Are GARCH models necessary for Expected Shortfall?

Erik Berggren

Bachelor thesis[1] at the Department of Statistics

January 2017

Abstract

Following the Basel Committee on Banking Supervision's decision to move from Value at Risk to Expected Shortfall, risk managers will have to alter their methods for reporting risk. This paper sheds light on the question of which volatility models and distributional assumptions that works best for this new method of risk measurement by evaluating forecasts for the Swedish index OMXS30. The empirical results indicate that the choice of model becomes less important for Expected Shortfall than for Value at Risk, and that the Student's t distributed Value at Risk model improves accuracy compared to a normally distributed model. The EWMA model, proposed by the RiskMetrics Group, as well as the IGARCH and GJR-GARCH models generates adequate Value at Risk estimates at the 97.5 percent confidence level.

KEYWORDS: Forecasting, Backtesting, Value at Risk, Expected Shortfall, GARCH models

1. Introduction

The aftermath of the recent financial crisis has led to major developments in the area of risk measurement and risk reporting for large financial institutions. A widely used method of quantifying market risk is the Value at Risk (VaR) framework, proposed by JP Morgan in 1994, which has been the conventional structure of reporting risk since the Basel Committee on Banking Supervision (BCBS) introduced their Basel II accord of 1999. Since then, banks are obliged to report daily VaR figures to disclose their exposure to the market with a 99 % percent confidence level. It is intuitively straightforward to interpret the VaR figure as an expected loss over a target horizon that will not be exceeded with a given confidence level. However, Artzner et al. (1997, 1999) criticised VaR for not being subadditive, i.e. the sum of the portfolio VaR may be greater than the sum of the individual VaR figures of the assets in the portfolio. Hence, the VaR framework fails to encourage portfolio diversification and thereby violates the classic conclusion of Markowitz (1952).

Delbaen (2002) and Artzner et al. (1997) developed the Expected Shortfall (ES) risk measure, which is defined as the expected loss given an exceedance of the VaR. Unlike VaR, ES is subadditive and Acerbi and Tasche (2001) describe it as a natural coherent alternative to VaR. Its usefulness for risk management is further discussed in Acerbi et al. (2001), in Inui and Kijima (2005) and in Yamai and Yoshiba (2005), where the authors advise risk managers to combine the two measures for a more comprehensive risk monitoring. The current Basel III accord has since 2012 been revised by the BCBS and in the 'fundamental review of the trading book' (FRTB) they now opt to replace Value at Risk with Expected Shortfall. The FRTB accord is due to be implemented in 2018 and includes a switch from a 99 to 97.7 percent confidence level for ES reporting. However, ES has proven not to be as easily backtested as VaR. Gneiting (2011) proved that ES does not possess the mathematical property of elicitability, as opposed to VaR, which raised questions on whether ES actually could be backtested. Acerbi and Szekely (2014) nonetheless conclude that elicitability in fact has to do with model *selection* and not with model *testing*, and therefore is irrelevant for the choice of a regulatory risk standard. They also introduce three new model-free and non-parametric tests for ES, which they show to be more powerful than the Basel VaR

1

tests. This paper will apply their second test as it is easily implemented and comes with clearly defined critical values.

The current Basel III regulation employs a traffic light system that determines the level of capital requirements banks are to fulfil. Dardac and Grigore (2011) discuss the various challenges of selecting an appropriate VaR methodology since different methods lead to different capital requirements. Pérignon et al. (2008) found that banks tend to be very cautious in their risk reporting and subsequently overstate their VaR, partly due to widespread underestimation of diversification effects. However, whilst many studies have focused on finding the optimal volatility model for VaR, there are fewer studies that probe the same question for ES. Angelidis and Degiannakis (2007) examine a selection of GARCH models for their usefulness in a VaR/ES setting and find the APARCH and EGARCH models with normal innovations to generate adequate forecasts for the S&P 500. Researchers seldom find a volatility model that always dominate other models but comprehensive studies by Brooks and Persand (2003) and Hansen and Lunde (2005) conclude that asymmetric models tend to generate more accurate VaR estimates than symmetric models. The picture is nonetheless far from clear as studies by Angelidis et al. (2004) as well as Orhan and Köksal (2012) find the simple ARCH model with leptokurtic innovations to generate the best VaR forecasts. As for the error term innovations, Wilhelmsson (2006) finds support for the use of leptokurtic distributions that allows for fat tails in asset returns rather than normal innovations. Whether these improvements are significant or not is disputed in Liu and Hung (2010), where the authors argue that the choice of model is more important than the error term distribution. A risk manager can nevertheless not only use fat-tailed distribution in the volatility models but also in the VaR framework itself. Lin and Shen (2006) therefore propose a Student's *t* distributed VaR, which they conclude improves accuracy and is more suitable than a normally distributed VaR.

This study builds on earlier research within the field and tests nine different volatility models with four different error term distributions as well as testing them in a VaR/ES forecasting setting as an attempt to find an adequate method that complies with the latest Basel accords. All models are examined using a

comprehensive data set of closing price data on the Swedish index OMXS30 for a period of more than 20 years, thereby covering both the Dotcom bubble and the recent financial crisis. To the best of my knowledge, there are no previous papers that apply this set of models and error term distributions to such a long period of the selected data and evaluate the models under both traditional VaR backtests as well as a backtest for ES. The paper is organised into five sections, of which this is the introduction. Section 2 presents the volatility models and their distributions followed by Section 3 where the theoretical framework of VaR and ES is presented together with the tests used to evaluate the forecasts. Section 4 presents descriptive statistics of the data set, the estimation procedure and the empirical results of the study, whereas Section 5 concludes.

