Volatility Forecasts for Finnish Equities

Markus Ehrnrooth, 62834B
21.11.2007

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Engineering Physics and Mathematics
Systems Analysis Laboratory
## Contents

1 Introduction

2 Volatility and Volatility Modeling in a Nutshell
   2.1 Notation and Notions of Volatility .......................... 3
   2.2 Models for Volatility Forecasting ............................ 5
      2.2.1 Past Standard Deviations - based Forecasts .......... 7
      2.2.2 GARCH-Class Volatility Models .......................... 8
      2.2.3 Stochastic Volatility .................................... 12
      2.2.4 Options-Implied Volatility ............................... 12

3 Estimation of Volatility Models and Options-Implied Volatility
   3.1 GARCH Model Estimation ....................................... 14
   3.2 Stochastic Volatility Model Estimation ...................... 17
      3.2.1 Markov Chain Monte Carlo ............................... 17
      3.2.2 Gibbs Sampling .......................................... 18
      3.2.3 Gibbs Sampling and SV Model Estimation ............... 20
   3.3 Creating Options Implied Volatility Forecasts ............. 22

4 Volatility Forecast Evaluation ................................. 24

5 Model Estimation for Four Finnish Equities
   5.1 The Data ..................................................... 27
   5.2 GARCH Volatility Forecasts ................................. 30
   5.3 Forecasts Based on Warrant Implied Volatility ............. 35
   5.4 Comparison of Obtained Volatility Forecasts ............... 37

6 Conclusions ..................................................... 40

A Tests for Heteroscedasticity .................................. 42

B Price Graphs .................................................... 43

C Return Graphs ................................................... 44

D Graphs of Daily Implied Standard Deviation .................. 46
1 Introduction

Volatility forecasting is an important topic in financial economics and in financial applications. It helps answering many questions such as efficient asset allocation, risk management and asset pricing, specially derivatives pricing.

As a consequence, volatility modelling has received attention from academics and practitioners. Since the seminal paper of Engle (1982), many different volatility models have been developed and vast literature has emerged. Together with growing computing power – and partly as a consequence of it – research has resulted in deeper understanding and better predictability of volatility which in turn has contributed to the explosive growth in markets for complex financial products (Christoffersen and Diebold, 2000).

While researchers tend to agree that financial asset return volatility is highly predictable1, there is a lively debate on how this volatility should be modeled and predicted. There are many methods and variations of these in the field of volatility forecasting. The practitioner seeks to assess what model performs best for her purposes. In this study we consider short term forecasts for the daily return volatility of four Finnish equities with liquid warrants written on them; Neste Oil, Nokia, Sampo and UPM2.

Specifically, we construct the necessary models and compare two different volatility forecasting methods; a GARCH model and an options-based volatility forecast for each of the equities. The forecasts will be compared to out-of-sample data, and the forecasting performance of the models will be compared.

Although several studies have been done that compare the forecasting performance of different volatility models3, the author has not found preceeding

---

1There is an equally strong concensus among researchers that financial asset returns are approximately unpredictable. The claim of volatility predictability is made in most articles with a reference to Bollerslev, Engle and Nelson, (1994)

2These equities were the underlying assets of the most traded equity warrants on the OMX between September 2006 and March 2007

3See Poon and Granger (2003) for an extensive survey
studies with Finnish equities. This study will describe the estimation procedure of GARCH, Stochastic Volatility and options-implied volatility models, and the difficulties of each method. These sections will be beneficial for practitioners in the future.

The remainder of this study is organized as follows. Section 2 provides an overview of volatility, volatility models and volatility modelling. Section 3 discusses the estimation of GARCH-models, Stochastic Volatility models and the construction of options-based volatility forecasts. Special issues related to volatility forecasts are discussed in Section 4. Section 5 describes the practical model estimation and forecasting for the four equities and compares the obtained forecasts. Finally conclusions are made in Section 6.
2 Volatility and Volatility Modeling in a Nutshell

2.1 Notation and Notions of Volatility

Let $P_t$ be the price of an asset at time index $t$. The corresponding one-period simple return is

$$ R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}. \quad (1) $$

Correspondingly, denote the return over the discrete interval $[t-n, t] n > 0$ by $R_t[n]$

$$ R_t[n] = \frac{P_t}{P_{t-n}} - 1 = \frac{P_t - P_{t-n}}{P_{t-n}}. \quad (2) $$

