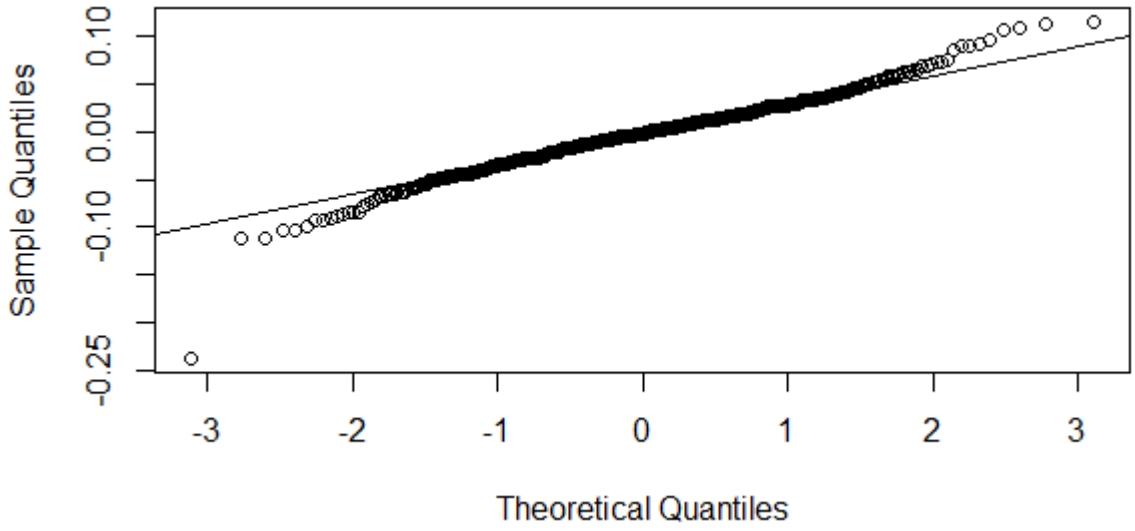

Figure 13: Normal Q-Q plot of subsample 20.

3.4 Fitted variance

Section 3.3 gave us an indication that our variance series can in fact be a proxy for the variable Z in the mean-variance mixture of the IG and normal distribution for daily returns. In this part we will therefore model our variance of returns series from (7) as an IG distribution (using the R-package *STAR* by Pouzat (2012)), going back to monthly observations. This will give us estimates of the parameters delta and gamma, which we in turn will use when fitting the NIG distribution to the return series. A particularly interesting approach would have been to estimate

an independent variance series, e.g. constructed from implied volatilities or a volatility index, and thereby nail down some of the free parameters of the NIG distribution for the return series. Such a volatility series would arguably be a better estimate of the instantaneous volatility, seeing as it is forward-looking, able to respond quickly to changes in the "true" volatility, and because it would typically be the volatility that makes an option price based on normality assumptions correct. A similar approach is pursued by Benth et al. (2005), where they develop an option-pricing model based on the NIG distribution and use this model to find an implied value for the parameter delta. In this paper however, we will use the variance series for monthly return data constructed from a 12-month moving standard deviation of returns (see sub-section 2.1.2).

Figure 14 illustrates that the IG distribution matches the sample distribution of variance from (7) well (we get similar results for daily and quarterly data, see Figure A9 and A10 in Appendix A). From Table 7 we see that the estimates for delta and gamma have both decreased relative to what we got in section 3.2. This implies that the resulting fitted NIG distribution of returns here will be less normal than what we got in section 3.2.

The results we get from our constructed variances from (7) are, however, likely to overestimate both gamma and delta relative to what we would get by using a variance series that responds faster to changes in the volatility. The reason is that our volatility series probably understates the highest volatility and overstates the lowest volatility – these observations are "averaged out". To get a feeling for how our results might change for another proxy of the variances, we also estimate an IG distribution for a series of residuals squared (the residuals squared are calculated using expression (8)). In this case, the estimate for gamma decreases to 0.3561 and the estimate for delta decreases to 0.00184, which is qualitatively in line with what we expect.

We now turn to estimating the NIG distribution for returns, but we keep gamma and delta fixed at their respective values calculated from the variance distributions.[26] All resulting parameter values in the case of IG distributed variance calculated as in (7) are reported in Table 7. Figure 15 shows the result of the three different ways of estimating the NIG distribution mentioned in this chapter, with the corresponding Gaussian distribution for comparison.

[26] In R using *ghyp*, we do this by a change of parametrization, keeping alpha bar (from the standard parametrization of NIG in the R-package, see Luethi and Breymann (2013)) fixed to the value implied by delta and gamma.

Table 8 provides the AIC values for the four models estimated in this chapter, calculated using equation (14). The AIC scores do not provide a clear indication that modelling NIG directly on the return data outperforms modelling return indirectly by pre-fitting an IG distribution to our variance series from expression (7). Using the likelihood ratio test approach from section 3.2 we cannot reject the null hypothesis that the two models fit the data equally well[27]. We do however notice in Table 8 that the approach of modelling residuals squared (from expression (8)) as IG, is generating a NIG distribution which performs much worse than even the Gaussian distribution.

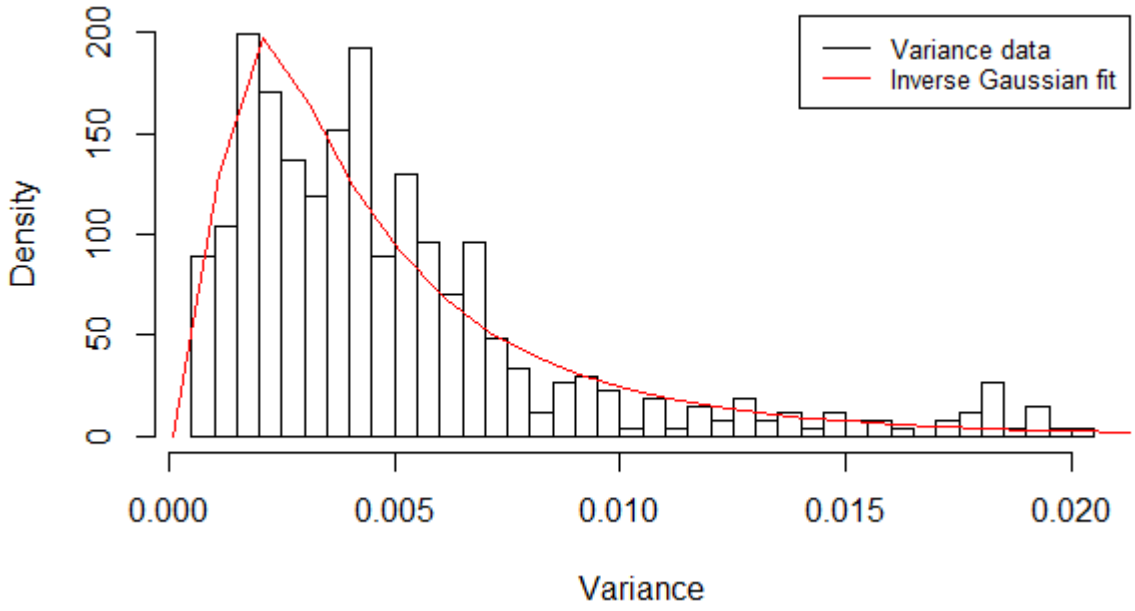

Figure 14: Monthly variance data and inverse Gaussian fit.

[27] We are testing whether our unrestricted model in 3.2 with 4 free parameters outperforms the model of section 3.4 where one of the parameters are pre-estimated through IG distributed variance (calculated using (7)). This test provides a P-value= 0.1423, and we cannot significantly discriminate between the two.

	Monthly NIG-fit with IG-estimated variance
α	16.956
δ	0.0847
β	-3.5649
μ	0.0268
γ=(α^2-β^2)^0.5	16.577

Table 7: Fitted NIG distribution after fitting variance as inverse Gaussian.

	Monthly NIG-fit	Monthly NIG-fit with IG-estimated	Monthly NIG-fit with IG-estimated residuals squared	Gaussian distribution
AIC	-1345.52	-1343.36	-1196.63	-1309.02

Table 8: Akaike (1974) information criterion (AIC) for the four models estimated in this chapter.

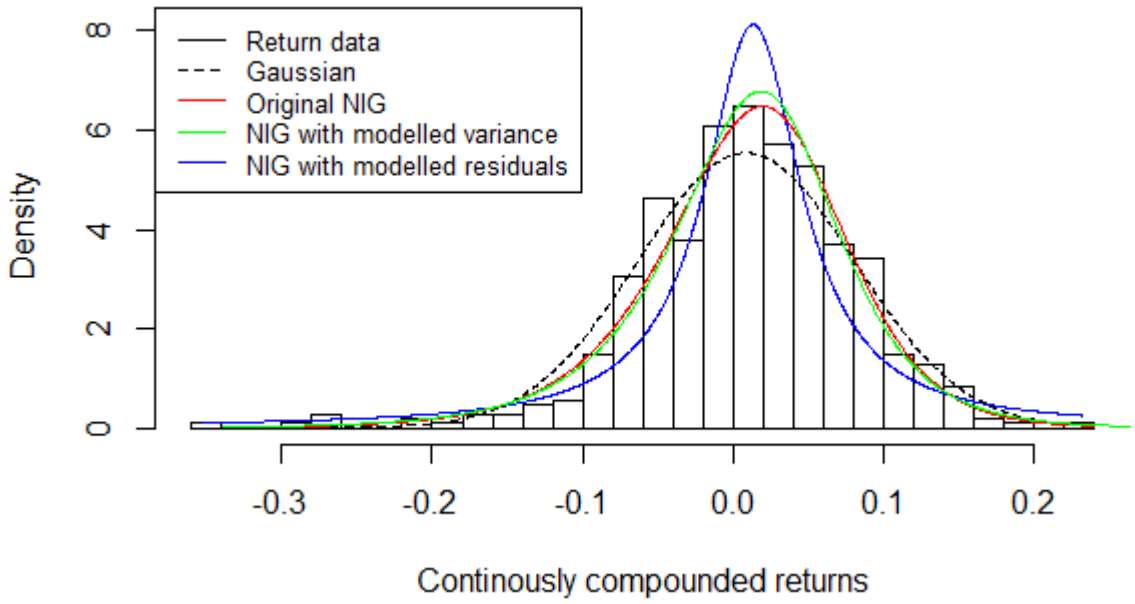

Figure 15: NIG distributions corresponding to our different ways of modelling variance.

3.5 NIG-triangle

In Table 9, we have calculated the measures steepness and asymmetry, which are alternative measures to skewness and kurtosis. The measures were suggested by Barndorff-Nielsen, Blæsild,

Jensen, and Sørensen (1985) and provide a way to compare steepness and asymmetry across different NIG models. Additionally, these measures can give an indication of how a particular model deviates from normality. The steepness parameter is given by

$$\xi = \left(1 + \delta\sqrt{\alpha^2 - \beta^2}\right)^{-\frac{1}{2}} \tag{15}$$

and the asymmetry parameter is given by

$$\chi = \frac{\beta}{\alpha}\xi. \tag{16}$$

As Barndorff-Nielsen and Prause (1999) points out the domain of variation for (χ, ξ), which is known as the NIG triangle in the literature because of its shape if graphed, is

$$\{(\chi, \xi) : -1 < \chi < 1, \qquad 0 < \xi < 1\}. \tag{17}$$

The pair $(\chi, \xi) = (0, 0)$ corresponds to the normal distribution (Barndorff-Nielsen & Prause, 1999). Barndorff-Nielsen and Prause (1999) also points out that the pair $(\chi, \xi) = (0, 1)$ corresponds to the Cauchy distribution. We include these special cases in Table 9 for comparison. Figure 16 plots the NIG-triangle with all the univariate models for return estimated in this paper.

NIG-triangle	Steepness (ξ)	Asymmetry (χ)
Theoretical Gaussian	0	0
Cauchy distribution	1	0
Daily (1972-2015)	0.774	-0.029
Daily (1972-1993)	0.778	0.033
Monthly (1970-2014)	0.542	-0.134
Monthly with pre-estimated variance	0.645	-0.136
Monthly with pre-estimated residuals	1.000	-0.995
Quarterly (1970-2014)	0.675	-0.179
US monthly (1970-2014)	0.588	-0.163

Table 9: NIG-triangle with benchmarks and all our different univariate NIG models.

From Table 9 we note that all estimated models for stock market returns in this paper has a negative skewness, except for the daily model for 1972-1993 used for evaluating NIG as a model out of sample in chapter 4. We do however, see that the degree of asymmetry seem to be lower, and maybe even negligible, for our models of daily return than for monthly and quarterly returns. Another general observation is that all our return models exhibit more steepness than the normal distribution. Daily returns exhibit more steepness than monthly and quarterly data, which means that it has more weight in its center and its tails relative to its shoulders compared to lower frequency data. With one exception the steepness parameter in all models lies in the neighborhood 0.5-0.8 indicating substantial non-normality. These levels are consistent with previous findings for financial return data (Bølviken & Benth, 2000; Venter & de Jongh, 2002). The model where we have fitted the residuals squared to an IG distribution and used our estimates to find an implied NIG distribution, is an extreme case where the distribution is as asymmetric and steep as possible in this framework. We also note that the US return model is slightly more asymmetric and steep than the Norwegian monthly return model, thus further away from a Gaussian model, according to these measures.

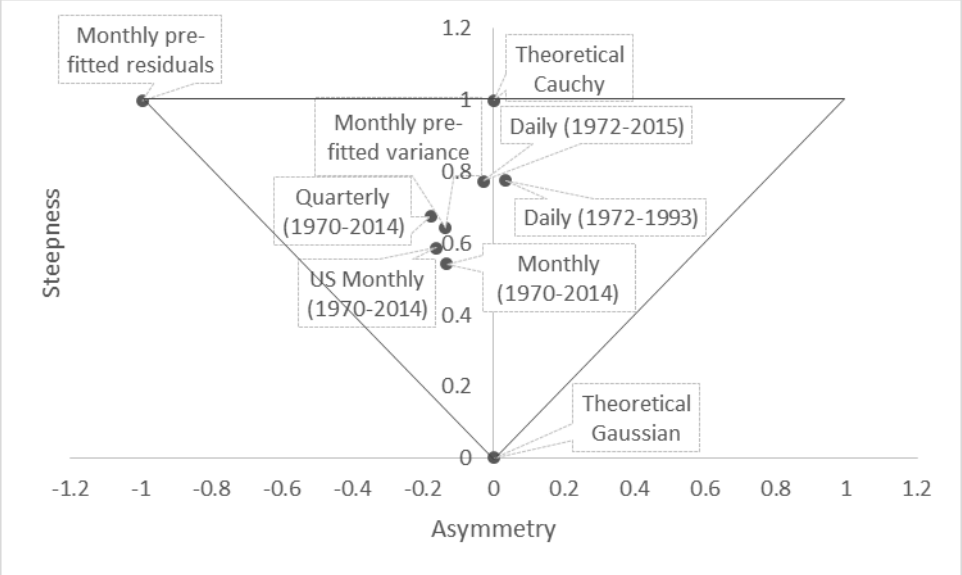

Figure 16: NIG-triangle with all estimated univariate models.

3.6 Parameter uncertainty in our NIG-fit – parametric bootstrapping

In this section, we evaluate the parameter uncertainty in the monthly return model of section 3.2. The method applied in this section can be used in other models as well, and even on different measures such as for instance the asymmetry and steepness parameters of section 3.5.