2. Volatility models

Let $y_t = \ln(P_t/P_{t-1})$ be the continuously compounded return series, where P_t is the closing price at trading day t. The return series is assumed to follow the stochastic process

$$y_t = \mu + \varepsilon_t$$

where μ is the mean and ε_t is the time-varying unpredictable innovation process given as

$$\varepsilon_t = z_t \sigma_t.$$

The term z_t is a sequence of independently and identically distributed variables with zero mean and unit variance and lastly, σ_t is the conditional standard deviation realised from the volatility models considered below.

Engle (1982) was first to develop a model that capture the property of time varying variance in returns. His autoregressive conditional heteroscedasticity model, ARCH(p) is defined as a linear function of the past p squared innovations and expressed as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \tag{1}$$

3

under the assumption that ω and α_i are strictly positive for $i = 1, ..., p$ in order to guarantee positive conditional variance. Bollerslev (1986) developed a generalised ARCH model, which in addition also takes past values of the conditional variance into account. His GARCH(p,q) model is given as

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

where $\omega > 0$, $\alpha_i \geq 0$ for $i = 1, ..., p$ and $\beta_j \geq 0$ for $j = 1, ..., q$. The process is covariance stationary if $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ and the unconditional variance is then equal to $\sigma^2 = \omega / \left(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j \right)$. However, Engle and Bollerslev (1986) noted that the sum of the estimated coefficients α_i and β_j often end up being close to unity and therefore proposed an integrated GARCH (IGARCH) where $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$ by construction. This will however results in a process that is not covariance stationary, but Nelson (1990) showed that the IGARCH(1,1) with a positive drift is still strictly stationary. The IGARCH(1,1) is closely related to the Exponentially Weighted Moving Average (EWMA) that was proposed by the company RiskMetrics in their original VaR methodology. Its volatility forecast is a weighted average of the previous periods variance forecast and the current period's squared return, without an intercept. The model is expressed as $\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda)\varepsilon_{t-i}^2$ where RiskMetrics propose $\lambda = 0.94$. Notably, the EWMA model does not require ML estimation and will serve as a benchmark model in this paper.

Some of the stylised facts of financial time series such as fat tailed returns and volatility clustering are successfully captured by the above models, but one drawback with the presented models is that they fail to capture potential asymmetries in the time series. As Black (1976) noted, there may be a leverage effect in asset returns meaning that volatility tend to increase after bad news, ($\varepsilon_t < 0$), and decrease after good news, ($\varepsilon_t > 0$). Schwert (1989) found that volatility clusters are more prominent during periods of economic recession and this set the tone for a new class of GARCH models that not only took the magnitude of the innovations, ε_t, but also the sign in front of it.

Engle (1990) proposed an asymmetric GARCH model to account for asymmetric effects of negative and positive innovations. The AGARCH is, in Sheppard (2013), expressed as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i(\varepsilon_{t-i} - \gamma)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (3)$$

where negative values of γ implies that positive shocks will result in smaller increases in future volatility than negative shocks of the same absolute magnitude. Apart from the leverage coefficient, similar restrictions as those for the GARCH model are imposed on the AGARCH in order to ensure positive variance. The AGARCH is closely related to the non-linear asymmetric GARCH, or the NAGARCH, proposed by Engle and Ng (1993). In the NAGARCH model, the magnitude of the leverage effect is also dependent on the lagged conditional variance, and the model is expressed as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \left(\varepsilon_{t-i} - \gamma \sqrt{\sigma_{t-j}^2} \right)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \qquad (4)$$

The most popular model to capture asymmetric effects is nonetheless the exponential GARCH, or the EGARCH, by Nelson (1991). The EGARCH differs substantially from the other GARCH models as it models the natural logarithm of the variance, rather than the variance directly. The structural form of the EGARCH in Sheppard (2013) also includes a multiplicative indicator variable to signal if returns are negative, thus accounting for asymmetries in the return series. The EGARCH(p,r,q) is expressed by Sheppard (2013) as

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{p} \alpha_i \left(\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{k=1}^{r} \gamma_k \frac{\varepsilon_{t-k}}{\sigma_{t-k}} I_{t-k} + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) \qquad (5)$$

where I_{t-k} is the indicator variable signalling whether the return is positive or negative. Formally the indicator is expressed as $I_t = \begin{cases} 1 & if \ y_t < 0 \\ 0 & otherwise \end{cases}$. The first

term, $\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} = |e_{t-i}| - \sqrt{\frac{2}{\pi}}$ is the absolute value of a normal random

variable, $|e_{t-i}|$, minus its expectation, $\sqrt{\frac{2}{\pi}}$, leading to a mean zero shock. The

second term, e_{t-k}, is also a mean zero shock and the final term is the lagged log variance.. The leverage coefficient, γ_k, is normally estimated to be less than zero and represents the rise in volatility following negative shocks. Since the EGARCH models the logarithm of the variance, the conditional variance can never be negative and the assumption of positive parameters is therefore no longer necessary.

Similar to the EGARCH, the GJR-GARCH named after the authors who introduced it, Glosten, Jagannathan and Runkle (1993), also includes the indicator variable used in the EGARCH. However, the GJR-GARCH models the variance directly instead of using natural logarithms, which makes it easy to implement. The GJR-GARCH is expressed as

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i} + \sum_{k=1}^{r} \gamma_k \varepsilon_{t-k}^2 I_{t-k} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \qquad (6)$$

where the parameters have to be restricted in order to ensure positive variance, just as for the GARCH model. These restrictions however, are difficult to define for a complete GJR-GARCH(*p,r,q*) but are, according to Sheppard (2013a) simpler to define for a GJR-GARCH(1,1,1). The conditional variance will then be positive if $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_1 + \gamma_1 \geq 0$ and $\beta_1 \geq 0$. The GJR-GARCH(1,1,1), under conditionally normal innovations, will also be covariance stationary if the parameter restrictions are satisfied and $\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1 < 1$.