Several other return measures than (1) can be considered, the probably most widely used return variable being the (one-period) log return of an asset:

$$ r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = p_t - p_{t-1} \quad (3) $$

where $p_t = \ln(P_t)$. Log returns are easier to work with than simple returns, especially when dealing with multi-period returns. To illustrate this, consider the $k$-period log return

$$ r_t[k] = \ln(1 + R_t[k]) = \ln[(1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1})] $$

$$ = \ln(1 + R_t) + \ln(1 + R_{t-1}) + \cdots + \ln(1 + R_{t-k+1}) $$

$$ = r_t + r_{t-1} + \cdots + r_{t-k+1} $$

$$ = p_t - p_{t-k}. $$

The multi-period log return can thus be simply expressed as the sum of the one-period log returns, or the difference between the logarithmic prices at the endpoints of the interval under consideration. In this study $p$ denotes the logarithmic price process of an asset, and $r$ is the corresponding return process as defined by (3).

Colloquially, volatility is used as a measure of the uncertainty of the return realized on an asset (see definition of volatility in Hull, 2003). Often,
“volatility”, refers to the standard deviation, \( \sigma \) of a set of return observations

\[
\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (r_i - \bar{r})^2
\]

where \( \bar{r} \) is the mean return. Hence, stock volatility is often considered to be the sample standard deviation of a time series of daily returns for the stock. The sample standard deviation, \( \hat{\sigma} \), represents the second moment characteristic of the sample. As such, it is not meaningful as a risk measure unless it is attached to a distribution or a pricing dynamic. When \( \sigma \) is used as a measure of uncertainty, most people, perhaps implicitly, have a normal distribution of stock returns in mind\(^4\).

In financial economics, volatility is often defined as the parameter scaling the fluctuations generated by the wiener process in a continuous-time diffusion price-model (Andersen et al., 2005). If the returns \( r_t \) arise from an underlying continuous-time (logarithmic) price process,

\[
dp = \mu dt + \sigma dz
\]

where \( \mu \) denotes the drift and \( dz \) denotes a standard Brownian motion, the spot volatility, or simply the volatility of the process is \( \sigma \).

Volatility cannot be directly observed. Suppose, for instance, that daily returns are considered. A measure of daily volatility is not directly observable from the returns, since there is only one observation in a trading day. Obviously, if intraday returns would be available, these could be used to estimate daily volatility. However, stock volatility is made up of intraday volatility and volatility between trading days, so the use of one to estimate the other deserves careful consideration.

There are several “stylized facts” about the volatility of asset returns (Poon and Granger, 2003). First, volatility tends to cluster, meaning that

\(^4\)This is consistent with most options pricing models such as the Black-Scholes model where the stock price \( S \) is assumed to follow the dynamic \( dS = \mu S dt + \sigma S dz \), where \( dz \) denotes a standard Brownian motion.
volatility tends to be high during some periods in time and low during others. Second, volatility tends to exhibit mean reversion so that it approaches its long-time average. Third, volatility jumps are rare, and volatility seems to evolve over time in a continuous fashion. Fourth, volatility reacts asymmetrically (i.e. differently) to large price drops and price increases. Fifth, volatility in financial markets exhibits strong comovement, spreading quickly between assets and geographical markets. Last but not least, it is known that the returns of risky assets follow fat-tailed distributions, rather than the normal distribution often assumed in price models.

There are several other possible measures of volatility, an overview of different volatility measures is given in Table 1

2.2 Models for Volatility Forecasting

The conditional mean and conditional variance of the (discrete) return process $r_t$ are needed in volatility forecasting. The conditional mean is

$$\mu_{t|t-1} = E[r_t \mid \mathcal{F}_{t-1}],$$  \hspace{1cm} (6)

and the conditional variance is

$$\sigma^2_{t|t-1} = \text{Var}[r_t \mid \mathcal{F}_{t-1}] = E[(r_t - \mu_{t|t-1})^2 \mid \mathcal{F}_{t-1}],$$  \hspace{1cm} (7)

where the information set $\mathcal{F}_{t-1}$ is assumed to reflect all relevant information up to time $t - 1$. Both conditional mean and variance may (usually do) differ from the unconditional measures, as they incorporate the most recent information available. The discretely sampled return process $r_t$ can now be decomposed into a conditional mean and an innovation-term

$$r_t = \mu_{t|t-1} + a_t.$$  \hspace{1cm} (8)

Furthermore, when $\mu_{t|t-1}$ is known or readily estimated\(^5\), equation (7) yields

$$\sigma^2_{t|t-1} = \text{Var}[r_t \mid \mathcal{F}_{t-1}] = \text{Var}[a_t \mid \mathcal{F}_{t-1}].$$  \hspace{1cm} (9)

\(^5\)In empirical studies $\mu_{t|t-1}$ will be estimated jointly with the conditional variance, for simplicity it can be assumed that $\mu_{t|t-1}$ follows some some linear time series model.
Table 1: Overview of different possible volatility definitions.