The approach we use in this section is a parametric bootstrap method as presented by Efron and Tibshirani (1993). We consider the model in Table 5 estimated on our original monthly return dataset to be the data generating process of monthly returns. The bootstrap then uses this model to generate k new datasets (each with identical size to the original dataset) of randomly drawn observations. For each newly generated dataset the maximum likelihood estimation is performed again. This gives us k estimates for each parameter in the model, which we in turn can use to calculate descriptive statistics, such as standard errors, for these parameters.

To illustrate the results that a bootstrap approach yields, we have provided a histogram of the different estimates of alpha from a bootstrapping taking $k = 1000$, in Figure 17. It is evident from Figure 17 that we have some observations in the right tail where alpha is extremely large. In these cases, the model is close to normal, as we pointed out in section 3.1. The observations with extremely high alphas in Figure 17 are bootstrap samples where extreme tail-observations do not occur among the randomly simulated NIG-variates.

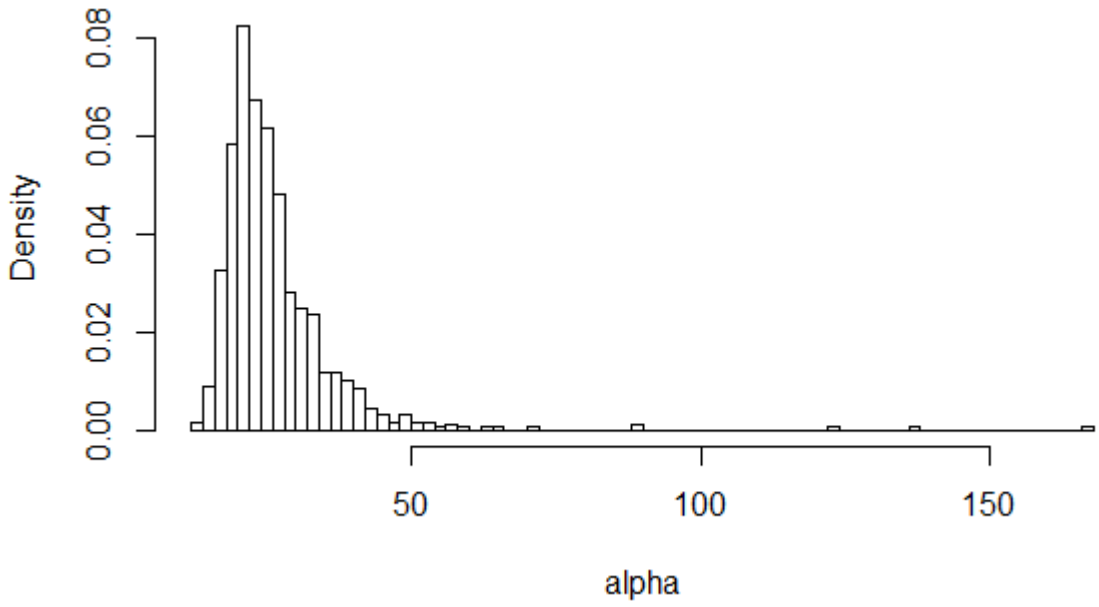

Figure 17: Distribution of alpha in parametric bootstrapping.

In Table 11, we provide the original parameter estimates, the standard error of the estimated parameters (calculated using expression (18)) and its minimum and maximum value. The uncertainty about the parameters is quite substantial. An important point made by Aase and Lillestøl (2015) is that the estimated parameters in Table 11 can be correlated. This correlation is not reflected in the bootstrap results in Table 11. In Table 10, we provide a correlation matrix of how the parameter estimates correlates over the 1000 randomly generated bootstrap samples. We note from Table 10 that α and β for instance are almost perfectly negatively correlated in the estimations. This implies that the bootstrap estimates with high α in Table 11 is corresponding to low values of β, possibly making the standard errors of the two biased upwards.[28]

	α	δ	β	μ
α	1	0.66	-0.99	0.91
δ	0.66	1	-0.57	0.75
β	-0.99	-0.57	1	-0.88
μ	0.91	0.75	-0.88	1

Table 10: Correlation matrix for parameter estimates.

	Original NIG-estimate (Table 5)	Standard error of bootstrap estimates	Minimum of bootstrap estimates	Maximum of bootstrap estimates
α	23.239	10.334	12.304	167.08
δ	0.1070	6.8205	0.0710	179.34
β	-5.7292	24.032	-600.79	55.656
μ	0.0359	0.0381	-0.0026	0.9162

Table 11: Results of bootstrap on NIG model for monthly returns.

$$\text{Standard error} = \sqrt{\frac{1}{540}\sum_{i=1}^{541}(X_i - \bar{X})^2}\,, \tag{18}$$

where $X = \alpha, \delta, \beta, \mu$.

[28] One approach to avoid unnecessarily high standard errors of two parameters that are almost perfectly correlated would be simply to fix one of the parameters at a reasonable level, and then estimate the remaining parameters.

Now that we have an impression about the uncertainty around each parameter of our baseline model – the NIG model directly estimated on monthly returns in section 3.2 – it would be interesting to compare the different models we have estimated. The NIG fit with IG-estimated variance in Table 7 have parameter estimates that somewhat differs from what we got in our baseline model in Table 5. When we consider the uncertainty of our parameter estimates, however, we see that this model is well within a 95% confidence band. The modelling approach where we model residuals squared as IG however, provides by comparison an α and δ lower than even our minimum bootstrap estimate of the two, and outside these two parameters' confidence band[29]. This suggests that modelling variance (from (7)) rather than residuals squared (from (8)) as inverse Gaussian is more reasonable[30]. This conclusion is also consistent with the AIC values of the different models provided in Table 8 of section 3.4[31].

4. Risk measures and NIG

In this part of the paper, we will turn to a practical application of the NIG distribution in risk measurement. The focus of our analysis will be on two common risk metrics: value at risk (VaR) and expected shortfall (ES). All the analysis will be performed on daily ex dividend returns, as opposed to the previous chapter, which mainly focused on monthly cum dividend return data. The risk measures are based on a long position in the underlying stock index. It is important to clarify the position of the investor because of the skewness in our return distribution. VaR and ES would be less extreme in a short position than in the long position due to the negative skewness.

4.1 Value at risk and expected shortfall – a short introduction

Value at risk is defined as the loss that is exceeded over a given time interval with p percent probability. Typically p is chosen between 0.1% and 5% and the time interval is usually somewhere between 1 and 14 days (Duffie & Pan, 1997). More formally, VaR is implicitly given by the following expression

$$\text{Prob}(X \leq x) = p, \quad \text{where } x := \text{VaR}. \tag{19}$$

[29] In the model where residuals squared are fitted as IG we get $\alpha = 0.17$, $\delta = 0.04$, $\beta = -0.17$ and $\mu = 0.01$.
[30] Variance and residuals as defined by (7) and (8), respectively.
[31] The comparison here only being dependent on L in (14) since both models have the same number of parameters, k.

One of the main issues with VaR is that it says nothing about how bad the losses get, conditional on times being bad. Expected shortfall is one approach to deal with this issue. It is defined as the expected loss conditional on the loss being greater than VaR. Mathematically, ES is given by

$$\text{ES} := E[X|X \leq x]. \tag{20}$$

Both VaR and ES are in other words looking at the left tail of the return distribution, which we have discussed in chapter 3 to be quite sensitive to distributional assumptions. The distributional assumptions will be our focus in this thesis. Other issues with VaR and ES, such as incentives for diversification and fire sales during bad times, will be pursued elsewhere. For instance, Duffie and Pan (1997) provide a nice overview of VaR.

4.2 Estimated value at risk and expected shortfall

We present value at risk and expected shortfall for daily returns from 1972 to 2015. Sample VaR is calculated using type 1 by Hyndman and Fan (1996)[32]. From Table 12 we see that the NIG distribution fits the sample distribution nicely in terms of VaR, with perhaps the exception of the 0.1 percent level where the sample VaR is more than 1.6 percentage points lower than our NIG model suggests. For all values of *p*, the NIG distribution performs better than the normal distribution. For all values of *p* lower than 2.5 percent, the normal distribution underestimates VaR, whereas it overestimates it at a 5 percent level, compared to the sample.

In Table 13, we have reported the values for daily ES. The sample ES is calculated by taking the average of the return beyond VaR at different levels of p[33]. We see that NIG performs better than the normal distribution, and that the normal distribution consistently underestimates ES, for all reported levels of *p*. It is also worth noting that even the NIG distribution fails to capture the severity of the worst outcomes (represented by ES at the lowest values of *p*).

[32] It is calculated using the *stats* package in R (R Core Team, 2014).
[33] The VaR and ES for the two models are calculated by using the *ghyp* package by Luethi and Breymann (2013).

Levels of p	Sample VaR	NIG model VaR	Normal model VaR
0.1%	-0.0879	-0.0714	-0.0455
0.5%	-0.0518	-0.0498	-0.0379
1%	-0.0423	-0.0411	-0.0342
2.5%	-0.0294	-0.0303	-0.0287
5%	-0.0220	-0.0227	-0.0241

Table 12: Value at risk for daily models and data.

Levels of p	Sample ES	NIG model ES	Normal model ES
0.1%	-0.1112	-0.0858	-0.0496
0.5%	-0.0732	-0.0633	-0.0425
1%	-0.0597	-0.0541	-0.0392
2.5%	-0.0446	-0.0426	-0.0343
5%	-0.0349	-0.0343	-0.0303

Table 13: Expected shortfall for daily models and data.

4.3 Out of sample estimates

In this section, we will try to get a feel for how the NIG distribution performs compared to the normal distribution as a model of returns out of sample. As mentioned in section 3.1, the NIG distribution has four free parameters, and the normal distribution obtains as a limiting case. This of course implies that we can never do worse, in the sense of fit, within sample with a NIG distribution than a normal distribution. This is the reason why we needed to punish the inclusion of new parameters in order to evaluate the models against each other in chapter 3.

In many practical applications, we are not so much concerned about in-sample fit, but rather how well a model can predict future return distributions. This is particularly the case when it comes to risk management and measurement. In this section, we split the data set into two subsets of exactly equal size – one estimation set (pre 1993) which we use to estimate the NIG and normal distribution and one testing set (post 1993) which is used to test the two models out of sample.

This approach allows us to get a feel for whether the results we got in the previous section is sensitive to the data sampling, which could be an issue in many practical applications of a model[34]. Testing the models out of sample also allows us to see whether we are over-fitting[35] the data using our models.

Figure 18 is a graphical representation of our results out of sample. It seems to be the case that the out-of-sample NIG distribution match the sample distribution closely, at least when compared to the out-of-sample Gaussian distribution. From Table 15 we see that the first two moments are quite similar for the normal and the NIG distribution while the third and fourth differ. The skewness is slightly positive for the out of sample NIG, whereas the sample skewness for post 1993 is negative. The positive skewness before 1993 is related to the positive correlation between volatility and returns observed before 1984, as we pointed out in chapter 2.

Tables 16 and 17 give us the out-of-sample estimates of VaR and ES respectively. Note that the NIG distribution performs better than the normal distribution for all reported significance levels. This is somewhat surprising, considering the positive skewness in the NIG model, but is explained by the NIG model capturing the high kurtosis in the data.

We have also calculated the AIC for the out-of-sample NIG model[36] using (14), which turns out to be lower than the AIC for the Gaussian in-sample model[37]. This suggests that not only does the out-of-sample NIG model of returns outperform the out-of-sample Gaussian model; it also outperforms the in-sample Gaussian model, according to the AIC. This conclusion is further confirmed by a likelihood ratio test[38] as explained in section 3.2. As we already noted, any model that aims to be useful in risk measurement should be robust to judgement calls about data sample,

[34] Consider the calculation of risk metrics such as VaR and ES – the pros of including extra observations further back in time in the data set has to be weighed against the diminishing relevance of the data set as a whole. This implies that a risk manager/compliance officer in 2007 (just before the financial crisis) would have to make some judgement calls whether the observations from the great banking and housing crisis in the late 80s beginning of the 90s are still relevant. One would ideally use a model that gives results that are robust to such judgement calls, i.e., where the results of the model is in the proximity to the "true" metrics whether the judgement goes one way or another.

[35] By over-fitting we mean modelling a property of a sample which is specific to that sample and not a population property.

[36] $\text{AIC}(\text{NIG out of sample}) = 2 * 4 - 15688.63 * 2 = -31369.3.$

[37] $\text{AIC}(\text{Gaussian in sample}) = 2 * 2 - 2 * 15159.87 = -30315.7.$

[38] We get the following test statistic in this case,

$$\lambda_{\text{LR}} = 2 * [15688.63 - 15159.87] = 1057.52.$$

This yields an extremely low P-value and we can conclude that NIG out of sample outperform Gaussian in sample.

which means that a model that can fit all sorts of special features of the sample period, e.g. positive skewness, might not be a good model for risk measurement purposes. It is therefore particularly interesting (and promising for NIG modelling in risk measurement) that the out-of-sample NIG model actually over-fitted the sample moments pre 1993 in the sense that it predicts a positive skewness, but it still outperformed the Gaussian distribution.

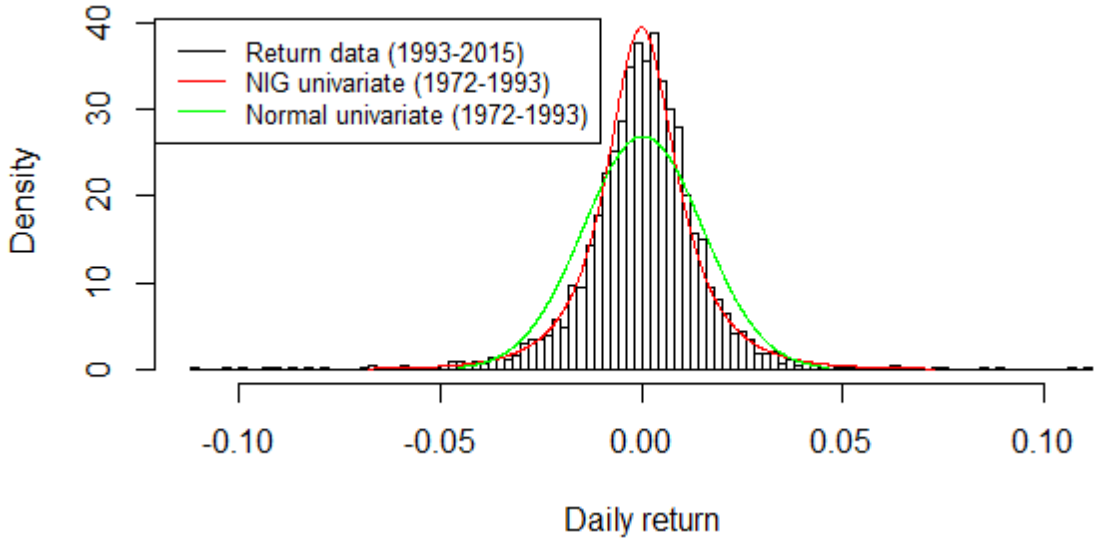

Figure 18: Out of sample for estimated models of daily returns.

	NIG model (1972-1993)
α	55.662
δ	0.0117
β	2.3776
μ	-0.00021
γ=(α^2-β^2)^0.5	55.612

Table 14: Parameter estimates of NIG out of sample.