Finally, the asymmetric power ARCH model (APARCH), by Ding, Engle and Granger (1993) presented a new technique to model the 'long memory' property of asset returns. As suggested by the efficient market hypothesis, returns themselves contain little or no serial correlation, but Taylor (1986) found that the absolute value of asset returns, $|r_t|$ contain significant positive serial correlation over long lags. Even the power transformations of the absolute values, $|r_t|^d$ have been found to exhibit quite high autocorrelations for long lags, which argues against the use of squared innovations as assumed in traditional ARCH specifications according to Ding et al. (1993). The authors therefore propose a Box-Cox power transformation of the conditional standard deviation process and can thus directly parameterise the non-linearity of the conditional variance with

the estimation of a new parameter, δ. This allows for greater flexibility in the memory of the model, without having to estimate more parameters for lagged innovations. The APARCH model is expressed as

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta \qquad (7)$$

where $\omega > 0$, $\delta \geq 0$, $\alpha_i \geq 0$ for $i = 1, \ldots, p$, $-1 \leq \gamma_i \leq 1$ for $= 1, \ldots, p$ and $\beta_j \geq 0$ for $j = 1, \ldots, q$ in order to ensure positive variance. The APARCH model also nests a variety of GARCH models as it can for example be reduced to a GARCH(p,q) by letting $\delta = 2$ and $\gamma_i = 0$.

As for the distributional assumptions imposed on ε_t in the innovation process, Engle (1982) assumed the standard normal distribution, which is described as

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{\varepsilon_t^2}{2}\right)}. \qquad (8)$$

Even though the assumption of conditionally normal innovations leads to strongly consistent parameters, there are reasons to assume other distributions than the normal since the empirical distribution of stock returns is thick tailed. The thick tails are allowed for in the Student's t distribution assumed by Bollerslev (1987), which is given as

$$f(\varepsilon_t, v, \sigma_t^2) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)\sigma_t^2}} (1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2})^{-\frac{(v+1)}{2}} \qquad (9)$$

where v is the degrees of freedom under the assumption $v > 2$ and $\Gamma(\cdot)$ is the gamma function. By making distributional assumptions that are more in line with empirical observations, the parameter estimates of the volatility process improves both precision and efficiency. The Generalised Error Distribution (GED), introduced by Subottin (1923) and applied in an ARCH framework by Nelson (1991), offers flexibility in the sense that it can easily transform a Normal density function into a leptokurtic (fat-tailed) or platykurtic (thin-tailed) distribution by altering v. Its density function is expressed as

$$f(\varepsilon_t, v, \sigma_t^2) = \frac{v \, exp\left(-\frac{1}{2}\left|\frac{\varepsilon_t}{\sigma_t}\right|^v\right)}{\lambda 2^{\frac{v+1}{v}}\Gamma\left(\frac{1}{v}\right)} \qquad (10)$$

$$\lambda = \left(\frac{2^{-\frac{1}{\nu}} \Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)} \right)^{\frac{1}{2}} \tag{11}$$

where $\nu = 2$ implies a normal distribution, $\nu < 2$ gives a platykurtic distribution and $\nu > 2$ leads to a leptokurtic distribution. However, the empirical distribution of stock returns often exhibit a positive or a negative skewness, i.e. an asymmetry in the density function indicating that movements in one direction is more frequently observed than movements in the other direction. Even though the unconditional distribution of the returns may exhibit small or no skewness, the conditional skewness may well be present in times of financial distress. Hence, Hansen (1994) extended the standardised Student's t distribution to allow for asymmetries. His skewed t distribution is given as

$$f(\varepsilon_t, \nu, \sigma_t^2, \lambda) = \begin{cases} bc\left(1 + \frac{1}{\nu - 2}\left(\frac{b\varepsilon_t + \sigma_t a}{\sigma_t(1 - \lambda)}\right)^2\right)^{-\frac{\nu+1}{2}}, & \varepsilon_t < -a/b \\ bc\left(1 + \frac{1}{\nu - 2}\left(\frac{b\varepsilon_t + \sigma_t a}{\sigma_t(1 + \lambda)}\right)^2\right)^{-\frac{\nu+1}{2}}, & \varepsilon_t \geq -a/b \end{cases} \tag{12}$$

where $2 < \nu < \infty$ and $-1 < \lambda < 1$. The constants a, b and c are given by

$$a = 4\lambda c\left(\frac{\nu - 2}{\nu - 1}\right), \tag{13}$$
$$b^2 = 1 + 3\lambda^2 - a^2$$

and

$$c = \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\sqrt{\pi(\nu - 2)}\Gamma\left(\frac{\nu}{2}\right)}. \tag{14}$$

Kurtosis and skewness is controlled by the two shape parameters, ν and λ, respectively. Hansen's skewed t distribution is the same as Student's t distribution when the skewness parameter $\lambda = 0$. All models and distributions are given in accordance to the MFE Toolbox for Matlab by Sheppard (2013b).