<table>
<thead>
<tr>
<th>Description</th>
<th>Equation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional standard deviation</td>
<td>[+\sqrt{\text{Var}(r_t</td>
<td>F_{t-1})}]</td>
</tr>
<tr>
<td>Time series volatility</td>
<td>(r_t = \mu_t + \sigma_t \epsilon_t)</td>
<td>The financial return (r_t) is made up of a drift (\mu_t) and the product of the volatility (\sigma_t) and the mean zero, variance one innovation (\epsilon_t). The innovations are for example iid Gaussian. Models for (\sigma_t) include ARCH/GARCH, Stochastic Volatility, Long Memory and Markov switching.</td>
</tr>
<tr>
<td>Spot volatility</td>
<td>(dp = \mu dt + \sigma dz)</td>
<td>The (log) price process is governed by a stochastic differential equation. (dz) denotes a standard Brownian motion, and (\sigma = \sigma(t)) is the spot volatility. This is the most common model assumed in option pricing.</td>
</tr>
<tr>
<td>Quadratic variation</td>
<td>(\lim_{t-(t+1) \to 0} \sum_i (p_i - p_{i-1})^2)</td>
<td>Consider a continuous time stochastic process over a given time period. If the (log) price process is a semimartingale, the quadratic variation is (under some mild assumptions) an unbiased estimator of conditional variance</td>
</tr>
<tr>
<td>Integrated volatility</td>
<td>(IV(t) = \int_{t-1}^t \sigma^2(\tau) d\tau)</td>
<td>If the (log) price process follows a stochastic differential equation, the quadratic variation equals the integrated volatility, which is of importance in options-pricing.</td>
</tr>
<tr>
<td>Realized volatility</td>
<td>(RV(t, \Delta) = \sum_{j=1}^{1/\Delta} \sum_{i=t-1}^{t+1} \Delta [\Delta])</td>
<td>Finite sample quantities of quadratic variation are often called 'realized volatility'. An example would be sampling with a frequency of 5 minutes. For finer and finer sampling, realized volatility and integrated volatility converge (RV(t, \Delta) \to IV(t))</td>
</tr>
<tr>
<td>Implied volatility</td>
<td>Given a specific option pricing formula, determine the volatility that matches the theoretical option prices from the model to the real life option prices in the market. This volatility is called the implied volatility. The Black Scholes model is often used for determining implied volatilities.</td>
<td></td>
</tr>
</tbody>
</table>
Volatility models are concerned with the evolution of $\sigma^2_{t|t-1}$. In the following section, various forecasting models based on the historical information set $\mathcal{F}_{t-1}$ will be presented (for details, see Poon and Granger, 2003). The partitioning and most of the content in this section is based on Poon and Granger (2003).

2.2.1 Past Standard Deviations - based Forecasts

The most simple model for forecasting future volatility assumes the volatility $\sigma_t$ to follow a random walk, yielding the forecast

$$\hat{\sigma}_t = \sigma_{t-1}. \quad (10)$$

An equally simple model is the historical average used as a forecast of future volatility

$$\hat{\sigma}_t = \frac{\sigma_{t-1} + \sigma_{t-2} + \ldots + \sigma_1}{t-1}, \quad t > 2. \quad (11)$$

Extending this idea, the moving average model that discards estimates that are older than $\varphi$

$$\hat{\sigma}_t = \frac{\sigma_{t-1} + \sigma_{t-2} + \ldots + \sigma_{t-\varphi}}{\varphi}, \quad \varphi \geq 2, \quad (12)$$

is obtained. As it is reasonable that more recent observations contain more information than older ones, the historical average can be built on to obtain the exponentially smoothened estimate

$$\hat{\sigma}_t = (1 - \alpha)\sigma_{t-1} + \alpha \hat{\sigma}_{t-1}, \quad \alpha \in [0, 1], \quad (13)$$

which places greater weight on more recent observations. Again, observations older than some finite $\varphi$ may be discarded, which yields the exponentially
weighted moving average model

\[ \hat{\sigma}_t = \frac{\sum_{i=1}^{p} \alpha_i \sigma_{t-i}}{\sum_{i=1}^{p} \alpha_i}. \]  

(14)