	NIG moments (1972-1993)	Normal moments (1972-1993)	Sample moments (1993-2015)
Expected value	0.000293	0.000294	0.0002490757
Variance	0.000212	0.000221	0.0002176176
Skewness	0.1585	0	-0.5133
Kurtosis	7.625	3	9.838

Table 15: Moments for out of sample evaluation of models on daily returns.

Levels of p	Sample VaR (1993-2015)	NIG model VaR (1972-1993)	Normal model VaR (1972-1993)
0.1%	-0.0909	-0.0677	-0.0456
0.5%	-0.0570	-0.0475	-0.0380
1%	-0.0448	-0.0394	-0.0343
2.5%	-0.0304	-0.0292	-0.0288
5%	-0.0228	-0.0221	-0.0242

Table 16: Out of sample value at risk for daily data.

Levels of p	Sample ES (1993-2015)	NIG model ES (1972-1993)	Normal model ES (1972-1993)
0.1%	-0.1135	-0.0812	-0.0497
0.5%	-0.0795	-0.0602	-0.0427
1%	-0.0633	-0.0516	-0.0393
2.5%	-0.0461	-0.0407	-0.0345
5%	-0.0361	-0.0330	-0.0304

Table 17: Out of sample expected shortfall for daily data.

5. Equity premium puzzle and multivariate NIG modelling

This chapter will analyze and model quarterly consumption growth and real returns from the second quarter of 1978 to the fourth quarter of 2014 for Norway. The first two sections will present stylized facts and more in-depth analysis of the data. Section 5.3 will present two

theoretical models for the equity premium and the risk-free rate in a continuous time version of the Lucas (1978) exchange economy. Firstly, we model the dividend and consumption growth processes with continuous sample paths, which will be our baseline model. Secondly, we instead model the dividend and consumption growth processes as pure jump processes, like Aase and Lillestøl (2015). Section 5.4 fits the theoretical expressions developed in 5.3 to the data and presents the implied risk-aversion parameter and impatience rate in three different cases: our baseline model, a Gaussian jump model and a NIG jump model. Section 5.5 explores the driving forces behind our results and suggests other explanations.

5.1 Stylized facts about quarterly consumption growth

In Table 18, we present the same type of descriptive statistics as in Table 1 for seasonally adjusted continuously compounded quarterly consumption growth per capita in Norway (henceforth simply consumption growth). The data set for per capita quarterly consumption growth is calculated by subtracting the continuously compounded population growth from the quarterly aggregated consumption growth. Per capita consumption growth is hence given by

$$\dot{c_t} := \log\left(\frac{C_{t+1}}{C_t}\right) - \log\left(\frac{N_{t+1}}{N_t}\right), \tag{21}$$

where C_t and N_t is Norwegian aggregate consumption and population, respectively, at time t. Ideally, the population growth should be calculated using the time intervals matching the aggregate consumption growth. This is the approach used for the period from 1998 to 2014. However, due to lack of quarterly population data for the period before 1998, we calculate consumption growth per capita as consumption growth less one fourth of annual population growth. Mathematically, consumption growth per capita pre 1998 is calculated as

$$\hat{\dot{c}}_t := \log\left(\frac{C_{t+1}}{C_t}\right) - \frac{1}{4}\log\left(\frac{N_{t+i}}{N_{t-(4-i)}}\right), \tag{22}$$

where $i \in [1,4]$ is the number of quarters remaining in the year.

We observe that the quarterly consumption growth vary between a growth of about negative three and a half percent to a positive four and a quarter percent. The average consumption growth in our data set has been 0.54% on a quarterly basis. Consumption growth has a lower coefficient of variation than the returns in the stock market, reflecting that the consumption growth process in the economy is less volatile than the stock market. Furthermore, the consumption growth is

slightly negatively skewed and has a positive excess kurtosis. Applying the Shapiro and Wilk (1965) normality test on consumption growth, we get a P-value of 0.348, which implies that we cannot reject normality. Figure 19 also confirms this conclusion – the data points seem to be located on a straight line[39]. We can therefore conclude that the negative skewness and the excess kurtosis are not statistically significant different from what normality suggests.

In addition to addressing normality we have investigated whether there are significant autocorrelations in the quarterly consumption growth or residuals squared of quarterly consumption growth (calculated as in equation (8)). The conclusion of these analyses is that there are no significant autocorrelations in either consumption growth or its residuals. The conclusion of this section is hence that an i.i.d. Gaussian process could be a good model for Norwegian consumption growth.

	Quarterly consumption growth per capita
Min	-0.03548
Max	0.04252
Mean	0.00538
Variance	0.00015
Coefficient of variation (CV)	2.27
Skewness	-0.0924
Kurtosis	3.611
N	147
P-value in normality test	0.3482

Table 18: Descriptive statistics and moments for quarterly consumption growth data.

[39] We also fitted a univariate NIG and a univariate Gaussian model to the consumption growth data and found that the AIC of the Gaussian model in fact were lower than in the NIG model (-875.4 and -873.3 for Gaussian and NIG, respectively). A likelihood ratio test, as explained in section 3.2, yields a P-value= 0.3704, and we cannot conclude that the two models differ in performance. These results suggest that the Gaussian model is a better choice than the NIG model for univariate modelling of consumption growth.

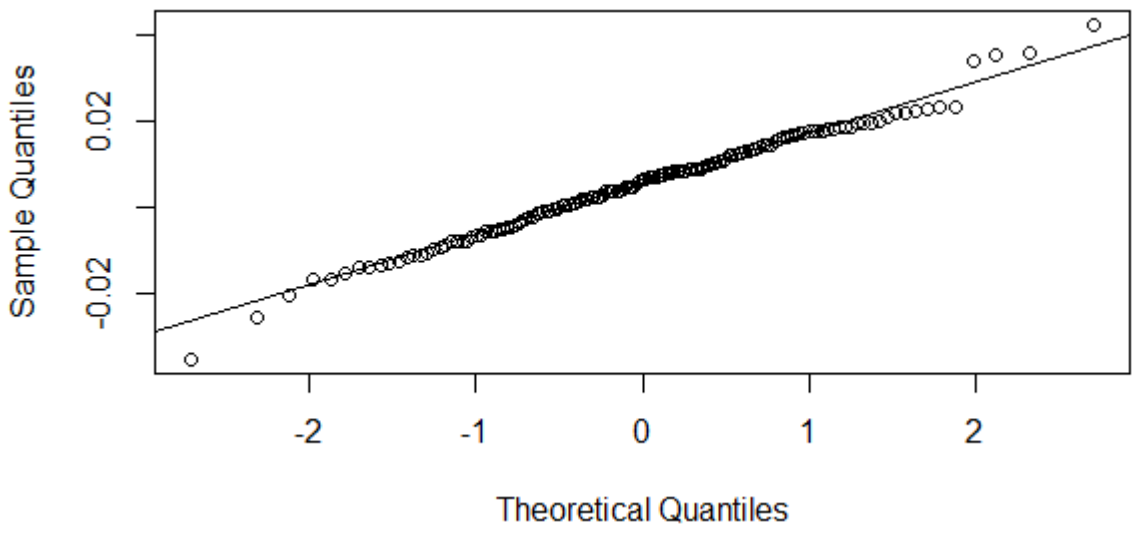

Figure 19: Normal Q-Q plot for quarterly consumption growth data.

5.2 Empirical analysis of multivariate data

In Table 19, we have provided sample cross central moments of consumption growth and returns. The first of these – correlation – is defined as

$$S_{r\dot{c}} = \frac{E[(r-\mu_r)(\dot{c}-\mu_{\dot{c}})]}{\sigma_r \sigma_{\dot{c}}},$$

where μ_X is the expected value of X, r is real return, $\dot{c}$ is consumption growth and σ_X is the (sample) standard deviation of X.

The two non-trivial coskewness measures defined by Miller (2014) are given by

$$S_{r\dot{c}\dot{c}} = \frac{E[(r-\mu_r)(\dot{c}-\mu_{\dot{c}})^2]}{\sigma_r \sigma_{\dot{c}}^2},$$

and

$$S_{rr\dot{c}} = \frac{E[(r-\mu_r)^2(\dot{c}-\mu_{\dot{c}})]}{\sigma_r^2 \sigma_{\dot{c}}},$$

Miller (2014) defines the three non-trivial cokurtosis statistics in the following way

$$S_{rrr\dot{c}} = \frac{E[(r-\mu_r)^3(\dot{c}-\mu_{\dot{c}})]}{\sigma_r^3 \sigma_{\dot{c}}},$$

$$S_{rr\dot{c}\dot{c}} = \frac{E[(r-\mu_r)^2(\dot{c}-\mu_{\dot{c}})^2]}{\sigma_r^2\sigma_{\dot{c}}^2},$$

and

$$S_{r\dot{c}\dot{c}\dot{c}} = \frac{E[(r-\mu_r)(\dot{c}-\mu_{\dot{c}})^3]}{\sigma_r\sigma_{\dot{c}}^3},$$

In Table 19, we see that correlation is positive and equal to about 0.27. This means that if return on the stock exchange during a quarter is high (low), then consumption growth is likely to be high (low) as well. We also note that both coskewness measures are negative. A negative coskewness suggests that the variability in one variable tends to be high when the expected realization of the other is low. For our purpose, this means that volatility in real returns is high when consumption growth is low and volatility of consumption growth is high when real returns are low[40]. An implication of this is that crises where both consumption growth and return is extremely low are more likely than their equivalent extreme positive outcomes.

From Table 19, we see that all three cokurtosis measures are positive. The interpretation of a positive cokurtosis is that extreme outcomes in one variable (negative or positive) tend to be occurring together with extreme outcomes in the other variable (negative or positive) (Ranaldo & Favre, 2005). Another way of saying this is that the correlation of the instantaneous variance of the two variables is positive. In our case, this implies that periods with high volatility in consumption growth is connected with high-volatility in the stock market.

[40] Another observation is that coskewness $S_{r\dot{c}\dot{c}}$ is smaller in absolute value than $S_{rr\dot{c}}$, suggesting that the effect of high volatility of returns when consumption is low is greater than the opposite effect of high volatility of consumption when the stock market performs poorly.

	Real return (r) and Consumption growth $(\dot{c})$
$S_{r\dot{c}}$ (correlation)	0.2747
$S_{r\dot{c}\dot{c}}$ (coskewness)	-0.1985
$S_{rr\dot{c}}$ (coskewness)	-0.3719
$S_{rrr\dot{c}}$ (cokurtosis)	1.4862
$S_{rr\dot{c}\dot{c}}$ (cokurtosis)	0.0027
$S_{r\dot{c}\dot{c}\dot{c}}$ (cokurtosis)	0.8751

Table 19: Sample standardized cross central moments for quarterly real returns and consumption growth.

We also provide a contour plot and a three-dimensional density plot for consumption growth and returns in Figure 20 and Figure 21, respectively (using the R-package *MVN* by Korkmaz, Goksuluk and Zararsiz (2015)). The two plots indicate that extreme negative outcomes in the two variables tend to occur simultaneously, which is supportive of our previous findings. To see this, one should notice that the density is higher for the outcomes where both variables are extremely negative than for outcomes where both variables are extremely positive. From the contour plot in Figure 20, for instance, one can see the bump in the south-west region where consumption growth is at its lowest and the stock market is substantially negative as well.

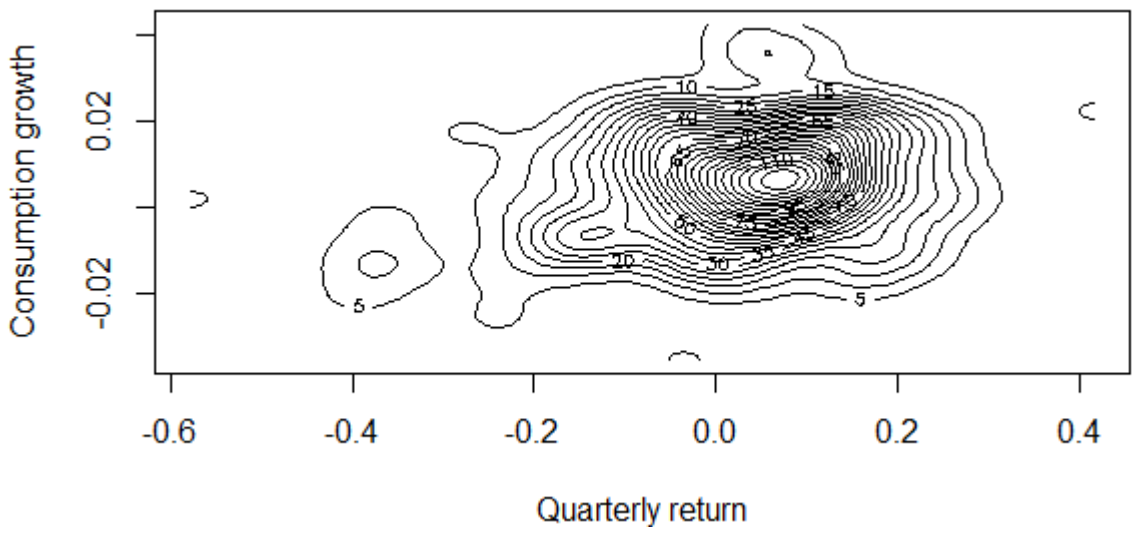

Figure 20: Contour plot of bivariate data consisting of quarterly real return and consumption growth.

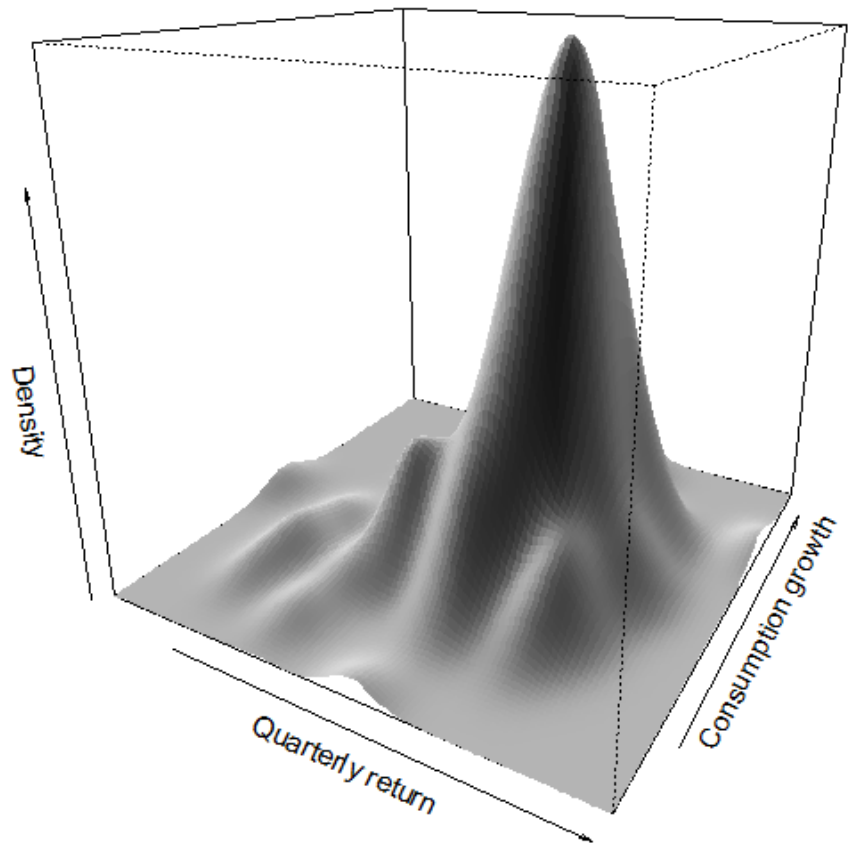

Figure 21: Perspective plot of bivariate data consisting of quarterly real return and consumption growth.