3. Theoretical framework

As mentioned earlier, the volatility forecasts from the models in the previous section will be applied to a Value at Risk and Expected Shortfall framework and backtested with appropriate tests. VaR is defined as a *quantile* in the distribution

of profit and loss that is expected to be exceeded only with a certain probability α, formally expressed as

$$\alpha = \int_{-\infty}^{-VaR(\alpha)} f_q(x)dx \qquad (15)$$

where $f_q(x)$ is the assumed distribution of profit and losses. Having obtained the parameter estimates from the volatility models, it is straightforward to calculate the daily VaR figure using the following equation:

$$VaR_{\alpha,t} = \sigma_t \Phi^{-1}(\alpha) \qquad (16)$$

where σ_t is the standard deviation and $\Phi^{-1}(\alpha)$ is the inverse of the normal cdf at probability α. The distribution of profit and loss may however violate the assumption of normality, as is not unlikely when dealing with financial data. Lin and Shen (2006) examined several major equity indices and suggest the use of Student's t distribution for estimating VaR as it improves the estimation of VaR, especially at high confidence levels. The t distributed VaR is computed as

$$VaR_{\alpha,t} = \sqrt{\frac{\nu - 2}{\nu}} \sigma_t t_{\alpha,\nu}^{-1} \qquad (17)$$

where ν is the degrees of freedom based on the excess kurtosis, $\kappa - 3 = \frac{6}{\nu - 4}$, when $\nu > 4$. Notably, kurtosis is not defined when $\nu \leq 4$, and $t_{\alpha,\nu}^{-1}$ is the inverse of the t distributed cdf . Generally, the degrees of freedom have to be estimated by ML, but when $\nu > 4$ the relationship between excess kurtosis and the degrees of freedom can be used as in the following equation to calculate ν:

$$\nu = \frac{4\kappa - 6}{\kappa - 3}. \qquad (18)$$

Again, Expected Shortfall is defined as the conditional expected loss given an exceedance of VaR, and formally expressed as

$$ES_\alpha = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_x \, dx. \qquad (19)$$

Similarly as with the computation of the VaR figure, computing the ES number is straightforward having already obtained the volatility forecast. Under the normal distribution, ES is calculated as

$$ES_{\alpha,t} = \sigma_t \frac{\phi(\Phi^{-1}(\alpha))}{1 - \alpha} \tag{20}$$

where ϕ is the density function of the standard normal probability function. Under the assumption of Student's t distributed losses, ES is calculated as

$$ES_{\alpha,t} = \sqrt{\frac{\nu - 2}{\nu}} \sigma_t \frac{g_\nu(t_{\alpha,\nu}^{-1})}{1 - \alpha} \left(\frac{\nu + t_{\alpha,\nu}^2}{\nu - 1} \right) \tag{21}$$

where g_ν is the density function of a standardised t distributed variable and $t_{\alpha,\nu}^{-1}$ is the α quantile of the same distribution.

3.1 Evaluation of VaR and ES forecasts

Finding satisfactory forecast models for VaR estimates requires a method for evaluating the predictions ex-post. This paper will use two tests to evaluate the statistical properties of the VaR estimates, and one test to examine the models performance in the ES setting. These tests however, cannot be used to rank the models since a greater p-value does not imply a significantly superior performance. Instead, this paper is more focused on finding a volatility model for VaR and ES that complies with the FRTB accord, rather than a model that minimises some loss function. Hence, tests and methods such as those developed by Lopez (1999), Sarma et al. (2003) and Hansen (2005) will not be used.

Henceforth, out-of-sample returns will be labelled according to an indicator function in order to define whether they exceeded the VaR estimate or not. The indicator variable is constructed as

$$\eta_t = \begin{cases} 1 & if \ y_t < -VaR_t \\ 0 & if \ y_t \geq -VaR_t \end{cases} \tag{22}$$

where 1 indicates a violation and 0 indicates a return less than the VaR. The total number of violations is thereafter divided by the total number of VaR estimates, thus obtaining the empirical VaR size.

The conventional test to examine whether the empirical number of violations is in line with the expected number of violations was developed by Kupiec (1995). The null hypothesis in the unconditional coverage test is that the number of violations is equal to its expectancy and the test statistic is given as

$$LR_{uc} = 2\ln\left[\left(1 - \frac{F}{T}\right)^{T-F}\left(\frac{F}{T}\right)^{F}\right] - 2\ln[(1 - \alpha)^{T-F}\alpha^{F}] \sim \chi^2_{(1)} \qquad (23)$$

where T is the number of out-of-sample estimates and F the observed number of violations. Hence, F/T is the empirical VaR size, which follows the binominal distribution, i.e. $F \sim B(T, \alpha)$. Rejection of the null hypothesis thus implies that the model generates too many or too few violations.

Ideally, a violation in time *t* does not reveal any information about the likelihood of a violation in time *t* +1, i.e. violations occurs independently of each other. Christoffersen (1998) developed a test to detect clusters of violations, where clusters are defined as two or more directly subsequent violations. The test statistic is given as   (24)

$$LR_{ind} = -2\ln[(1 - \alpha)^{T-F}\alpha^{F}] + 2\ln[1 - \pi_{01})^{\eta_{00}}\pi_{01}^{\eta_{01}}(1 - \pi_{11})^{\eta_{10}}\pi_{11}^{\eta_{11}} \sim \chi^2_{(1)}$$

where η_{ij} is the number of observations with the value i followed by j for $i, j = 0, 1$ and $\pi_{ij} = \frac{\eta_{ij}}{\sum_j \eta_{ij}}$ are the corresponding probabilities. Rejection of the null hypothesis thus implies that the violations cannot be viewed as independent. Naturally, a drawback with the test is that it fails to detect clusters with violations every other day, but the above two tests together ought to provide a suitable ground for evaluating the models usefulness for VaR.