Linear time series models such as the AR\((p)\) model can be applied to volatility modelling, through

\[ \hat{\sigma}_t = \sum_{i=1}^{p} \alpha_i \sigma_{t-i} + \epsilon_t, \]  

(15)

where the volatility is expressed as a function of its past values and an independent and identically distributed \((iids)\) Gaussian error term. If past errors are included into the model, the ARMA\((p, q)\) model is obtained

\[ \hat{\sigma}_t = \sum_{i=1}^{p} \alpha_i \sigma_{t-i} + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} + \epsilon_t. \]  

(16)

The practical application of the models above (with the exception of the random walk and the historical average) involve finding the optimal lag length and/or weighting scheme for forecasting purposes. The optimization is usually done by minimizing a suitable error measure on in-sample forecasts in an estimation period. The optimal parameters are then used for forecasting. It is less common to constantly update the parameter estimates as new information becomes available. For simple moving average models, the search for an optimal time window involves a tradeoff between variance and bias of the estimator, as longer lags reduce the variance while increasing the bias. For instance, lags of five years for monthly data are commonly used in the empirical finance literature for simple moving average models (Andersen et al., 2005). For more information on estimation of linear time series models, see Box et al. (1994).

2.2.2 GARCH-Class Volatility Models

Engle’s (1982) ARCH model was the first to provide a systematic approach to volatility forecasting. The idea of ARCH models is that the mean cor-
rected asset return of (8), \( a_t \), can be represented as

\[
a_t = \sigma_t \epsilon_t,
\]

(17)

where the dependence of the \( a_t \)'s can be described by a quadratic function of its lagged values,

\[
\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_k a_{t-k}^2.
\]

(18)

Here, \( \epsilon_t \) is a sequence of iid, zero-mean and variance 1 random variables, and the parameters are non-negative, \( \alpha_0 > 0 \) and \( \alpha_i \geq 0 \) for \( i > 0 \). For practical purposes, \( \epsilon_t \) is often assumed to follow the standard normal distribution or standardized student-\( t \) distribution. In the model specification (18), large past shocks tend to produce large estimates for the conditional variance. This quality is in accordance with the volatility clustering observed in return series of financial assets.

An extension to the ARCH model was developed by Bollerslev (1986) known as generalized ARCH (GARCH). Let again \( a_t \) be the mean-corrected return of (8). In this case, \( a_t \) follows a GARCH\((p,q)\) process if (17) holds, with \( \epsilon_t \) a sequence of iid, zero-mean and variance 1 random variables, and

\[
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j|t-j+1}^2 \|
\]

(19)

where \( \alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0 \) and \( \sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_j) < 1 \), where \( \alpha_i = 0 \) for \( i > p \) and \( \beta_j = 0 \) for \( j > q \). As in ARCH models, \( \epsilon_t \) is often assumed to follow the standard normal distribution or a normalized student \( t \)-distribution. If \( q = 0 \) it can be seen that the GARCH model reduces to an ARCH model.

The most common empirical GARCH model, the GARCH\((1,1)\) model

\[
\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1|t-2}^2,
\]

(20)

\[\text{The latter constraint on } \alpha_i + \beta_j \text{ secures that the unconditional variance of } a_t \text{ is finite.}\]
helps illustrate some general important properties of GARCH-models. Again, it can be seen that large past innovations give rise to large $\sigma_{t}^{2}$, which leads to behavior similar to the volatility clustering observable in return series. Second, the GARCH(1,1) process can be shown to be leptocurtotic,

$$\frac{E[a_{t}^{4}]}{E[a_{t}^{2}]^2} = \frac{3[1-(\alpha_1 + \beta_1)^2]}{1-(\alpha + \beta_1)^2 - 2\alpha_1^2} - 3 > 0$$

and hence GARCH models tend to have heavier tails than the normal distribution, a phenomenon that also is observable in return series. The GARCH models have some weaknesses, however. First, even with standardized $t$-innovations, the tail behaviour of GARCH models generally remains too short (Tsay 2002), and GARCH models respond in the same manner to positive and negative past return shocks, which is not the case for empirical financial return series. A large number of different GARCH models have been developed in order to correct for different weaknesses\(^7\) in the original GARCH model. For an overview, see Andersen et al. (2005) and the references therein.