We have also tested whether consumption growth and real returns are bivariate normally distributed, using the test by Rizzo and Szekely (2005) in the R-package *energy* (Rizzo & Szekely, 2014). The null-hypothesis of the test is that consumption growth and real returns are bivariate normal. This multivariate normality test uses a bootstrap framework (similar to the bootstrap approach applied in section 3.6), which involves resampling the data randomly (Rizzo & Szekely, 2005). This approach to testing bivariate normality provides a different P-value each time we run the test. We always get low P-values however, which enables us to reject the null hypothesis of bivariate normality at even a 0.01% significance level. We obtain similar results applying other tests as well[41]. The conclusion of non-normality in this bivariate distribution makes it worthwhile to consider opening up for non-normality in this chapter.

5.3 Utility maximization problem – a theoretical model

We consider an endowment economy where Arrow-Debreu securities for all states and dates up to T are trading at time 0. Suppose there exists a representative consumer[42] and that her optimization problem is to choose a non-negative consumption plan in order to maximize a von Neumann-Morgenstern (expected[43]) utility function. Furthermore, suppose the felicity index belongs to the constant relative risk aversion[44] (CRRA) class of utility functions. The consumption plan has to belong to the set of feasible consumption plans, i.e. the present value of consumption cannot exceed the present value of the endowment process. The representative consumer's problem is

$$\sup_{c\in L_+} U(c) = E\left[\int_0^T e^{-\rho t}\frac{c_t^{1-\theta}}{1-\theta}dt\right] \quad \text{s.t. } \Pi(c) \le \Pi(e) \text{ where } \theta > 0 \text{ and } \theta \neq 1. \qquad (23)$$

[41] The other multivariate normality tests all reject bivariate normality with P-values of no more than 0.0007. The other tests we applied are the ones by Royston (1983), Henze and Zirkler (1990), and Mardia (1974).

[42] Gorman form is necessary and sufficient. In short, Gorman form requires the expenditure function of each consumer to be an affine function with respect to utility: $e^i\left(\boldsymbol{p}, u^i(c)\right) = f^i(\boldsymbol{p}) + u^i(c)g(\boldsymbol{p})$, where e^i is the expenditure needed for consumer i to reach utility u^i. Both $f^i(\boldsymbol{p})$ and $g(\boldsymbol{p})$ are homogenous of degree one in the price vector $\boldsymbol{p}$.

[43] Note that expected utility does not have an axiomatic foundation in a multiperiod setting; the reason is that the substitution axiom breaks down, see e.g. Mossin (1969).

[44] The relative risk aversion is defined as $R(c) := -\frac{u''(c)}{u'(c)}c$. In the case of felicity index $u(c) = \frac{c^{1-\theta}-1}{1-\theta}$, which becomes log(c) by LHôpital's rule when θ=1, $R(c) = \theta$.

Here ρ is a utility discount factor or impatience rate, θ is the parameter of relative risk aversion[45] and c_t is the consumption per capita at period t. L_+ is the set of non-negative consumption processes that satisfies $E\left[\int_0^T c_t^2 dt\right] < \infty$ and $\Pi(\cdot)$ is a pricing functional with the Riesz representation

$$\Pi(x) = E\left[\int_0^T \pi_t x_t dt\right], \tag{24}$$

where π is the state price deflator. The Lagrangian for this problem is

$$\mathcal{L}(c,\eta) = E\left[\int_0^T \left(e^{-\rho t}\frac{c_t^{1-\theta}}{1-\theta} + \eta\pi_t(e_t - c_t)\right)dt\right]. \tag{25}$$

The first order condition for this problem is found by setting the directional derivative to zero, see e.g. Aase and Lillestøl (2015). In our case, the first order condition will be both necessary and sufficient for optimality, the latter coming from the concavity of the utility function and convexity of the feasible consumption set. Optimal consumption is given by

$$c_t = e_t = (\eta e^{\rho t}\pi_t)^{-\frac{1}{\theta}}. \tag{26}$$

We can, without loss of generality, normalize η to 1. Note that this implies an inverse relationship between consumption and the state price deflator – whenever consumption is low the state price deflator is high, and vice versa. The intuition for this is that whenever consumption is scarce, marginal utility of consumption is high, thus the price of consumption is driven up. This price mechanism is needed in a decentralized economy to induce consumers to act in a way consistent with equilibrium. The degree to which consumers responds to price changes depends on θ – a high value implies a lower willingness to reduce consumption in states or dates where consumption is expensive.

Since Arrow-Debreu securities do not trade, we introduce a securities market of the type discussed by Aase and Lillestøl (2015) as a means of financing the consumer's consumption plans. In this market, N+1 securities are traded ex dividend at competitive prices *S* and have accumulated dividend processes *D*,

[45] θ also has a connection to the intertemporal elasticity of substitution (IES), more specifically $\frac{1}{\theta}$ = IES.

$$S(t) = \left(S^0(t), S^1(t), \dots, S^N(t)\right) \tag{27}$$

and

$$D(t) = \left(D^0(t), D^1(t), \dots, D^N(t)\right), \tag{28}$$

where $t \in [0, T]$. We let the zeroth security be the riskless asset by convention. The real price process of the riskless asset is $S^0(t) = e^{\int_0^t r_u du}$, where r is a bounded short rate process. Let a gains process be given by $G(t) = S(t) + D(t)$. In the framework of Aase and Lillestøl (2015), the gains process is an Ito-jump-diffusion process in $\mathbb{R}^{N+1}$.

Let $\varphi \in \varkappa^2(G)$, where $\varkappa^2(G)$ is the set of square integrable trading strategies. We impose this restriction on the trading strategies in order to avoid arbitrage in continuous time. One may loosely think of this restriction as limiting the number of trades in a given time interval, thus avoiding arbitrage arising from doubling strategies[46]. Let p be the spot price of consumption. The trading strategy is said to finance the household's consumption plan if

$$\varphi_t S_t = \int_0^t \varphi_s dG_s - \int_0^t p_s c_s ds \tag{29}$$

and $\varphi_T S_T = 0$ (Aase & Lillestøl, 2015). To see why the boundary condition makes sense, consider the no-Ponzi game restriction – the household must have a non-negative terminal wealth. Noting that our utility function is strictly increasing, it will never be optimal to leave a bequest. In this setting, the household has to choose an optimal portfolio and consumption in each period. The optimization problem thus becomes

$$\sup_{(c,\varphi)} U(c) \quad \text{where } (c, \varphi) \in L_+ \times \varkappa^2(G) \quad \text{and } \varphi \text{ finances } (c - e) \tag{30}$$

The dynamics of the consumption and dividend process is of the same jump-diffusion type as discussed by Aase and Lillestøl (2015)

[46] Doubling strategies like the Martingale betting system where a gambler (for example) participates in a fair coin toss and double her bet each time the coin toss is lost, providing a sure win (arbitrage) if unlimited amount of trades is possible and there is no credit constraints.

$$\frac{dc(t)}{c(t-)} = \mu_c(t)dt + \sigma_c(t)dB(t) + \int_Z \gamma_c(t,z)\tilde{N}(dz,dt) \tag{31}$$

$$dD^i(t) = \mu_D^i(t)dt + \sigma_D^i(t)dB(t) + \int_Z \gamma_D^i(t,z)\tilde{N}(dz,dt), \qquad i = 0,1,\dots,N \tag{32}$$

where $\tilde{N}(dz,dt)$ is a 2-dimensional compensated Poisson measure[47], $B(t)$ is a d-dimensional Brownian motion and $\sigma_c(t)$ and $\sigma_D^i(t)$ are vectors of appropriate dimension scaling the Brownian motion. In the continuous version of the processes in (31) and (32), $|\sigma_X(t)|$ is the instantaneous volatility of X. γ_c and γ_D^i are the stochastic jump-sizes, and $\mu_c(t)$ and $\mu_D^i(t)$ are the deterministic growth rates of the consumption and dividend processes, respectively (Aase & Lillestøl, 2015).

To get a feeling for what the correct price of a risky asset in this economy should be, we begin by considering the discrete time version of the Lucas (1978) pure exchange economy. He finds that the equilibrium price of any risky asset, *i*, assuming zero end-value, should be

$$S_t^i = \frac{1}{u'(c_t)} E_t\left[\sum_{s=t}^{T} \frac{\Delta D_{s+1}^i u'(c_{s+1})}{(1+\rho_d)^{s+1-t}}\right]. \tag{33}$$

Here $u(\cdot) := \frac{c^{1-\theta}}{1-\theta}$ is a CRRA[48] felicity function and $\Delta D_{s+1}^i := D_{s+1}^i - D_s^i$ is the dividend paid out in the beginning of period $s+1$. ρ_d is a discrete time version of the utility discount rate. It is important to note that the price of the asset and the dividends are both in terms of units of the consumption good in the corresponding time period. To interpret the meaning of (33), we multiply both sides by the marginal utility of consumption at time t

$$S_t^i u'(c_t) = E_t\left[\sum_{s=t}^{T} \frac{\Delta D_{s+1}^i u'(c_{s+1})}{(1+\rho_d)^{s+1-t}}\right]. \tag{34}$$

We now see that the left-hand side of (34) gives us the product of S_t^i units of the consumption good and the (marginal) utility per unit of the consumption good. One may loosely think of this as the utility value of the security at time t. Similarly, the right-hand side gives us the (sum of) expected utility value of the dividends paid out by the security, discounted by the utility discount rate. If the equality in (34) does not hold, the representative consumer could either buy or sell

[47] $\tilde{N}(dz,dt) = N(dz,dt) - v(dz)dt$, where $N(dz,dt)$ is a counting process for the jumps and $v(dz)$ is a Levy-measure.

[48] Note that Lucas (1978) did not assume CRRA, so (33) is more general than what we consider here.

(depending on which way the inequality goes) the asset and increase her expected lifetime utility. Seeing as the representative consumer is a stand-in for the entire economy, this mechanism would tend to cause the price to revert back to its equilibrium value as given by (33).

Note that (34) is closely related to the Euler equation for optimal intertemporal consumption. To see this, we recognize that (34) can be re-written as

$$S_t^i u'(c_t) = E_t\left[\frac{(\Delta D_{t+1}^i + S_{t+1}^i)u'(c_{t+1})}{(1+\rho_d)}\right]. \tag{35}$$

Divide both sides of (35) by S_{t+1}^i to obtain the (stochastic) Euler equation

$$u'(c_t) = E_t\left[\frac{(1+r_{d,t+1}^i)}{(1+\rho_d)}u'(c_{t+1})\right], \tag{36}$$

where $r_{d,t+1}^i := \frac{\Delta D_{t+1}^i + S_{t+1}^i}{S_t^i} - 1$ is the discrete time cum-dividend rate of return.

Let $\pi_t := e^{-\rho t}u'(c_t)$, where $\rho := \log(1+\rho_d)$ is the continuously compounded utility discount rate. π_t is then the state price deflator. To see what happens in a continuous time version of (33), rewrite it in terms of the state price deflator and let the number of trading times between t and T go to infinity[49], following Aase (2008),

$$S_t^i = \frac{1}{\pi_t}\lim_{n\to\infty} E_t\left[\sum_{j=tn}^{Tn} \Delta D_{\frac{j+1}{n}}^i \pi_{\frac{j+1}{n}}\right] =$$

$$\frac{1}{\pi_t}\lim_{n\to\infty} E_t\left[\sum_{j=tn}^{Tn}\left(\left(D_{\frac{j+1}{n}}^i - D_{\frac{j}{n}}^i\right)\left(\pi_{\frac{j+1}{n}} - \pi_{\frac{j}{n}}\right) + \left(D_{\frac{j+1}{n}}^i - D_{\frac{j}{n}}^i\right)\pi_{\frac{i+1}{n}}\right)\right] =$$

$$\frac{1}{\pi_t}E_t\left[\int_t^T \left(dD_s^i d\pi_s + dD_s^i \pi_s\right)\right]. \tag{37}$$

We note that (37) is similar to (33), but that the continuous time version contains an additional covariance term – the covariance between the dividends and the state price deflator. The

[49] One can think of n as the number of trading times in the original time unit, e.g. if time originally was measured in months, n could be the number of seconds in a month. Of course, as n approaches infinity, the length of a time interval goes to zero, i.e. it becomes infinitesimal. We know this limit exists because both the consumption process and cumulative dividend process are right continuous with left limits (this is a property of all Itô-processes), i.e. the left limit $X_{t-} := \lim_{s\uparrow t} X_s$ and the right limit $X_t := \lim_{s\downarrow t} X_s$ both exist for all $t \in [0,T]$.

covariance term is shown to be of particular importance when jumps are introduced in the dividend and consumption processes, for further elaboration on this see Aase (2015).

In our baseline model, the continuous time version of the Lucas (1978) model without jumps, i.e. the only source of uncertainty in (31) and (32) is the standard Brownian motion, we get the familiar expressions for the equilibrium equity premium on a stock market index and interest rate (Aase, 2002), as shown in Appendix B,

$$\mu_{r_M}(t) - r_f(t) = \theta \sigma'_{r_M}(t)\sigma_c(t) \tag{38}$$

and

$$r_f(t) = \rho + \theta\mu_c(t) - \frac{1}{2}\theta(1+\theta)\sigma'_c(t)\sigma_c(t). \tag{39}$$

Here $\mu_{r_M}(t)$ is the expectation of cum dividend return on the stock market index conditional on all information available up to time *t*, and $r_f(t)$ is the risk free interest rate. $\left|\sigma_{r_M}(t)\right|$ is the instantaneous volatility of total return on the stock market index. $\sigma'_{r_M}(t)\sigma_c(t)$ is the instantaneous covariance between the return on the index and the consumption growth process. Similarly, $\sigma'_c(t)\sigma_c(t)$ is the instantaneous variance of the consumption growth process. When we estimate this model in the next section, we will take both the covariance and variance mentioned here as constant in time.

Notice that both (38) and (39) has nice, intuitive interpretations. Firstly, (38) tells us that the investors only requires compensation for the "systematic risk", i.e. the covariance between consumption and the asset return (as opposed to the variance of the asset), associated with holding risky assets. The size of this compensation is determined by the risk aversion – a high risk aversion implies a high risk-premium. Secondly, (39) tells us that the interest rate is increasing in the impatience rate rho and in the expected consumption growth and decreasing in the variance of consumption growth. A higher rho, all else equal, implies that the consumers wish to borrow against future income to consume more today. Similarly, if expected future consumption rises, consumers would wish to bring some of that consumption to today when the marginal utility of that consumption is higher. The size of this effect is determined by the intertemporal elasticity of substitution (IES), $\frac{1}{\theta}$ – a high IES means that the marginal utility of consumption is less responsive to changes in consumption levels; thereby making consumers

more willing to accept variation in consumption. The final term in (39) is related to precautionary savings – risk averse consumers wish to save more in the face of uncertainty. The precautionary savings is increasing in the risk aversion of the consumer. Since a single step-in agent can represent all consumers, moving consumption in time (by either borrowing or saving) would be impossible and any attempts of doing so would cause interest rates to rise in order to clear the markets.