Backtesting ES it not intuitively as straightforward as backtesting VaR since it measures losses beyond VaR. Neither has ES historically been given as much attention as VaR, although there are methods that under certain circumstances can be able to backtest ES. McNeil and Frey (2000) suggest a 'residual approach', Berkowitz (2001) propose a method referred to as the Guassian approach and Kerkhof and Melenberg (2004) suggested a method denoted as the functional delta method. These methods however, require parametric assumptions and large samples, which makes them inappropriate for regulatory purposes. Acerbi and Szekely (2014) introduce three different non-parametric methods to backtest ES that can be used to select appropriate models without the need for large samples. This paper will use apply their second test, as the authors found that the critical value of the test is very stable and therefore does not require Monte Carlo

simulations of the test statistics in order to find the critical value. Their test is based on a test statistic defined as

$$Z = \frac{1}{\alpha T} \sum_{t=1}^{T} \frac{y_t \eta_t}{ES_{\alpha,t}} + 1 \qquad (25)$$

where the numerator contains the loss beyond VaR when a violation occurs and the denominator contains the forecasted ES figure. Hence, the test statistic calculates the average ratio of the loss on a violation day and the estimated ES. Under the null hypothesis, $E_{H_0}[Z] = 0$ and the ES estimate is assumed to belong to the same (true) distribution as the stochastic loss variable. The authors simulated critical values from a potential loss distribution using the Student's t distributions with different degrees of freedom varying from very thick tails, $v = 3$, to the case of a normal distribution, $v = \infty$. Following their simulation, they propose critical values of $Z_{crit,0.05} = -0.70$ for the 5 percent significance level and $Z_{crit,0.01} = -1.80$ for the 1 percent significance level. Rejection of the null hypothesis thus implies that the predicted ES density is believed to come from another distribution than the actual loss distribution.

4. Empirical results

To evaluate the volatility models, one-step-ahead out-of-sample VaR and ES forecasts are generated for the Swedish equity index OMXS30. This paper will use the 97.5 percent confidence level for all VaR and ES estimates in line with the latest guidelines of the Basel Committee. The data is collected from Nasdaq OMX and covers the period January 1st 1996 to November 1st 2016, yielding more than 5000 observations of closing price data. Log returns are computed and displayed in Figure 1 below.
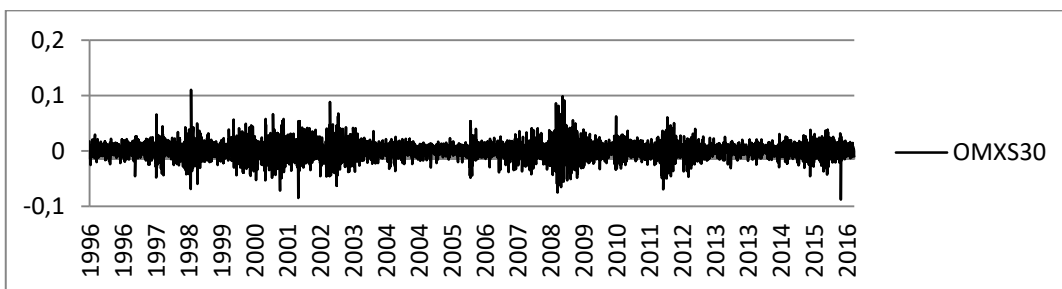

Figure 1 – Daily log returns

Clearly, there are prominent volatility clusters in the return series and especially around the time of the Dotcom bubble in the early 2000 and the recent financial crisis of 2007 – 2009. Table 1 reports descriptive statistics of the return series as well as the Jarque-Bera value.

Table 1 – Descriptive statistics

	Mean	St.d.	Min	Max	Skewness	Kurtosis	Jarque-Bera
OMXS30	0	0,015	-0,09	0,11	0,04	6,66	2922

The return series is positively skewed and exhibits significant excess kurtosis thus the null hypothesis of normality is clearly rejected at any level of significance following the large Jarque-Bera value. Forecasts are produced using a rolling window estimation method where parameters are re-estimated each day. The length of the estimation window is found to play an important role for the forecasting result, but there is no clear method for deciding the optimal window length. Engle et al. (1993) noted that "some restrictions on the forecasting sample may be profitable". Angelidis et al. (2004) conclude that the forecasting results improve when the sample size gets smaller since a smaller size captures only the latest market movements and the estimates are thus less based on longer volatility trends. Hence, the window length in this paper is set to 500 in order for the parameters to be stable yet reactive to market movements. Further studies are nonetheless encouraged to investigate the optimal window length more closely.

As for the order of *p* and *q* (and *r* when applicable) in the volatility models, the conventional method is to choose order after looking at some information criterion. However, So and Yu (2006) conclude that the best fitted model according to Akaike's Information Criterion (AIC) and Schwartz Bayesian Information Criterion (BIC) does not necessarily lead to better VaR estimates. Furthermore, Degiannakis and Xekalaki (2006) showed that the best-performing model for volatility forecasting could not be selected according to any in-sample model selection criterion. Therefore, all models in this paper have $p = q = 1$, (and *r,* when applicable) in line with convention in the GARCH forecasting literature. Finally, all GARCH models are estimated by ML using the MFE Toolbox for Matlab by Sheppard (2013b).