The asymmetrical response observable in return series to positive and negative past shocks is generally referred to as the “leverage” effect, since a price drop of the equity of a company increases its debt-to-equity ratio, making the company more likely to go bankrupt and its stock hence more volatile. It is unlikely that this explanation fully describes the observed phenomenon (see for instance Bouchaud et al., 2001), however several different GARCH formulations have been developed to describe this kind of asymmetry. The most common model of these is the EGARCH model of Nelson (1991). Encouraged of the results of a recent study of Hansen and Lunde (Hansen and Lunde, 2005) revealing that models incorporating the “leverage” effect clearly outperformed a traditional GARCH(1,1) model, an EGARCH model will be evaluated in the later chapters, and we will therefore here briefly present the EGARCH model.

In the EGARCH models, the conditional variance $\sigma_{t|t-1}^{2}$ is modeled by

\(^{7}\)Mostly, efforts have been made to incorporate the assymetrical response of real return series to positive and negative shocks, and long-term effects of past shocks into the model.
making \( \ln(\sigma^2_{t|t-1}) \) linear in some function of lagged \( \epsilon_t \)'s in order to ensure that \( \ln(\sigma^2_{n|t-1}) \) remains non-negative. The function that is chosen is

\[
g(\epsilon_t) = \theta \epsilon_t + \gamma \left( |\epsilon_t| - E[|\epsilon_t|] \right) = \begin{cases} 
(\theta + \gamma) \epsilon_t - \gamma E[|\epsilon_t|] & \text{if } \epsilon_t \geq 0 \\
(\theta - \gamma) \epsilon_t - \gamma E[|\epsilon_t|] & \text{if } \epsilon_t < 0
\end{cases},
\]

(22)

where the asymmetry is obvious. Generally, the EGARCH(\( p,q \)) model can be written as

\[
\ln(\sigma^2_{t|t-1}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^{q} \beta_j \ln(\sigma^2_{t-j|t-(j+1)})
\]

(23)

Again, to illustrate the EGARCH model, we study the EGARCH(1,1) model

\[
\ln(\sigma^2_{t|t-1}) = \alpha_0 + \alpha_1 g(\epsilon_{t-1}) + \beta_1 \ln(\sigma^2_{t-1|t-2}).
\]

(24)

If we assume \( \epsilon_t \) to be iid standard normal, we have \( E(|\epsilon_t|) = \sqrt{2/\pi} \) and we get the model

\[
\sigma^2_{t|t-1} = \begin{cases} 
\sigma^2_{t-1|t-2} e^{\alpha_0 + \alpha_1 ((\theta+\gamma) \epsilon_{t-1} - \gamma \sqrt{2/\pi})} & \text{if } \epsilon_t \geq 0 \\
\sigma^2_{t-1|t-2} e^{\alpha_0 + \alpha_1 ((\theta-\gamma) \epsilon_{t-1} - \gamma \sqrt{2/\pi})} & \text{if } \epsilon_t < 0
\end{cases}.
\]

(25)

To get an idea of the size of the leverage, we can study the ratio of the estimates for \( \sigma^2_{t|t-1} \) in case of a past shock \( \epsilon_{t-1} \) of equal magnitude, but with different signs. We get

\[
\frac{\sigma^2_{t|t-1}(+\epsilon_{t-1})}{\sigma^2_{t|t-1}(-\epsilon_{t-1})} = e^{\alpha_1 (\theta+\gamma) \epsilon_{t-1}} = e^{2\alpha_1 \theta \epsilon_{t-1}}.
\]

(26)

Because the model reflects the stronger volatility impact of a negative past shock seen in empirical return series, it would be expected that the above fraction should be less than one in any practical application. Therefore it can be expected that \( \alpha_1 \) and \( \theta \) have different signs. Note that by the construction of the EGARCH model, the parameter positivity constraints observed in GARCH models can be relaxed.
2.2.3 Stochastic Volatility

If a stochastic component is included into the equation of the conditional variance of $a_t$, the resulting model is called a stochastic volatility (SV) model. A SV model is somewhat more narrowly defined as

$$a_t = \sigma_t \epsilon_t \quad \ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln(\sigma_{t-i}^2) + \eta_t,$$

(27)

where the $\epsilon_t$ are iid standard normal, $\eta_t$ are iid $N(0, \sigma^2)$ and $\epsilon_t$ and $\eta_t$ are independent. In other words, $\ln(\sigma_t^2)$ is assumed to follow an AR($p$) process. The difference between the SV model (27) and the GARCH-class models in equations (18),(19) and (23) can be seen in that the latter do not include a stochastic component in the specification of the conditional variance of $a_t$, $\sigma_t^2$, whereas in the SV model (27) such a component $\eta_t$ is included. This changes the technicalities of the variance process. With a stochastic component, the exact value of the current volatility is non-measurable, even if the exact specification of the model and all past information is available. The estimation of SV models changes from being based on past information as is the case for GARCH models, to incorporating all available future information. Thus, the SV model is not only theoretically attractive; its dynamic specification makes it more able to incorporate fat tail distributions of returns, and have residuals that are closer to standard normal (Poon and Granger, 2003). The SV framework has been studied extensively in the field of options pricing.