Aase and Lillestøl (2015) finds the equilibrium interest rate and equity premium expressed by moment generating functions in the two cases when the dividend and consumption growth processes in (31) and (32) are pure jump processes[50] and the jump-sizes are either normally or NIG distributed,

$$\mu_{r_M}(t) - r_f(t) = -\lambda[M(-\theta, 1) - M(-\theta, 0) - M(0,1) + 1] \tag{40}$$

and

$$r_f(t) = \rho + \theta\mu_c(t) - \lambda[\theta(M(1,0) - 1) + M(-\theta, 0) - 1]. \tag{41}$$

Here μ_c is the expected consumption growth, estimated as the average consumption growth in the data set. $M(u_c, u_r)$ is the moment generating function and λ is the expected jump frequency. Note that the expressions in (38) and (39) are only dependent on the first two moments of the multivariate distribution. By introducing jumps into this framework, the equity premium and risk free rate will depend on moments of all orders, as seen in (40) and (41).

The moment generating function of a joint normal distribution is given by (Aase & Lillestøl, 2015)

$$M^N(u_c, u_r) = exp\left\{\hat{\mu}_c u_c + \hat{\mu}_r u_r + \frac{1}{2}(u_c^2\hat{\sigma}_c^2 + 2u_c u_r\hat{\rho}_n\hat{\sigma}_c\hat{\sigma}_r + u_r^2\hat{\sigma}_r^2)\right\}, \tag{42}$$

where $\hat{\rho}_n$ is the correlation coefficient between consumption growth and the return on the risky asset and the $\hat{\mu}$s and $\hat{\sigma}$s are parameters of the bivariate normal jump model. We use the notation $\hat{X}$ to distinguish the parameters of our bivariate normal and NIG distribution from the parameters in (31) and (32).

The moment generating function for the NIG distribution is given by (Aase & Lillestøl, 2015)

[50] In other words, we are ignoring the uncertainty arising from the Brownian motion in (31) and (32).

$$M^{NIG}(\boldsymbol{u}) = exp\left\{\boldsymbol{u}'\boldsymbol{\mu} + \delta\left[\sqrt{\alpha^2 - \boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta}} - \sqrt{\alpha^2 - (\boldsymbol{\beta}+\boldsymbol{u})'\boldsymbol{\Delta}(\boldsymbol{\beta}+\boldsymbol{u})}\right]\right\}, \quad (43)$$

where δ and α are scalars related to scale and peakedness of the joint NIG distribution. $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ are two-dimensional vectors related to location and skewness and $\boldsymbol{\Delta}$ is a 2x2-matrix related to, although in a complicated manner, the covariation of consumption growth and the return on the risky asset (Lillestøl, 1998).

5.4 Model estimation and calibration of impatience and risk aversion

5.4.1 Multivariate model estimates

In this sub-section, we have estimated two bivariate models of real return and consumption growth. One of the models we have estimated is a bivariate Gaussian model and the other one is a bivariate NIG model. As we did in section 3.2, we compare the bivariate Gaussian model with the bivariate NIG model using the AIC (Akaike, 1974) as defined by (14). The AIC of our two models are provided in Table 20. According to the AIC the bivariate NIG model outperforms the bivariate Gaussian model[51]. This implies that our multivariate NIG model fits the density in Figure 20 and Figure 21 better than the multivariate Gaussian model even to a degree where it also compensates the punishment that occurs when introducing more parameters to the model. We also perform a likelihood ratio test, as explained in section 3.2, between our two candidate models and we conclude that the bivariate NIG model significantly[52] outperforms the bivariate Gaussian model.

We have also provided the parameter estimates in our two models in Table 20. The two models have similar expectations, variances and covariance values but differ in other aspects of the bivariate distributions. Corresponding to the findings in the empirical analysis in section 5.1 and 5.2, both consumption growth and real returns have positive expectations and the returns have a substantially higher variance. Furthermore, we notice that the covariance between the consumption growth and the real returns are positive in both models, which is as expected considering the positive sample correlation found in section 5.2. Both models are estimated using the maximum likelihood approach (like in the univariate case in section 3.2) for multivariate datasets given in the R-package *ghyp* by Luethi and Breymann (2013).

[51] This is not all that surprising considering the clear rejection of the null hypothesis of bivariate normality in the data in section 5.2.

[52] The likelihood ratio test yields a P-value$= 2.88 * 10^{-6}$.

	Multivariate Gaussian	Multivariate NIG
AIC	-1043.603	-1063.114
$(\mu_{\dot{c}} \quad \mu_r)$	(0.0054 0.0208)	(0.0054 0.0208)
$\begin{pmatrix} \hat{\sigma}_c^2 & \hat{\sigma}_{rc} \\ \hat{\sigma}_{rc} & \hat{\sigma}_r^2 \end{pmatrix}$	$\begin{pmatrix} 0.00015 & 0.00047 \\ 0.00047 & 0.01949 \end{pmatrix}$	$\begin{pmatrix} 0.00016 & 0.00042 \\ 0.00042 & 0.01862 \end{pmatrix}$
α		40.343
δ		0.0526
(βc βr)		(-11.033 -5.5648)
$(\hat{\mu}_c \quad \hat{\mu}_r)$		(0.0086 0.1060)
$\begin{pmatrix} \Delta_{cc} & \Delta_{rc} \\ \Delta_{rc} & \Delta_{rr} \end{pmatrix}$		$\begin{pmatrix} 0.1030 & 0.1861 \\ 0.1861 & 10.045 \end{pmatrix}$

Table 20: Parameters for the estimated multivariate models of quarterly consumption growth and real stock return.

5.4.2 Estimates of risk aversion and impatience – the equity premium puzzle

In this sub-section, we will investigate the equity premium puzzle for Norway using three approaches. Firstly, we consider a continuous process for cumulative dividends and consumption and calculate the implied risk aversion and impatience. Secondly, we will use the bivariate Gaussian jump model to describe the dynamics of the underlying processes. Thirdly, we allow for non-normality in the jump-sizes by using the bivariate NIG-jump model of sub-section 5.4.1.

As pointed out earlier we use quarterly data when addressing the equity premium puzzle, like for instance Lettau (2002). This is opposed to Mehra and Prescott (1985) who use annual data. The reason why we make this decision is the larger dataset available for quarterly return and consumption growth. One issue with quarterly data that needs to be addressed, however, is seasonality. When calibrating the models we are using seasonally adjusted consumption growth. This makes sense because optimizing households would not systematically plan to consume more in certain parts of the year unless their marginal utility of consumption is higher during these periods for a given consumption level. Our fix for this seasonal effect is therefore to adjust the

units of consumption[53]. To investigate how our choice of quarterly frequency affects our results, we will discuss our findings in this section in relations to annual descriptive statistics as well.

The continuous process model delivers a relative risk aversion parameter of 30.64 and a utility discount factor of –35.2%, as shown in Table 21. Notice that the utility discount factor is provided at an annual basis. A significantly negative utility discount factor is a puzzle in itself, because it implies that people prefer consumption in the future over consumption today, all else equal[54]. Additionally, the relative risk aversion parameter is way too high compared to what is considered reasonable.

Equations (40) and (41) in section 5.3 give us the relative risk aversion, θ, and the utility discount factor, ρ, as implicit functions of our parameter estimates, equity premium and risk-free rate in the case of pure jump-processes (also see Appendix C for how they are calculated). Notice that for our purposes, we can normalize the expected jump frequency (λ in equations (40) and (41)) per quarter to one. The multivariate Gaussian jump process yields a risk aversion parameter of 26.09 and a utility discount factor of -30.4%. Introducing jumps hence seem to move the parameter estimates in the right direction, but quantitatively the effect is rather small.

By moving from the multivariate Gaussian jump process to the multivariate NIG jump process, we investigate what the effect is of allowing for non-normality in the bivariate model of consumption growth and real returns. In Table 21, we see that the relative risk aversion parameter becomes 25.25 and that the utility discount factor is -28.4%. The result of opening up for non-normality is that both parameters moves further in the right direction, but quantitatively the effect is still small. In addition to being unreasonable, a relative risk aversion above 20 is clearly contradicting with the extensive study of Aarbu and Schroyen (2014), also for Norway, who find an average relative risk aversion of 3.7 in their survey with thought (simple) lotteries.

[53] Seasonality of consumption is a result of planned consumption shifting (happens every year). For instance, people most likely choose to consume more during Christmas times because their marginal utility of consumption (for a given level of consumption) is higher this time of year. There are two ways of handling this. One is to increase the complexity of the utility function. The other is to say that one unit of consumption during Christmas time is more than one unit of consumption during the rest of the year, which is done here.

[54] Our estimate implies that the representative consumer would rather consume 100 of a consumption good in one year than consuming 130 of the consumption good today.

	Multivariate Gaussian continuous process (Norway)	Multivariate Gaussian jump process (Norway)	Multivariate NIG jump process (Norway)	Multivariate NIG jump process (US) (Aase and Lillestøl, 2015)
θ	30.64	26.09	25.25	22.20
ρ	-0.352	-0.304	-0.284	0.008

Table 21: Results from our analysis on the equity premium puzzle, given on an annual basis.

Aase and Lillestøl (2015) use the dataset of Mehra and Prescott (1985) to estimate these models for the US. In Table 21 we also provide the results by Aase and Lillestøl (2015) for their multivariate NIG jump process. They get estimates for the risk aversion parameter of the same order of magnitude as we get here, which implies an equity premium puzzle. Their utility discount factor is, however, in the proximity of zero on a yearly basis. Qualitatively, they also get slightly more realistic parameter estimates by introducing jumps and then allowing for non-normality, but also in their case the effect is quantitatively small.

In Table 22, we provide descriptive statistics on an annual basis from our dataset, all related to our resulting equity premium puzzle. We provide similar statistics for the US in parentheses (Aase & Lillestøl, 2015). Notice that the implied impatience rate is affected in two different ways in our data set compared to Aase and Lillestøl (2015). Firstly, the consumption growth has been more than 20 percent higher per capita in Norway[55] while the variance of consumption growth in Norway is about 40 percent of what it is in the US, both of which cause a lower implied impatience rate. Secondly, with the opposite effect, the real interest rate has been more than two percentage points higher in our dataset (on a yearly basis), which cause a higher implied impatience rate. The total effect depends on theta – a high theta (as is the case here) means that the effect of consumption growth and variance dwarfs the effect of higher real interest rate.[56]

[55] A high consumption growth in Norway in the period 1978-2014 can partly be explained by the development of the petroleum sector in Norway during this period.

[56] To illustrate each effect, consider a theta of 25. The difference in consumption growth is 0.4 percentage points, which reduces the implied impatience rate in the baseline model by $25 \times 0.4 = 10$ percentage points. The difference in consumption variance is -0.077 percentage points, causing the implied impatience rate to fall by $\frac{1}{2} \times 25 \times (25 + 1) \times 0.077 = 25$ percentage points. These two effects sum up to -35 percentage points, which is much greater than the positive effect of the interest rate of 2 percentage points.

Another observation to be made from Table 22 is that the covariance between consumption growth and return is higher in the US, possibly explained by S&P500 being a better proxy for the theoretical market portfolio. The higher covariance, in combination with lower equity premium in the US, explains why we get a somewhat higher estimated relative risk aversion for the Norwegian representative consumer.

	Expectation	Standard deviation	Covariances
Consumption growth $(\dot{c})$	2.15% (1.75%)	2.22% (3.55%)	$\mathrm{cov}(r,\dot{c}) = 0.00155$ (0.00227)
Real stock market return (r)	8.30% (5.53%)	32.27% (15.84%)	$\mathrm{cov}(r,r_f) = 0.00002$ (0.00148)
Real risk-free rate (r_f)	2.66% (0.64%)	2.83% (5.74%)	$\mathrm{cov}(\dot{c},r_f) = -0.00002$ (-0.00015)
Equity premium	5.64% (4.89%)	32.39% (15.95%)	

Table 22: Descriptive statistics for Norway (1979-2014) on an annual basis. The corresponding results of Aase and Lillestøl (2015) for the US (1889-1978) are given in parentheses.

To illustrate what we mean by an equity premium puzzle, consider a lottery that either doubles an investor's initial wealth, or leaves initial wealth unchanged. The two outcomes are equally probable. The certainty equivalent of this lottery is implicitly given by the following expression

$$u(W_0 + CE) = E[u(XW_0)],$$

where

$$X = \begin{cases} 2 \text{ with probability } 0.5 \\ 1 \text{ with probability } 0.5. \end{cases}$$

A risk neutral investor would be willing to pay the expected payoff of the lottery, i.e. half her initial wealth. Typically, a value around two is seen as reasonable for the relative risk aversion, theta. In this case, the certainty equivalent is two thirds of the expected payoff of the lottery. In the case of a relative risk aversion parameter of 30.64, however, she would only be willing to pay around two percent of her wealth to participate in the lottery. The relative risk aversion decreases slightly when we use the Gaussian jump model, hence increasing the certainty equivalence slightly. In our NIG jump process of consumption and real return, her valuation of the gamble is 2.9% of her wealth. This implies that she would rather have 103 percent of initial wealth ($W_0 +$

$0.03W_0$) for sure than participating in a lottery where she ends up with twice her wealth ($2W_0$) or initial wealth (W_0) with equal probability. Thus, the equity premium puzzle is neither solved by introducing jumps, nor by allowing for non-normality in return and consumption growth.

Relative risk aversion	Certainty equivalent
$\theta = 0$	$0.5000W_0$
$\theta = 2$	$0.3333W_0$
$\theta = 25.25$	$0.0290W_0$
$\theta = 26.09$	$0.0280W_0$
$\theta = 30.64$	$0.0237W_0$

Table 23: Implications of our risk aversion estimates for certainty equivalence.

5.5 The equity premium – why is it a puzzle?

The baseline model of the previous two sections – a continuous time version of the rational expectations model by Lucas (1978) – has three main underlying assumptions. Firstly, both the consumption growth process and cumulative dividend process are continuous and normally distributed. Secondly, a representative consumer maximizes a time-separable and additive utility function, i.e. expected utility, in a multiperiod context. Thirdly, there are no market frictions. In this paper, we have loosened the first assumption, using the same approach as Aase and Lillestøl (2015). A second approach is to increase the complexity of the consumer preferences, e.g. by introducing recursive utility. The latter approach will indeed help explain this puzzle, as commented by Aase and Lillestøl (2015). Mehra and Prescott (1985) suggested in their original paper that introducing market frictions, e.g. some form of credit constraint, could help explain the puzzle.

As noted in section 2.3 and section 5.2 – both the stock market returns and the joint distribution of stock market returns and consumption growth exhibit fatter tails than the normal distribution. Tail-events tend to occur simultaneously (as measured by the cokurtosis) and extreme negative tail-events are more likely than extreme positive tail-events (as measured by the negative coskewness). Even in the absence of negative coskewness, the presence of a positive cokurtosis increases the required equity premium as long as the utility function is concave, i.e. risk-averse

consumers. The intuition is that an increase in cokurtosis increases the probability of consumption and returns being either extremely low or extremely high at the same time. These two extreme outcomes are however, weighted differently – extreme low outcomes have very high marginal utilities, whereas extremely high outcomes have very low marginal utilities. Similarly, the presence of negative coskewness places more mass of the distribution on outcomes of moderately higher consumption and returns, which are weighted with moderately lower marginal utilities, but at the same time a negative coskewness place more mass in the very negative end of the distribution, which in turn is weighted with a very high marginal utility.