Table 2 – Empirical VaR and ES results for OMXS30

Normal VaR / ES					Student's t VaR / ES					
		Emp. VaR	Kupiec	Ind. test	ES test			Emp. VaR	Kupiec	Ind. test	ES test
ARCH	N	3,64%	0,00	0,07	-0,26	ARCH	N	3,64%	0,00	0,07	-0,09
	t	3,59%	0,00	0,06	-0,25		t	3,49%	0,00	0,19	-0,10
	GED	3,66%	0,00	0,08	-0,26		GED	3,64%	0,00	0,07	-0,10
	Skew t	3,55%	0,00	0,12	-0,24		Skew t	3,43%	0,00	0,31	-0,09
GARCH	N	3,38%	0,00	0,50	-0,11	GARCH	N	3,17%	0,00	0,32	0,01
	t	3,32%	0,00	0,44	-0,11		t	3,15%	0,01	0,30	0,01
	GED	3,36%	0,00	0,48	-0,11		GED	3,17%	0,00	0,32	0,01
*	Skew t	3,19%	0,00	0,27	-0,13	*	Skew t	3,04%	0,03	0,18	0,00
IGARCH	N	2,94%	0,06	0,36	-0,11	IGARCH	N	2,79%	0,21	0,50	0,00
	t	2,90%	0,09	0,61	-0,11		t	2,71%	0,37	0,77	-0,01
	GED	2,92%	0,07	0,63	-0,11		GED	2,81%	0,18	0,52	0,00
	Skew t	2,85%	0,13	0,30	-0,10		Skew t	2,66%	0,47	0,39	-0,01
AGARCH	N	3,17%	0,00	0,32	-0,11	AGARCH	N	2,96%	0,05	0,19	0,00
	t	3,13%	0,01	0,29	-0,11		t	2,94%	0,06	0,18	0,00
	GED	3,15%	0,01	0,30	-0,11		GED	2,94%	0,06	0,18	0,00
	Skew t	3,07%	0,02	0,25	-0,11		Skew t	2,92%	0,07	0,17	0,01
NAGARCH	N	3,21%	0,00	0,61	-0,12	NAGARCH	N	3,00%	0,03	0,41	0,00
	t	3,19%	0,00	0,59	-0,12		t	2,92%	0,07	0,34	-0,01
	GED	3,26%	0,00	0,66	-0,12		GED	2,94%	0,06	0,36	-0,01
	Skew t	3,17%	0,00	0,57	-0,12		Skew t	2,85%	0,13	0,30	-0,01
EGARCH**	N	3,85%	0,00	0,07	-0,18	EGARCH**	N	3,69%	0,00	0,04	0,00
	t	3,78%	0,00	0,06	-0,19		t	3,57%	0,00	0,06	-0,01
	GED	3,81%	0,00	0,06	-0,18		GED	3,62%	0,00	0,07	0,00
	Skew t	3,83%	0,00	0,07	-0,19		Skew t	3,57%	0,00	0,06	-0,01
GJR	N	3,09%	0,01	0,44	-0,12	GJR	N	2,85%	0,13	0,64	-0,01
	t	3,02%	0,03	0,49	-0,12		t	2,77%	0,24	0,73	-0,02
	GED	3,07%	0,02	0,45	-0,12		GED	2,79%	0,21	0,70	-0,02
	Skew t	2,90%	0,09	0,60	-0,12		Skew t	2,75%	0,28	0,75	-0,02
APARCH	N	3,34%	0,00	0,90	-0,12	APARCH	N	3,09%	0,01	0,80	-0,01
	t	3,30%	0,00	0,95	-0,13		t	3,05%	0,02	0,77	-0,01
	GED	3,26%	0,00	0,99	-0,13		GED	3,09%	0,01	0,81	-0,01
	Skew t	3,19%	0,00	0,93	-0,13		Skew t	3,09%	0,01	0,81	0,00
EWMA	-	2,28%	0,33	0,65	-0,02	EWMA	-	1,95%	0,01	0,92	-0,02

*= estimated using a 750 observation window, **= estimated using a 1000 observation window
The two panels in Table (1) reports the results of the normal VaR/ES (to the left) and the Student's t distributed VaR/Es (to the right). The empirical VaR size for each model estimated under the four error term distributions are reported in the first column. The second column reports the p values from Kupiec's test of unconditional coverage and the third column reports the p values from Christoffersen's test of independence. The fourth column reports the test statistic from the ES backtest by Acerbi & Szekely (2014).

Table 2 reports the empirical VaR sizes, the p values from the Kupiec and the Christoffersen's test as well as the test statistics from the test by Acerbi and

Szekely, for the normally and the *t* distributed VaR/ES respectively. Notably, very few models produced accurate VaR estimates and all GARCH models, to various extents, underestimated the risk, as the empirical size is larger than the selected nominal size of 2.5 percent. This hypothesis is strengthened by the low *p* values from Kupiec's test of unconditional coverage, which indicate that the empirical violation ratio is significantly different from 2.5 percent for the majority of the models. As for the conditional coverage, the opposite is true as the overall large *p* values indicate that all models, except for the EGARCH with normal innovations under the *t*-distributed VaR, passed Christoffersen's independence test. The same is true for the Acerby and Szekely test for Expected Shortfall as there was no model with a Z statistic below -0.7, which is the critical value suggested by the authors. Hence, a risk manager could use any of the examined volatility models in this paper and still satisfy the Basel Committee's requirements for Expected Shortfall.

As for Value at Risk figures, the choice of model and distribution becomes more important. Overall, the *t*-distributed VaR dominated the normal VaR/ES as the empirical size was always lower for the VaR-t than for VaR-N. For some models (the IGARCH, AGARCH, NAGARCH and GJR-GARCH), the difference between passing and failing Kupiec's test would depend on the choice between VaR-N and VaR-t. From a risk managers perspective however, it appears that there are gains from using a *t* distributed VaR model rather than a normal VaR model. The difference was striking for the EWMA model as the empirical VaR-N figures was very close to 2.5 percent whereas the same figure for VaR-t was significantly below 2.5 percent, and thus the only model that overestimated the risk.