2.2.4 Options-Implied Volatility

In Black-Scholes options pricing (Black and Scholes, 1973), the price of an asset is assumed to follow the dynamic

$$dS = \mu S dt + \sigma S dz$$

(28)

where $dz$ denotes a standard Brownian motion. By constructing a risk free portfolio consisting of only the derivative and the underlying security (which

\footnote{In Ghysels et al. (1995) the model (27) with $p = 1$ is referred to as the discrete time SV model. The generalisation from an AR(1) to an AR($p$) model is straightforward.}
must earn the risk-free rate of return), a pricing formula for the option can be constructed using Itô’s lemma. The formula is the following Black-Scholes-Merton differential equation,
\[
\frac{\partial f}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - rf = 0,
\]
(29)
where \( r \) is the risk free interest rate and \( f = f(S,t) \) denotes the price of a derivative contingent on \( S \). The assumptions in the Black-Sholes model are that the parameters \( \mu \) and \( \sigma \) are constant and short selling with full proceeds is permitted; there are no transaction costs, taxes or restrictions on security divisibility and no dividends are paid during the life of the derivative; arbitrage opportunities do not exist and trading is continuous and the risk-free interest rate \( r \) is constant.

In the Black-Sholes model, the price of an European call option is a function of \( S_t \), the current price of the underlying asset, \( K \), the strike price, \( T \), the time to expiration of the option, \( r \), the risk free interest rate and \( \sigma \), the volatility of the underlying asset. As \( S_t, K, T \) and \( r \) are observable, and the price is monotonically increasing in \( \sigma \), once the market produces a price for the option backwards induction can be used to figure out what \( \sigma \) the market price implies. This value for the volatility is called implied volatility. Because the time window under consideration is \([t, t + T]\), implied volatilities are often interpreted as the market expectation of \( \sigma \) over this period. In theory, each asset governed by the process (28) should have only one \( \sigma \). However, options with the same times to maturity but different strike prices in practice tend to produce different implied volatilities. The surface-plot of the implied volatility against its strike price and time to maturity is called an implied volatility surface (it usually has a non-flat profile), and the resulting plots in the \((K, \sigma)\) -plane are known as volatility smiles or volatility smirks. The most common theoretical explanation for the observed volatility smiles is the leptocurtotic behaviour of financial returns.

To illustrate why a fat tailed distribution of returns imply a volatility smile, consider first an option that is deep out of the money\(^9\). Since the

\(^9\)A call (put) option is said to be out of the money when \( K > S_t \) (\( K < S_t \)), at the
actual return distribution is leptocurtotic, the probability of the option to move into the money is greater than the theoretical price function assumes. The price of the option will thus be greater than the theoretical price, thus implying a greater implied volatility. Similarly, if an option is deep in the money, the probability of it moving even further into the money is greater than the theoretical distribution assumes. Again, thus, its price will be higher than the theoretical price, implying a higher implied volatility. This reasoning for a volatility smile works where empirical returns have symmetric thick tails, such as the currency market. In the stock market, a volatility skew is however mostly observed, because empirical returns are skewed to the left, producing higher implied volatilities for low strike prices (Hull, 2003).

3 Estimation of Volatility Models and Options-Implied Volatility

Next, a look at estimating GARCH and SV models, and how to produce options-implied volatility forecasts will be taken. For GARCH models maximum likelihood estimates for model parameters can be produced. For SV models ML parameter estimates cannot be produced directly. The incorporation of two innovation terms into the SV model, $\epsilon_t$ and $\eta_t$, complicates the estimation of SV models. The method discussed here will be the Monte Carlo Markov Chain (MCMC) simulation technique, which will also later be used for practical model estimation. The methods of calculating implied volatilities from a theoretical Black-Scholes options-pricing formula will be discussed, as well as issues that need to be taken into account when using these.

3.1 GARCH Model Estimation

GARCH model estimation is fairly straightforward. It can be viewed as a three stage process. First, any linear dependence in the return data has to be removed. This corresponds to building an econometric model for the return conditional mean $\mu_{t|t-1}$ such as an AR or ARMA model. Once this has been

money when $K = S_t$, and in the money when $K < S_t$ ($K > S_t$)
done, the residuals of this model have to be tested for heteroscedasticity (also called ARCH effect). Second, if heteroscedasticity is detected, the order of the GARCH model to be fitted has to be specified and the model parameters then estimated. Finally, the model has to be checked and refined if necessary.