We saw in chapter 3 and chapter 4 that the univariate NIG distribution is able to capture the skewness and kurtosis of the univariate return distribution quite well. In chapter 4, we in particular note that the NIG distribution came close to matching the empirical tail-mass, which was substantially higher than what one assumes in a Gaussian model. This is anecdotal evidence that the bivariate NIG distribution can be the right extension of a model that tries to explain the equity premium puzzle. As the results in the previous section showed, NIG does indeed take us qualitatively in the right direction, but the quantitative effect is almost (economically) insignificant.

The second approach, namely loosening up the assumptions about the representative consumer's utility function, is likely to yield better results. To see why the utility assumptions of the baseline model are problematic, we have to look at the meaning of the parameter theta. This parameter is related to two properties of the model: the risk aversion is equal to theta, and in addition the IES is equal to one over theta. This of course implies that a value for theta determines both the degree to which the representative consumer wishes to smooth consumption across states of the world, and across time. There are however, no economic reasons to why these two properties of the model should be the same. In fact, as noted by Mossin (1969), the expected utility representation has no axiomatic foundations in a multi-period setting. One natural extension is therefore to allow the risk aversion and IES to be determined separately by two different parameters, which is in fact what recursive utility introduces (see for instance Aase (2015)).

6. Conclusions

This thesis shows that there are significant deviations from normality in Norwegian stock returns. More specifically, returns are shown to be significantly both leptokurtic and negatively skewed. These observations invalidate "mean-variance" analysis for risk and performance measurement of a stock portfolio. In chapter 3, we therefore present a normal inverse Gaussian (NIG) model of returns that allows for higher order moment effects. Our modelling approach is not however, able to capture the volatility clustering that we observe, but we do provide a rough sketch of how this can be accomplished.

In chapter 2 and 3 we find two additional properties in our stock return data that suggests NIG to be a promising candidate for univariate modelling. Firstly, a property of the NIG model is that when controlling for the instantaneous variance, we obtain the Gaussian distribution. This property of the model is found to be consistent with data. Secondly, the mixed mean-variance structure of NIG suggests the instantaneous variance to be IG distributed. We do find in the data that the instantaneous variance of returns is non-constant, and the IG distribution is shown to fit our calculated variance series well. In our estimated NIG model the parameter β implies a negative relationship between instantaneous variance and returns, making the distribution of returns negatively skewed. We find a similar negative relationship in the data by running regressions of return on its squared residuals.

We find that the NIG model of returns (not surprisingly) clearly, and significantly, outperforms the Gaussian model in sample. We argue however that the models should be evaluated out of sample when assessing their usefulness in risk management. It turns out that the out-of-sample NIG model outperforms even the in-sample Gaussian model, both in terms of risk metrics such as value at risk and expected shortfall, and in terms of overall fit measured by AIC. The likelihood ratio test confirms that our conclusion is statistically significant. The promising performance of NIG out of sample suggests that it is robust to judgement calls about sample period, which is an attractive feature for models that are used for risk measurement purposes. However, compared to actual observations of daily return in our data sample, even our NIG model underestimates the severity of the most extreme outcomes. An interesting extension to our model for risk measurement would be to allow for volatility clustering, using e.g. a NIG-GARCH model for the returns, suggested by Forsberg and Bollerslev (2002).

In chapter 5, we find that quarterly consumption growth in Norway does not deviate significantly from normality. We do however, find significant non-normality for the joint consumption growth and return distribution. In particular, there is evidence of negative coskewness and excess cokurtosis. Using a theoretical model for the equilibrium interest rate and equity premium under the expected utility hypothesis suggested by Aase and Lillestøl (2015), we are able to move beyond "mean-variance" analysis and allow for higher-order moment effects. As expected, the presence of negative coskewness and excess cokurtosis (and possibly even higher-order moment effects) reduces the implied risk aversion parameter and increases the impatience rate, moving both parameters in the right direction. The effects are however, quantitatively small and we are therefore not able to resolve the equity premium puzzle or the risk-free rate puzzle. Given the added complexity and relatively small quantitative effects of including higher-order moments, we conclude that "mean-variance" analysis might in fact be the right level of abstraction when modeling equilibrium interest rates and equity premium.

Several alternative models have been suggested in order to explain the equity premium and risk free rate puzzles. A promising approach is to allow the parameter governing consumption smoothing across time (IES) to differ from the parameter governing consumption smoothing across states (risk aversion). This can be achieved by modeling consumer preferences with recursive utility (thereby leaving the expected utility hypothesis), see e.g. Aase (2015). Another approach, suggested by e.g. Mehra and Prescott (1985), is to allow for market imperfections, which could cause the covariance, negative coskewness, excess cokurtosis etc. between consumption growth and stock returns to be more pronounced on an individual level than at the aggregate level.

Appendix A – Plots and graphs

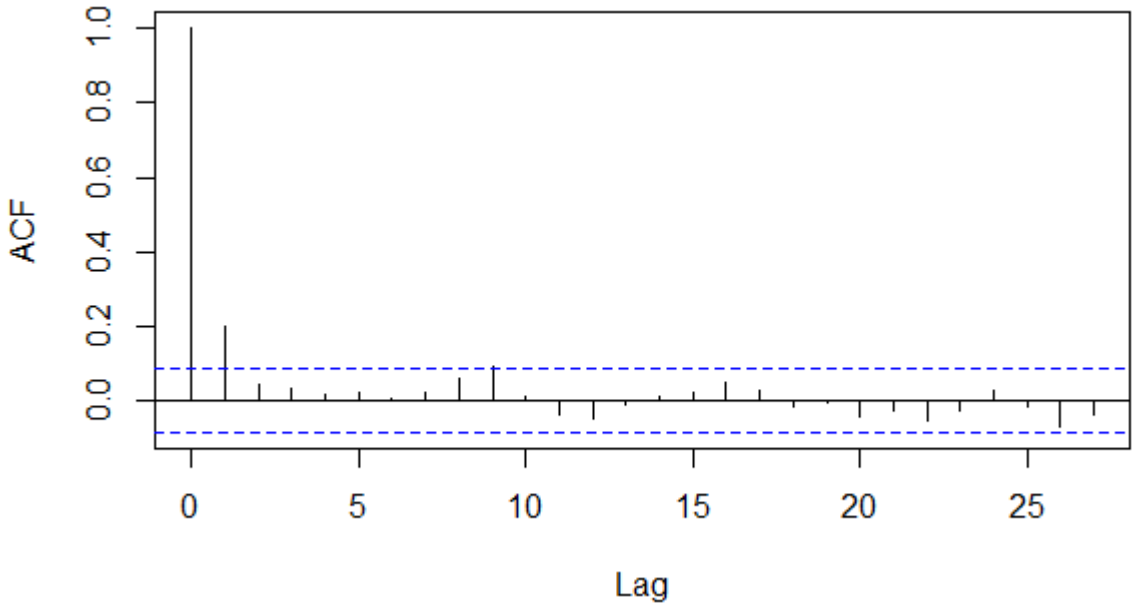

Figure A 1: ACF of residuals squared of monthly Norwegian stock returns.

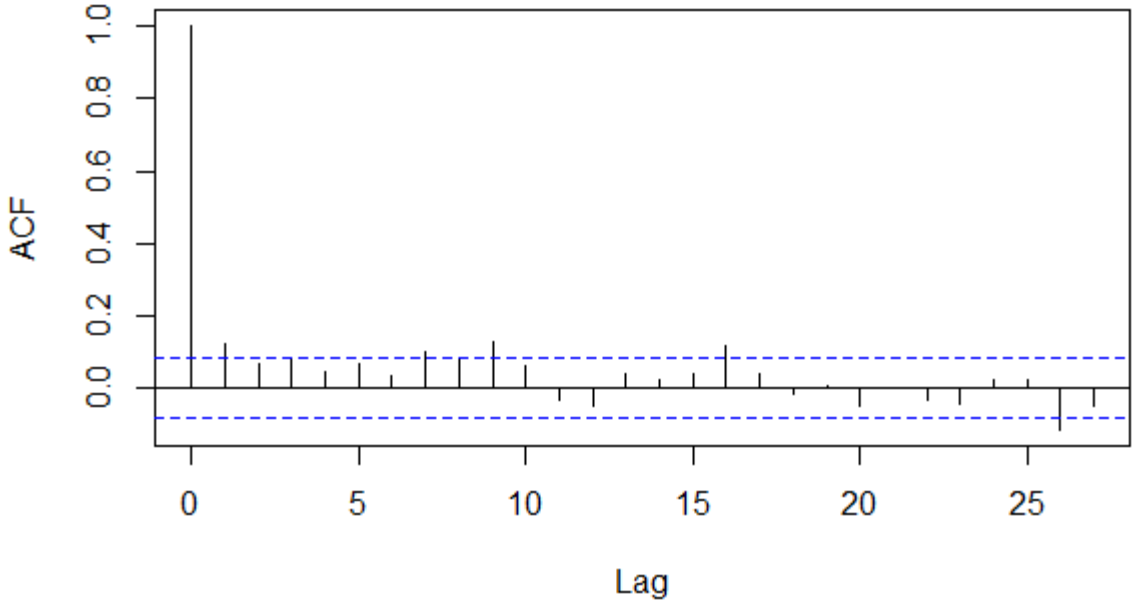

Figure A 2: ACF of absolute values of residuals of monthly Norwegian stock returns.

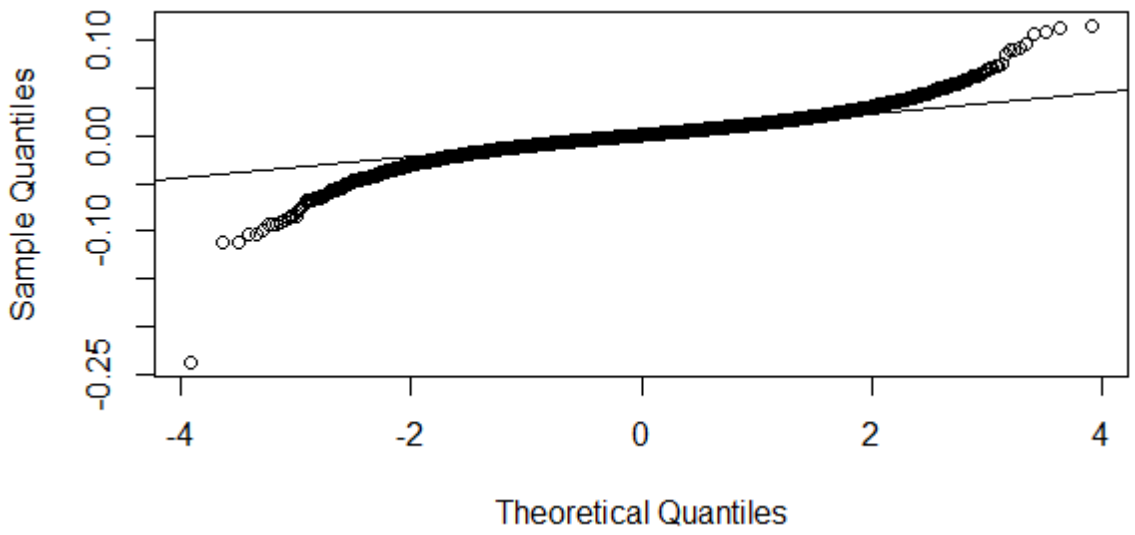

Figure A 3: Normal Q-Q plot for Norwegian daily returns.

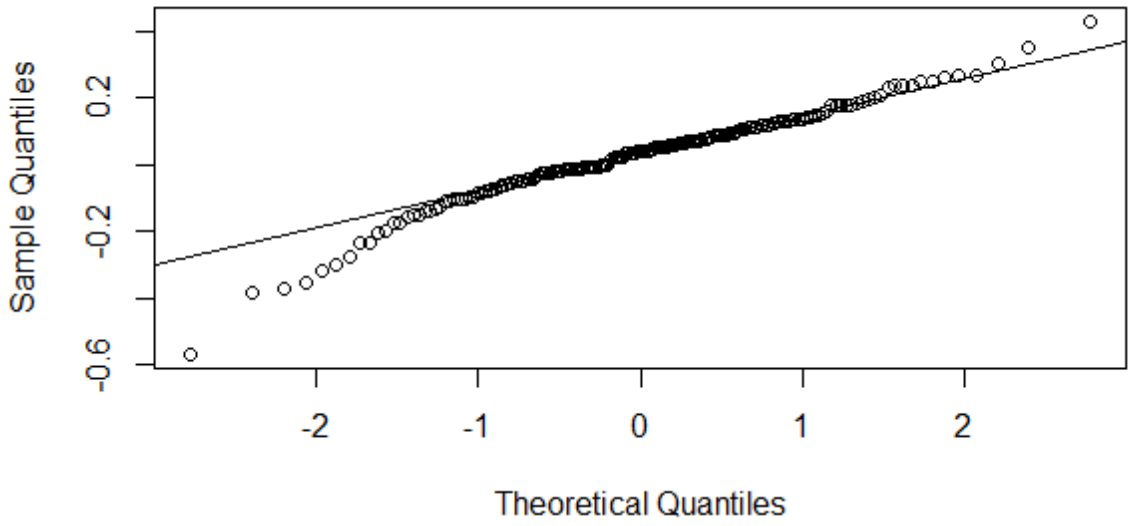

Figure A 4: Normal Q-Q plot for Norwegian quarterly returns.

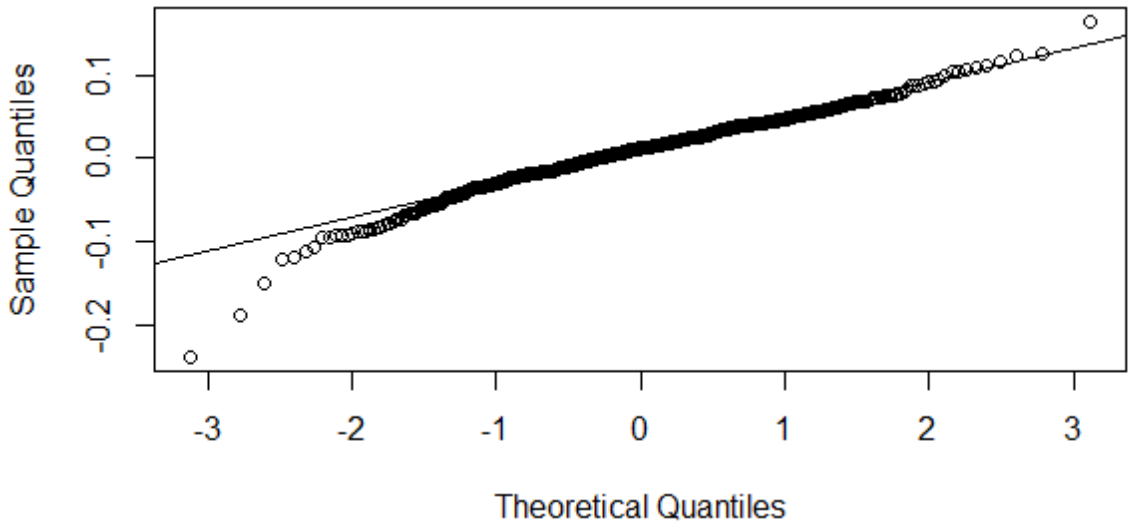

Figure A 5: Normal Q-Q plot for US monthly returns.