The accuracy of the GARCH models varied greatly, but the IGARCH and the GJR-GARCH produced the most accurate VaR figures. These two models are the only GARCH specifications that passed Kupiec's test both for VaR-N and VaR-t. Asymmetrical models do not appear to dominate symmetrical according to the results in this paper, although the performances of the original ARCH models were particularly dissatisfying, and the GARCH model did not produce accurate VaR figures either. These results indicate that general ARCH structures do not

work well over a longer period and that allowing for asymmetries in the model may improve forecasts, but also that not necessarily all asymmetrical models are good enough. The good performance of the IGARCH and the EWMA model indicate that a certain ratio of lagged squared returns and the lagged variance successfully models volatility, and that the ratio does not necessarily have to be altered at all. This is very good news for risk manager since the EWMA model is very easy to implement and does not require any distributional assumptions of the innovations, as the model is not estimated by ML.

As for the distributional assumptions on the innovations, leptokurtic innovations thoroughly dominated normal and GED innovations. The skewed t distribution worked particularly well and produced the most accurate VaR estimates for all but the APARCH VaR-t model. This indicates that there are gains from the allowance of conditional skewness in the return series. However, the skewed t distribution can impose an impeding effect on the usefulness of the models in times of quiet markets as the estimation could break down due violations of the distribution restrictions. Notably, the GARCH model under skewed t innovations had to be estimated using a 750 observation estimation window due to restriction violations, which caused the conditional variance to become negative. The exact reason for the breakdown is unclear but it occurred during a period of very low volatility, which indicates that usefulness of skewed t innovations may be dependent on the current market conditions. Overall, the t distribution suggested by Bollerslev (1987) did almost equally well and did not cause the models to break down at any point.

The results in this paper also suggest that technically more advanced EGARCH and APARCH models do not improve forecasts relative to simpler models. Both the EGARCH and the APARCH models transformations of the variance rather than the variance itself, which for this index and time period only appears to introduce redundant complexity to the modelling. This result is in line with Ané (2006), who conclude that the additional flexibility of the APARCH provide little, if any, improvement for risk management. However, the APARCH model produced VaR estimates that yielded very high p values from the Christoffersen test, which could indicate that the long-memory property of the APARCH makes

the model more reactive than other models and thus more prone to avoid clustered VaR violations. The EGARCH on the other hand, produced rather clustered VaR violations, as indicated by the low p values from the Christoffersen test of independence. Furthermore, the EGARCH, under certain leptokurtic innovations, had to be estimated using a longer estimation window in order to produce functional results. The exact reason for this is most likely related to the problem with the GARCH model, but the estimation window was set to 1000 observations for all EGARCH models in order to make them comparable for different innovation assumptions.

5. Concluding remarks

The purpose of this paper has been to evaluate VaR and ES estimates generated by a selection of GARCH models under four different error term distributions for the Swedish index OMXS30 over a period of more than 20 years of trading days. Empirical findings suggest that more complex models are not necessarily required for adequate ES estimates. For VaR estimates however, the EWMA model generated the most accurate forecasts followed by the IGARCH and the GJR-GARCH models. The results of this paper suggest that leptokurtic innovations are preferred over Gaussian error terms, and especially the skewed t distribution worked best for almost every model. The main take for risk managers is that adaption of Expected Shortfall instead of Value at Risk leads to relaxed demands of the choice of volatility models as the results of this study shows that all of the tested models works just fine for ES. Further research is nonetheless encouraged to investigate whether these results also holds for equities, commodities and other derivatives.

References

Acerbi, C., Nordio., C., Sirtori, C. 2001, "Expected Shortfall as a Tool for Financial Risk Management". *Working Paper*, https://arxiv.org/pdf/cond-mat/0102304v1.pdf.

Acerbi, C., Tasche, D. 2002, "On the coherence of expected shortfall", *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1487-1503.

Acerbi, C. Szekely, B. 2014. "Backtesting expected shortfall". *Risk Magazine*, December 2014.

Ané, T. 2006, "An analysis of the flexibility of Asymmetric Power GARCH models", *Computational Statistics and Data Analysis*, vol. 51, no. 2, pp. 1293-1311.

Angelidis, T., Benos, A., Degiannakis, S. 2004, "The use of GARCH models in VaR estimation". *Statistical Methodology* 1, pp. 105–128.

Angelidis, T., Degiannakis, S. 2007, "Backtesting VaR models: An Expected Shortfall Approach". No 701, *Working Paper*. University of Crete, Department of Economics

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. 1997, "Thinking Coherently". *Risk*, 10, pp. 68-71.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. 1999, "Coherent Measures of Risk". *Mathematical Finance*, 9, pp. 203-228.

Berkowitz, J. 2001, "Testing Density Forecasts, with Applications to Risk Management", *Journal of Business & Economic Statistics*, vol. 19, no. 4, pp. 465-474.

F. Black. 1976, "Studies of stock market volatility changes", *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, 1976, pp. 177–181.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroscedasticity". *Journal of Econometrics* 31, pp. 307–327.

Bollerslev, T. 1987, "A conditionally heteroskedastic time series model for speculative prices and rates of return", *The review of economics and statistics*, vol. 69, no. 3, pp. 542-547.

Brooks, C. & Persand, G. 2003, "The Effect of Asymmetries on Stock Index Return Value-at-Risk Estimates", *The Journal of Risk Finance*, vol. 4, no. 2, pp. 29-42.

Christoffersen, P. 1998, "Evaluating Interval Forecasts". *International Economic Review* 39, pp. 841-862.

Degiannakis, S. and Xekalaki, E, 2007, "Assessing the performance of a prediction error criterion model selection algorithm in the context of ARCH models", *Applied Financial Economics*, vol. 17, no. 2, pp. 149-171.