The first step described removes the process mean from the sample. Sometimes an AR or ARMA model has to be fitted in order to remove all linear dependence from the data. The squared residuals of the (possibly) fitted ARMA model, \( a_t^2 = (r_t - \mu_{t|t-1})^2 \) are used to test for conditional heteroscedasticity. The Ljung-Box-Pierce Q-test for the squared residual series, and Engle’s ARCH test (Engle, 1982, pp. 999-1000) can be used, see appendix A for a brief description of the tests. Volatility clustering observable in most financial returns imply strong autocorrelation in squared returns (Tsay, 2001).

For the identification of the correct order of a GARCH model the same methods can be used as in ARMA model order selection, since GARCH models essentially ARMA models for the evolution of the conditional variance of \( a_t \). The procedure usually involves studying the autocorrelation function (ACF) and partial autocorrelation function (PACF) of \( a_t^2 \) to determine the dependencies. However, it is common to proceed by trial and error, testing and comparing simple common models with short (\( \leq 3 \)) lags.

Maximum likelihood estimation (MLE) of the model parameters is used. Under the assumption that the innovations \( \epsilon_t \) of (17) are iid standard the conditional density function of \( a_t \) is

\[
f(a_t|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi \sigma_t^2}} \exp \left[ -\frac{a_t^2}{2\sigma_t^2} \right].
\]

Letting \( \theta \) denote the unknown vector of parameters to specify the conditional variance equation to be estimated, the log-likelihood function of the observed sample \( a_N, a_{N-1}, \ldots, a_1 \) is

\[
\ell(a_N, a_{N-1}, \ldots, a_1; \theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{t=\max(p,q)+1}^{N} \left[ \log \sigma_{t|t-1}^2(\theta) - \frac{a_t^2}{\sigma_{t|t-1}^2(\theta)} \right],
\]

(30)
where $\sigma^2_{t|t-1}(\theta)$ is the estimate according to the model to be fitted\textsuperscript{10}. The argument that maximizes the likelihood of the observed sample, 
$$\hat{\theta} = \arg \max \ell(a_N, a_{N-1}, \ldots, a_1; \theta)$$

can be found with a suitable numerical optimization technique.

The assumption that $\epsilon_t$ is iid standard normal coupled with time varying volatility implies leptocurtotic returns $r_t$. The result is however normally not fat-tailed enough to reflect the properties observable in daily return series and the above log-likelihood function hence becomes misspecified. The estimates can be improved by approximating the distribution of $\epsilon_t$ by a standardized student-$t$ distribution with degrees of freedom $\nu > 2$. In this case, the conditional density function of $a_t$ becomes

$$f(a_t|F_{t-1}) = \frac{\Gamma(\frac{1}{2}\nu + 1)}{\Gamma(\frac{\nu}{2})\sqrt{(\nu - 2)\pi}} \left(1 + \frac{a^2_t}{(\nu - 2)\sigma^2_{t|t-1}}\right)^{-\frac{(\nu+1)}{2}},$$

where $\Gamma(\cdot)$ denotes the Gamma function $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$. The corresponding log likelihood function of the observed sample $a_N, a_{N-1}, \ldots, a_1$ and parameter vector $\theta_t = (\theta, \nu)$ where the degrees of freedom of the $t$-distribution are estimated jointly becomes

$$\ell(a_N, a_{N-1}, \ldots, a_1; \theta_t) = -\sum_{t=\max(p,q)}^{N} \left[ \frac{\nu + 1}{2} \log \left(1 + \frac{a^2_t}{(\nu - 2)\sigma^2_{t|t-1}}\right) + \log \left(\frac{\sigma_{t|t-1}}{\sqrt{(\nu - 2)\pi}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\right) \right]. \quad (31)$$

Model checking can be done in several ways, the most common being examining the series $\tilde{a}_t = a_t/\sigma_t$ which under the assumptions of the model specification should follow either a student-$t$ distribution or a standard normal distribution. Standard techniques such as quantile to quantile plots and skewness and kurtosis checking can be used. Tests for heteroscedasticity may also be employed at this stage in order to quantify and confirm the correct model specification.