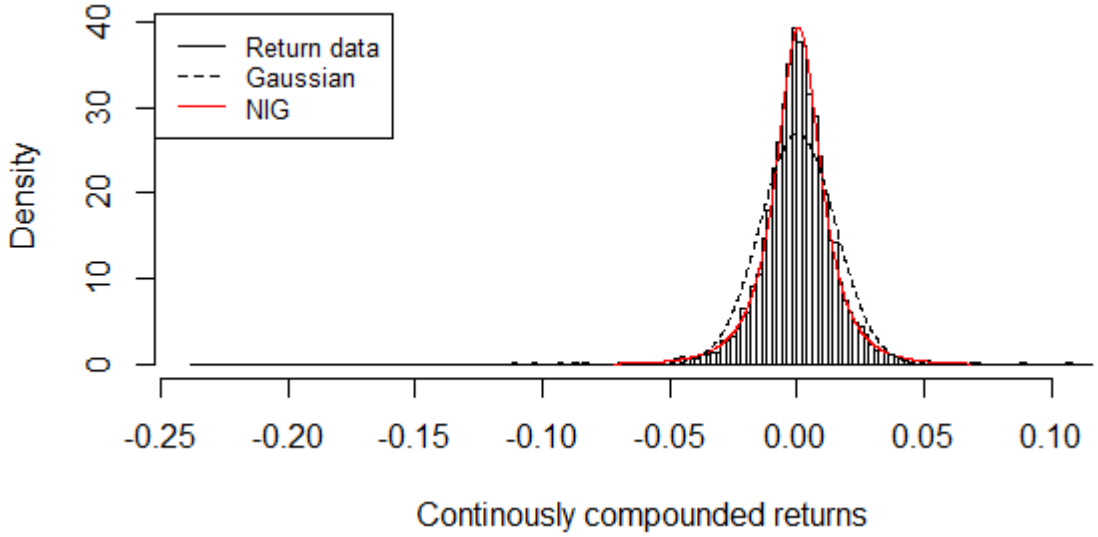

Figure A 6: Distribution of Norwegian daily price returns with NIG-fit (1972-2015).

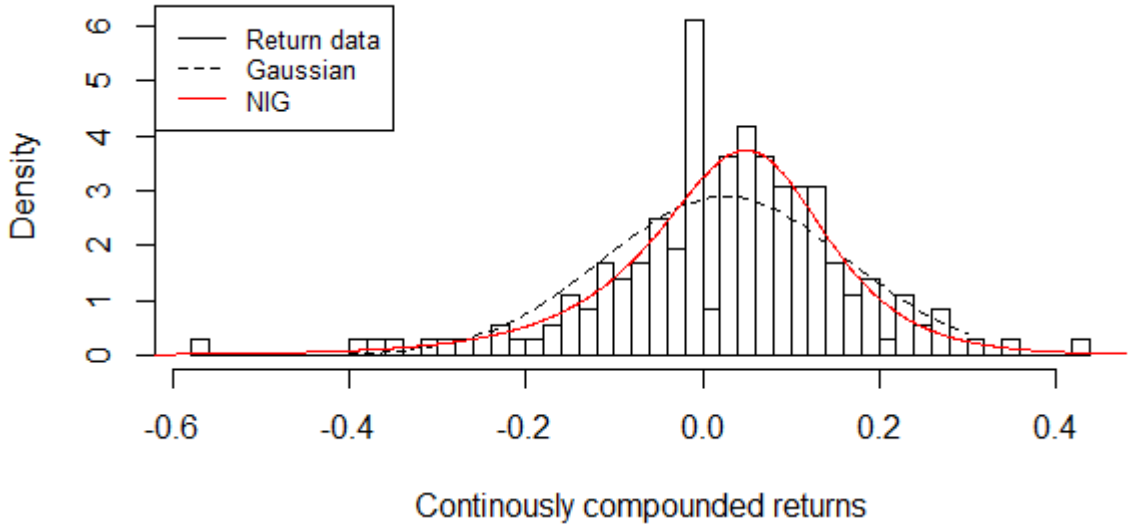

Figure A 7: Distribution of Norwegian quarterly returns with NIG-fit (1970-2014).

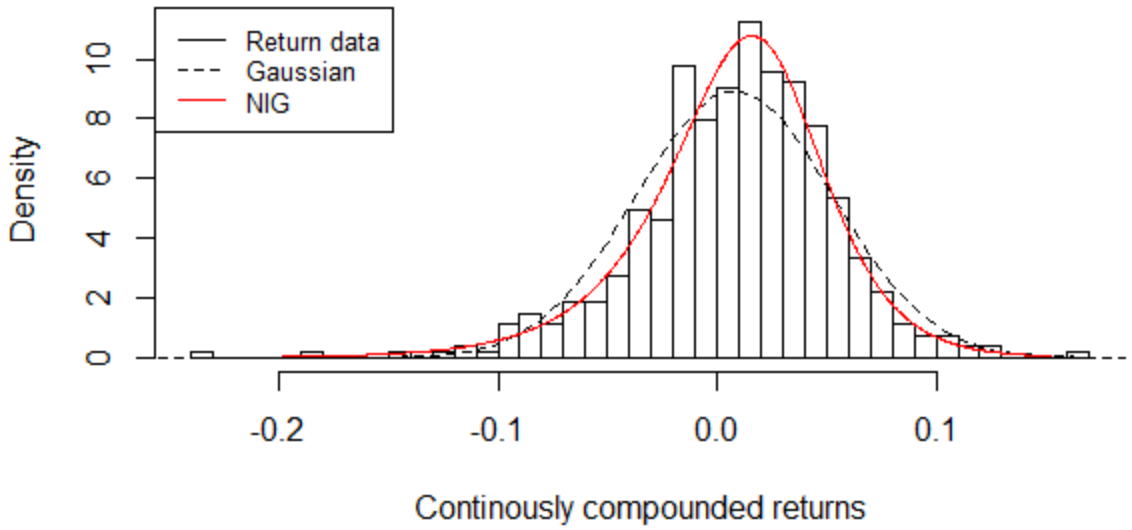

Figure A 8: Distribution of US monthly returns with NIG-fit (1970-2014).

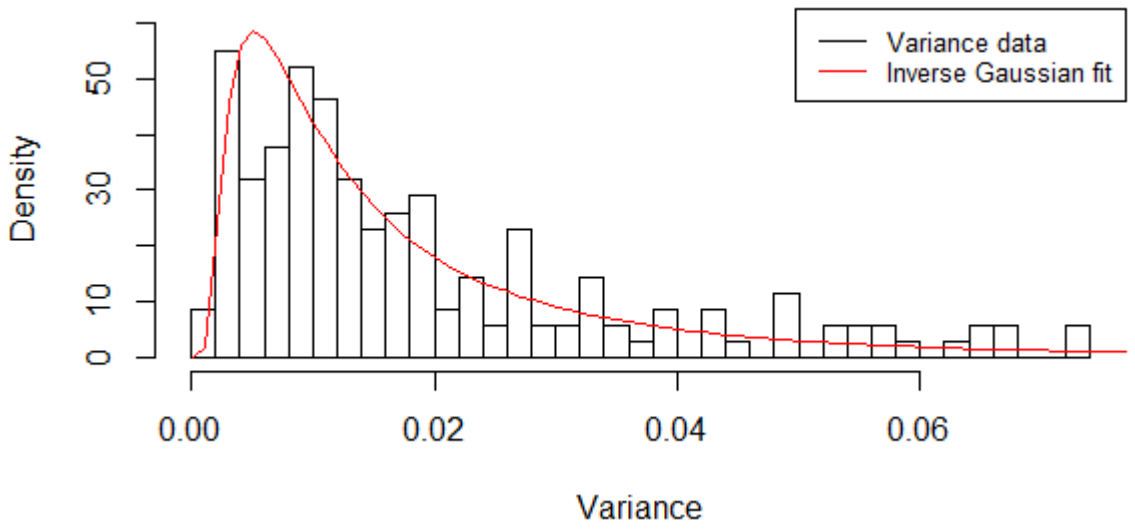

Figure A 9: Norwegian quarterly variance of returns and inverse Gaussian fit.

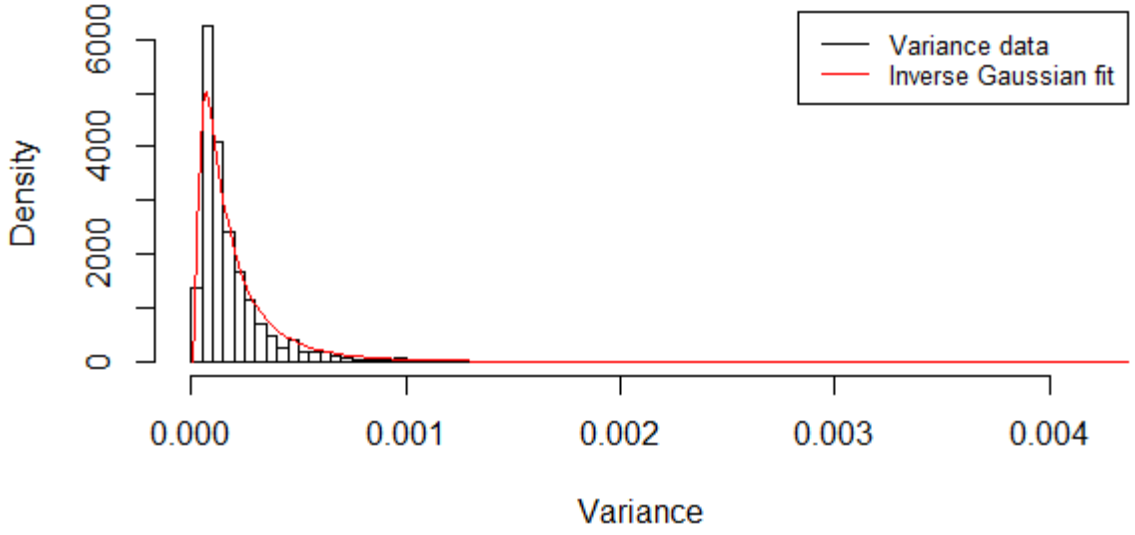

Figure A 10: Norwegian daily variance of returns and inverse Gaussian fit.

Appendix B – Equilibrium interest rate and equity premium in our baseline model

In this appendix, following Aase (2002), we derive the equilibrium interest rate and equity premium for our baseline model and start out with the equilibrium price of any risky asset in the economy, from expression (37),

$$S^i(t) = \frac{1}{\pi_t} E_t \left[\int_t^T \left(dD_s^i d\pi_s + dD_s^i \pi_s \right) \right]. \tag{B1}$$

Let $\tilde{S}^i(t) := \pi_t S^i(t)$ and rewrite (B1) as

$$\tilde{S}^i(t) = E_t \left[\int_t^\tau \left(dD_s^i d\pi_s + dD_s^i \pi_s \right) + \tilde{S}^i(\tau) \right], \qquad \tau \in [t, T]. \tag{B2}$$

Since both the dividends and consumption are (continuous sample path) Itô processes, π will also be an Itô process, assuming $u(\cdot)$ is sufficiently smooth, that is: its first three derivatives exist. In the CRRA case (as applied in chapter 5), derivatives of all orders exist. We can therefore write the dynamics of π and $\tilde{S}^i$ as

$$d\pi_t = \mu_\pi(t)dt + \sigma_\pi(t)dB_t \text{ and } d\tilde{S}^i(t) = \mu_{\tilde{S}^i}(t)dt + \sigma_{\tilde{S}^i}(t)dB_t.$$

Recall from section 5.3 that the continuous sample path process for the dividends is[57]

$$dD^i(t) = \mu_{D^i}(t)dt + \sigma_{D^i}(t)dB_t.$$

Then (B2) can be written as

$$\tilde{S}^i(t) = E_t \left[\int_t^\tau \left(\sigma'_{D^i}(s)\sigma_\pi(s) + \mu_{D^i}(s)\pi_s \right) ds + \int_t^\tau \pi_s \sigma_{D^i}(s) dB_s + \tilde{S}^i(t) \right.$$
$$\left. + \int_t^\tau \left(\mu_{\tilde{S}^i}(s)ds + \sigma_{\tilde{S}^i}(s)dB_s \right) \right]$$

$$\Downarrow$$

[57] Ignoring the jump part in expression (32).

$$E_t\left[\int_t^\tau \left(\sigma_{D^i}(s)\sigma_\pi(s) + \mu_{D^i}(s)\pi_s\right) ds + \int_t^\tau \pi_s \sigma_{D^i}(s) dB_s + \int_t^\tau \left(\mu_{\tilde{S}^i}(s)ds + \sigma_{\tilde{S}^i}(s)dB_s\right)\right] = 0$$

$$\Downarrow$$

$$\mu_{\tilde{S}^i}(s) + \sigma_{D^i}(s)\sigma_\pi(s) + \mu_{D^i}(s)\pi_s \equiv 0$$

$$\Updownarrow$$

$$\mu_{\tilde{S}^i}(t) = -\left(\sigma_{D^i}(t)\sigma_\pi(t) + \mu_{D^i}(t)\pi_t\right) \forall\, t \in [0, T]. \tag{B3}$$

By the product rule, we can write the dynamics of $\tilde{S}^i$ as

$$d\tilde{S}^i(t) = d\left(\pi_t S^i(t)\right) = d\pi_t S^i(t) + \pi_t dS^i(t) + d\pi_t dS^i(t) =$$

$$\left(\mu_\pi(t)S^i(t) + \mu_{S^i}(t)\pi_t + \sigma'_\pi(t)\sigma_{S^i}(t)\right)dt + \left(\sigma_\pi(t)S^i(t) + \sigma_{S^i}(t)\pi_t\right)dB_t.$$

This implies that

$$\mu_{\tilde{S}^i}(t) \equiv \mu_\pi(t)S^i(t) + \mu_{S^i}(t)\pi_t + \sigma'_\pi(t)\sigma_{S^i}(t) \;\; and \;\; \sigma_{\tilde{S}^i}(t) \equiv \sigma_\pi(t)S^i(t) + \sigma_{S^i}(t)\pi_t.$$

Inserting the expression for $\mu_{\tilde{S}^i}(t)$ into (B3) and rearranging yields

$$\frac{\mu_{S^i}(t) + \mu_{D^i}(t)}{S^i(t)} = -\left(\frac{\mu_\pi(t)}{\pi_t} + \frac{\sigma'_\pi(t)}{\pi_t}\frac{\left(\sigma_{D^i}(t) + \sigma_{S^i}(t)\right)}{S^i(t)}\right). \tag{B4}$$

Recall that the continuous sample path process for c can be written as

$$dc(t) = \mu_c(t)c(t)dt + \sigma_c(t)c(t)dB_t.$$

Recall from section 5.3 that $\pi_t := e^{-\rho t}u'(c_t)$. By Itô's lemma we can then write the dynamics of π as

$$\begin{aligned} d\pi_t = &\left(-\rho e^{-\rho t}u'(c_t) + e^{-\rho t}u''(c_t)\mu_c(t)c(t) + \frac{1}{2}e^{-\rho t}u'''(c_t)\sigma'_c(t)\sigma_c(t)c(t)^2\right)dt \\ &+ e^{-\rho t}u''(c_t)c(t)\sigma_c(t)dB_t. \end{aligned} \tag{B5}$$

Using the functional form $u(c_t) = \frac{c_t^{1-\theta}}{1-\theta}$ we can rewrite (B5) as

$$d\pi_t = -\left(\rho + \theta\mu_c(t) - \frac{1}{2}\theta(1+\theta)\sigma_c'(t)\sigma_c(t)\right)\pi_t dt - \theta\pi_t\sigma_c(t)dB_t. \quad \text{(B6)}$$

Matching $\mu_\pi(t)$ and $\sigma_\pi(t)$ in (B4) with the first and second term in (B6) respectively, yields

$$\frac{\mu_S(t) + \mu_D(t)}{S(t)} = \rho + \theta\mu_c(t) - \frac{1}{2}\theta(1+\theta)\sigma_c'(t)\sigma_c(t) + \theta\sigma_c'(t)\frac{\left(\sigma_{D^i}(t) + \sigma_{S^i}(t)\right)}{S^i(t)}.$$

Let $\sigma_{r_i}(t) := \frac{\left(\sigma_{D^i}(t) + \sigma_{S^i}(t)\right)}{S^i(t)}$. If there exists a riskless asset, its covariance with the consumption process should be zero. The return on the riskless asset can then be written as

$$r_f(t) = \rho + \theta\mu_c(t) - \frac{1}{2}\theta(1+\theta)\sigma_c'(t)\sigma_c(t). \quad \text{(B7)}$$

The risk premium on asset *i* is then

$$\mu_{r_i}(t) - r_f(t) = \theta\sigma_c'(t)\sigma_{r_i}(t) \quad \text{(B8)}$$

We can write (B8) for the special case, $i = M$, which is our Norwegian MSCI (2015a) index,

$$\mu_{r_M}(t) - r_f(t) = \theta\sigma_c'(t)\sigma_{r_M}(t). \quad \text{(B9)}$$

Appendix C – Calculation of impatience rate and risk aversion

In this appendix, we will express the equity premium and risk-free rate in the case of a pure jump model of consumption growth and equity returns, where the jump sizes are NIG distributed, in a way that easily lends itself to numerical or graphical solutions using standard software or a graphical calculator.