Ding, Z., Granger, C.W.J. and Engle, R.F. 1993, "A long memory property of stock market returns and a new model", *Journal of Empirical Finance*, vol. 1, no. 1, pp. 83-106.

Gneiting, T. 2011, "Making and Evaluating Point Forecasts", *Journal of the American Statistical Association* 106, (494).

Glosten, L.R., Jagannathan, R. and Runkle, D.E. 1993, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *The Journal of Finance*, vol. 48, no. 5, pp. 1779-1801.

Hansen, B.E. 1994, "Autoregressive Conditional Density Estimation", *International Economic Review*, vol. 35, no. 3, pp. 705-730.

Hansen, P.R., Lunde, A. 2005, "A forecast comparison of volatility models: does anything beat a GARCH (1,1)?", *Journal of Applied Econometrics*, vol. 20, no. 7, pp. 873-889.

Hansen, P.R. 2005, "A Test for Superior Predictive Ability". *Journal of Business and Economic Statistics*, 23, pp. 365-380.

Kerkhof, J. and Melenberg, B. 2004, "Backtesting for risk-based regulatory capital", *Journal of Banking and Finance*, vol. 28, no. 8, pp. 1845-1865.

Kupiec, P.H. 1995, "Techniques for verifying the accuracy of risk measurement models". Board of Governors of the Federal Reserve System (U.S.), *Finance and Economics Discussion Series*: 95-24.

The Basel Committee. Fundamental review of the trading book: a revised market risk framework, 2013. http://www.bis.org/publ/bcbs265.htm. [2017-02-09]

Dardac, N. and Grigore, A. 2011, "Modeling the Market Risk in the Context of the Basel III Acord", *Theoretical and Applied Economics*, vol. 11, pp. 5-20.

Engle, R.F. 1982, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". *Econometrica* 50, pp. 987–1007.

Engle, R.F. 1990, "Discussion: Stock Market Volatility and the Crash of '87," *Review of Financial Studies*, 3, pp. 103-106.

Engle, R.F., Hong, C.H., Kane, A., Noh, J. 1993, "Arbitrage valuation of variance forecasts with simulated options", *Advances in Futures and Options Research* 6 pp. 393–415.

Engle, R.F. and T. Bollerslev. 1986, "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5, pp. 1-50.

Engle, R.F. and V.K. Ng. 1993, "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, pp. 1749-1778.

Delbaen, F. 2002, "Coherent Risk Measures on General Probability Spaces". *Advances in Finance and Stochastics*. Essays in Honour of Dieter Sondermann, in: K. Sandmann and P. J. Schönbucher (eds.), Springer, pp. 1-38.

Inui, K.,Kijima, M. 2005, "On the Significance of Expected Shortfall as a Coherent Risk Measure". *Journal of Banking and Finance*, 29, pp. 853–864.

Lin, C., Shen, S. 2006, "Can the student-t distribution provide accurate value at risk?", *The Journal of Risk Finance*, vol. 7, no. 3, pp. 292-300.

Liu, H., Hung, J. 2010, "Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models", *Expert Systems with Applications*, vol. 37, no. 7, pp. 4928-4934.

Lopez, J.A. 1999, "Methods for Evaluating Value-at-Risk Estimates". *Federal Reserve Bank of New York, Economic Policy Review*, 2, pp. 3-17.

Markowitz, H. 1952, "Portfolio selection", *The Journal of Finance*, vol. 7, no. 1, pp. 77-91.

McNeil, A. J., Frey, R. 2000, "Estimation of tail-related risk measures for heteroscedastic Financial time series: an extreme value approach", *Journal of Empirical Finance* 7, pp. 217-300.

Nelson, D.B. 1990, "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, pp. 318-334.

Nelson, D.B. 1991, "Conditional heteroskedasticity in asset returns: a new approach", *Econometrica*, vol. 59, no. 2, pp. 347-370.

Orhan, M., Köksal, B. 2012, "A comparison of GARCH models for VaR estimation", *Expert Systems with Applications*, vol. 39, no. 3, pp. 3582- 3592.

Pérignon, C., Deng, Z.Y., Wang, Z.J. 2008, "Do banks overstate their Value-at-Risk?", *Journal of Banking and Finance*, vol. 32, no. 5, pp. 783-794.

Sarma, M., Thomas, S. and Shah, A. 2003, "Selection of VaR models". *Journal of Forecasting*, 22(4), pp. 337–358.

Schwert, G.W. 1989, "Why does stock market volatility change over time?", The *Journal of Finance*, vol. 44, no. 5, pp. 1115-1153.

Sheppard, K. 2013 a, Financial Econometrics Notes. University of Oxford: 2013, pp. 385-435.

Sheppard, K. 2013 b, MFE MATLAB Function Reference Financial Econometrics. University of Oxford: 2013, pp. 84-97.

So, M.K.P. and Yu, P.L.H. 2006. "Empirical analysis of GARCH models in Value at Risk Estimation". *Journal of International Markets, Institutions and Money*, 16(2), pp. 180-197.

Subbotin, M.T. 1923, "On the Law of Frequency of Error". *Matematicheskii Sbornik*, 31, pp. 296-301.

Taylor, S., 1986, Modelling financial time series, New York, John Wiley & Sons.

Wilhelmsson, A. 2006, "Garch forecasting performance under different distribution assumptions", *Journal of Forecasting*, vol. 25, no. 8, pp. 561-578.

Yamai, Y. and Yoshiba, T. 2005, "Value-at-risk Versus Expected Shortfall: A Practical Perspective". *Journal of Banking and Finance*, 29(4), pp. 997-1015.