Recall that the equity premium and risk-free rate can be expressed by moment generating functions in the following way (Aase & Lillestøl, 2015)

$$\mu_{r_M} - r_f = -\lambda[M(-\theta,1) - M(-\theta,0) - M(0,1) + 1] \tag{C1}$$

and

$$r_f = \rho + \theta\mu_c - \lambda[\theta(M(1,0) - 1) + M(-\theta,0) - 1]. \tag{C2}$$

We have suppressed the time-subscripts as we now treat the equity premium and risk-free rate as constants. We will also normalize the expected jump-frequency lambda to one. Note that this is reasonable because we only have one observation per quarter, and we are not trying to explain how consumption growth or stock returns move between these observations.

The moment generating function of the bivariate NIG distribution has the following representation (Aase & Lillestøl, 2015)

$$M^{NIG}(\boldsymbol{u}) = exp\left\{\boldsymbol{u}'\boldsymbol{\mu} + \delta\left[\sqrt{\alpha^2 - \boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta}} - \sqrt{\alpha^2 - (\boldsymbol{\beta} + \boldsymbol{u})'\boldsymbol{\Delta}(\boldsymbol{\beta} + \boldsymbol{u})}\right]\right\}, \tag{C3}$$

where δ and α are scalars, $\boldsymbol{\mu}' = (\hat{\mu}_c, \hat{\mu}_r)$, $\boldsymbol{\beta}' = (\beta_c, \beta_r)$ and $\boldsymbol{\Delta} = \begin{pmatrix} \Delta_{cc} & \Delta_{cr} \\ \Delta_{cr} & \Delta_{rr} \end{pmatrix}$.

Notice that for any vector $\boldsymbol{u}' = (x, u_2)$, where x is some unknown and u_2 is some constant, we can represent (C3) by

$$M^{NIG}\big((x, u_2)'\big) = exp\big\{Ax + B - \delta\sqrt{Cx^2 + Dx + E}\big\}, \tag{C4}$$

where A, B, C, D and E are all some constants. Using (C4), we can rewrite (C1) and (C2) as

$$1 + R_e = -e^{\left\{A_1\theta + B_1 - \delta\sqrt{C_1\theta^2 + D_1\theta + E_1}\right\}} + e^{\left\{A_2\theta + B_2 - \delta\sqrt{C_2\theta^2 + D_2\theta + E_2}\right\}} + e^{\left\{A_3\theta + B_3 - \delta\sqrt{C_3\theta^2 + D_3\theta + E_3}\right\}} \quad \text{(C5)}$$

and

$$r_f = 1 + \rho + \theta\mu_c + \theta - \theta e^{\left\{A_4\theta + B_4 - \delta\sqrt{C_4\theta^2 + D_4\theta + E_4}\right\}} - e^{\left\{A_2\theta + B_2 - \delta\sqrt{C_2\theta^2 + D_2\theta + E_2}\right\}}, \quad \text{(C6)}$$

where R_e is the average equity premium in our dataset. μ_c is estimated as the average growth rate of consumption and r_f as the average risk-free rate in our quarterly dataset. All the constants in (C5) and (C6) consists solely of estimated parameters of our model:

$$A_1 = -\hat{\mu}_c, \quad B_1 = \hat{\mu}_r + \delta\gamma, \quad C_1 = -\Delta_{cc},$$

$$D_1 = 2\big((1 + \beta_r)\Delta_{cr} + \beta_c\Delta_{cc}\big), \;\; E_1 = \gamma^2 - 2(\beta_c\Delta_{cr} + \beta_r\Delta_{rr}) - \Delta_{rr},$$

$$A_2 = -\hat{\mu}_c, \quad B_2 = \delta\gamma, \quad C_2 = -\Delta_{cc}, \quad D_2 = 2(\beta_r\Delta_{cr} + \beta_c\Delta_{cc}), \;\; E_2 = \gamma^2,$$

$$A_3 = 0, \quad B_3 = \hat{\mu}_r + \delta\gamma, \quad C_3 = 0, \quad D_3 = 0, \quad E_3 = \gamma^2 - 2(\beta_c\Delta_{cr} + \beta_r\Delta_{rr}) - \Delta_{rr},$$

and

$$A_4 = 0, \quad B_4 = \hat{\mu}_c + \delta\gamma, \quad C_4 = 0, \quad D_4 = 0, \quad E_4 = \gamma^2 - 2(\beta_c\Delta_{cc} + \beta_r\Delta_{cr}) - \Delta_{cc}.$$

Here in the multivariate case of the NIG distribution, $\gamma := \sqrt{\alpha^2 - \boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta}}$ (as opposed to $\gamma := \sqrt{\alpha^2 - \beta^2}$ in the univariate case).

By first solving for theta in (C5), (C6) is an equation in only one unknown – rho.

References

Aarbu, K. O., & Schroyen, F. (2014). *Attitudes towards income risk in an extensive welfare state.* Bergen: Norwegian School of Economics.

Aase, K. K. (2002). Equilibrium Pricing in the Presence of Cumulative Dividends Following a Diffusion. *Mathematical Finance, Vol. 12, No. 3*, ss. 173-198.

Aase, K. K. (2008). On the Consistency of the Lucas Pricing Formula. *Mathematical Finance, Vol. 18, No. 2*, ss. 293-303.

Aase, K. K. (2015). *Recursive utility and jump-diffusions.* Bergen: Norwegian School of Economics.

Aase, K. K., & Lillestøl, J. (2015). *Beyond the local mean-variance analysis in continuous time. The problem of non-normality.* Bergen: Norwegian School of Economics.

Abramowitz, M., & Stegun, I. A. (1974). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover Publications.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, Vol. 19, Iss. 6*, ss. 716-723.

Anderson, R. M., Eom, K. S., Hahn, S. B., & Park, J. H. (2012). *Sources of stock return autocorrelation.* Berkeley: University of California at Berkeley, Department of Economics.

Andersson, J. (2001). On the Normal Inverse Gaussian Stochastic Volatility Model. *Journal of Business & Economic Statistics, Vol. 19, No. 1*, ss. 44-54.

Anscombe, F. J., & Glynn, W. J. (1983). Distribution of the kurtosis statistic b2 for normal samples. *Biometrika, Vol. 70, No. 1*, ss. 227-234.

Asparicio, F., & Estrada, J. (1997). *Empirical Distributions of Stock Returns: Scandinavian Securities Markets, 1990-1995.* Madrid: Carlos III University.

Balanda, P. K., & MacGillivray, H. L. (1988). Kurtosis: A Critical Review. *The American Statistician, Vol. 42, No. 2*, ss. 111-119.

Barndorff-Nielsen, O. (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics, Vol. 24*, ss. 1-13.

Barndorff-Nielsen, O., & Prause, K. (1999). *Apparent Scaling.* Aarhus: Department of Theoretical Statistics, University of Aarhus.

Barndorff-Nielsen, O., Blæsild, P., Jensen, J. L., & Sørensen, M. (1985). The Fascination of Sand. *A Celebration of Statistics*, ss. 57-87.

Benth, F. E., Groth, M., & Kettler, P. C. (2005). *A Quasi-Monte Carlo Algorithm for the Normal Inverse Gaussian Distribution and Valuation of Financial Derivatives.* Oslo: Department of Mathematics, University of Oslo.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference - Understanding AIC and BIC in Model Selection. *Sociological Methods & Research, Vol. 33, No. 2*, ss. 261-304.

Bølviken, E., & Benth, F. E. (2000). *Quantification of Risk in Norwegian Stocks via the Norman Inverse Gaussian Distribution.* Tromsø: Proceedings of the AFIR 2000 Colloquium.

Campbell, J. Y., & Shiller, R. J. (1988). Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance, Vol. 43, No. 3*, ss. 661-676.

D'Agostino, R. B., Belanger, A., & D'Agostino, R. B. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician, Vol. 44, No. 4*, ss. 316-321.

DeCarlo, L. T. (1997). On the Meaning and Use of Kurtosis. *Psychological Methods, Vol. 2, No. 3*, ss. 292-307.

Duffie, D., & Pan, J. (1997). An overview of Value at Risk. *Journal of Derivatives, Vol. 4, No. 3*, ss. 7-49.

Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* London: Chapman & Hall/CRC.

Eitrheim, Ø., Klovland, J. T., & Qvigstad, J. F. (2007). Historical Monetary Statistics for Norway - Part II. Oslo. Retrieved from http://www.norges-bank.no/Statistikk/Historisk-monetar-statistikk/

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics, Vol. 20, No. 3*, pp. 339-350.

Eriksson, A., Ghysels, E., & Wang, F. (2009). The normal inverse Gaussian distribution and the pricing of derivatives. *Journal of Derivatives, Vol. 16*, ss. 23-37.

Field, A. (2009). *Discovering Statistics Using SPSS, Third edition.* Los Angeles: SAGE Publications.

Forsberg, L., & Bollerslev, T. (2002). Bridging the Gap Between the Distribution of Realized (ECU) Volatility and ARCH Modelling (of the Euro): The Garch-NIG Model. *Journal of Applied Economics, Vol. 17*, ss. 535-548.

Friend, I., & Blume, M. E. (1965). The Demand for Risky Assets. *American Economic Review, Vol. 65, No. 5*, ss. 900-922.

Henze, N., & Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods, Vol. 19, Iss. 10*, ss. 3595-3617.

Hyndman, R. J., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician, Vol. 50, No. 4*, ss. 361-365.

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician, Vol. 47, Part 1*, pp. 183-189.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2015). *MVN: An R Package for Assessing Multivariate Normality*. Retrieved from http://cran.r-project.org/web/packages/MVN/vignettes/MVN.pdf

Lettau, M. (2002). Idiosyncratic Risk and Volatility Bounds, or Can Models with Idiosyncratic Risk Solve the Equity Premium Puzzle? *The Review of Economics and Statistics, Vol. 84, No. 2*, ss. 376-380.

Lillestøl, J. (1998). *Fat and skew: Can NIG cure? On the prospects of using the Normal Inverse Gaussian distribution in finance.* Bergen: Institute of Finance and Management Science, The Norwegian School of Economics and Business Administration.

Lucas, R. E. (1978). Asset Prices in an Exchange Economy. *Econometrica, Vol. 46, No. 6*, ss. 1429-1445.

Luethi, D., & Breymann, W. (2013). *ghyp: A package on the generalized hyperbolic distribution and its special cases. R package version 1.5.6.* Retrieved from http://CRAN.R-project.org/package=ghyp

Lütkepohl, H. (2006). *New Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag.

Macrobond. (2015). *Macrobond Financial. [Online]*. Available at: Subscription service.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business, Vol. 36, No. 4*, ss. 394-419.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhy A: The Indian Journal of Statistics, Vol. 36, Series B, Pt. 2*, ss. 115-128.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools.* Princeton: Princeton University Press.

Mehra, R., & Prescott, E. C. (1985). The Equity Premium - A Puzzle. *Journal of Monetary Economics, Vol. 15, No. 2*, ss. 145-161.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2014). *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*. Retrieved from R package version 1.6-4: http://CRAN.R-project.org/package=e1071

Miller, M. B. (2014). *Mathematics & Statistics for Financial Risk Management, Second edition.* Hoboken: John Whiley & Sons, Inc.

Mossin, J. (1969). A Note on Uncertainty and Preferences in a Temporal Context. *American Economic Review Vol. 59, No. 1*, ss. 172-174.

MSCI. (2015a). *MSCI*. Retrieved from https://www.msci.com/end-of-day-data-country

MSCI. (2015b). *MSCI Index Calculation Methodology.* Retrieved from https://app.msci.com/eqb/methodology/meth_docs/MSCI_Feb15_IndexCalcMethodology.pdf

Pouzat, C. (2012). *STAR: Spike Train Analysis with R. R package version 0.3-7*. Retrieved from http://CRAN.R-project.org/package=STAR

R Core Team. (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from http://www.R-project.org/

Ranaldo, A., & Favre, L. (2005). *How to Price Hedge Funds: From Two- to Four-Moment CAPM.* UBS Research Paper.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis, Third edition.* Belmont: Thomson.

Rizzo, M. L., & Szekely, G. J. (2005). A New Test for Multivariate Normality. *Journal of Multivariate Analysis, Vol. 93, No. 1*, ss. 58-80.

Rizzo, M. L., & Szekely, G. J. (2014). *energy: E-statistics (energy statistics). R-package version 1.6.2.* Retrieved from http://CRAN.R-project.org/package=energy

Royston, J. P. (1983). Some Techniques for Assessing Multivariate Normality Based on the Shapiro-Wilk W. *Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 32, No. 2*, ss. 121-133.

Schwert, G. W. (2011). *Stock Volatility During the Recent Financial Crisis, Working Paper 16976.* Cambridge: National Bureau of Economic Research (NBER).

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika, Vol. 52, No. 3/4*, ss. 591-611.

Shiller, R. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review, Vol. 71, No. 3*, ss. 421-436.

Statistics Norway. (2015). *Statistikkbanken.* Retrieved from http://ssb.no/statistikkbanken

van Binsbergen, J., Brandt, M., & Koijen, R. (2012). On the Timing and Pricing of Dividends. *American Economic Review, Vol. 102, No. 4*, ss. 1596-1618.

Venter, J. H., & de Jongh, P. J. (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk, Vol. 4*, ss. 1-24.