\textsuperscript{10}For example, if a GARCH(1,1) model is fitted, $\theta = (\alpha_0, \alpha_1, \beta_1)$ and correspondingly $\sigma^2_{t|t-1}(\theta) = \alpha_0 + \alpha_1 a^2_{t-1} + \beta_1 \sigma^2_{t-2|t-2}$. 

16


3.2 Stochastic Volatility Model Estimation

The latent volatility process in the SV model has consequences for the model estimation, because estimation cannot be done through direct application of maximum likelihood. We will only discuss the Monte Carlo Markov Chain method for SV model estimation. Many alternative estimation methods for SV models have been proposed in the literature, for a review see Broto and Ruiz (2004). The remainder of this section is based on Verhofen (2005) and Tsay (2001).

3.2.1 Markov Chain Monte Carlo

A discrete-time stochastic process $X_t$ where $X_t \in \Omega \forall t$ is a Markov process if

$$P(X_h|X_t, X_{t-1}, \ldots) = P(X_h|X_t), \quad h > t.$$  

For $\Phi \subset \Omega$, the transition probability function of the Markov process is the function

$$P_t(\omega, h, \Phi) = P(X_h \in \Phi|X_t = \omega), \quad h > t.$$  

If the above probability depends only on $h - t$, the process has a stationary transition distribution. Markov Chain simulation is a method of sampling from a probability distribution by constructing a Markov Chain that has the distribution in question as its stationary transition distribution. In a normal inference problem setting a parameter vector $\phi \in \Omega$ needs to be estimated conditional to the vector of observations $X$. In order to estimate the parameter vector $\phi$ the density function $f(\phi|X)$ needs to be known. The Markov Chain simulation simulates a Markov process $Y_t \in \Omega$ that converges to a stationary distribution equal to $f(\phi|X)$. For a specified $f(\phi|X)$, many Markov Chains can be constructed that behave in the desired fashion. Consider a SV model, where the mean equation of the asset return is a normal ARMA(p,q) model, and the volatility model is a SV model of order 1.

$$r_t = \beta_0 + \sum_{i=1}^{p} \beta_i r_{t-i} + \sum_{i=1}^{q} \phi_i a_{t-i} + a_t, \quad a_t = \sigma_t \epsilon_t \quad (32)$$

$$\ln \sigma^2_{t|t-1} = \alpha_0 + \alpha_1 \ln \sigma^2_{t-1|t-2} + \nu_t \quad (33)$$

17
where $\epsilon_t$ is standard normal iid, $\nu_t \sim N(0, \varsigma^2)$ is iid and $\epsilon_t \perp \nu_t$. The notation $\beta = (\beta_0, \beta_1, \ldots, \beta_p, \phi_1, \phi_2, \ldots, \phi_q)^T$ for the parameter vector of the mean equation will be used, and similarly $\alpha = (\alpha_0, \alpha_1, \varsigma^2)^T$. The observed return series will be denoted $R = (r_1, r_2, \ldots, r_N)^T$ and the vector of unobservable volatilities $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)^T$. The estimation of the model would be complicated in a maximum likelihood setting, because the likelihood function $f(R|\beta, \alpha) = \int f(R|\beta, \sigma) f(\sigma|\alpha) d\sigma$ would require integration over the $N$ dimensional vector $\sigma$. A Gibbs sampling method can be used instead.

### 3.2.2 Gibbs Sampling

Gibbs sampling involves generating draws from the joint probability distribution of several random variables. As an example, let there be three parameters $\theta_1, \theta_2$ and $\theta_3$, a set of observed data $X$ and a model $M$ to be fitted. The three unknown parameters need to be estimated in order to fit the model, but suppose the likelihood function is hard to obtain. However, suppose that we can draw random draws from the conditional distributions

$$f_1(\theta_1|\theta_2, \theta_3, X, M), \quad f_2(\theta_2|\theta_1, \theta_3, X, M), \quad f_3(\theta_3|\theta_1, \theta_2, X, M).$$

One Gibbs sampling iteration proceeds as follows; Let $\theta_{2,0}, \theta_{3,0}$ be arbitrary starting values for $\theta_2, \theta_3$.

1. Draw $\theta_{1,1}$ from $f_1(\theta_1|\theta_{2,0}, \theta_{3,0}, X, M)$
2. Draw $\theta_{2,1}$ from $f_2(\theta_2|\theta_{1,1}, \theta_{3,0}, X, M)$
3. Draw $\theta_{3,1}$ from $f_3(\theta_3|\theta_{1,1}, \theta_{2,1}, X, M)$

Repeating the iteration with the obtained iterates as starting values, a sample

$$(\theta_{1,1}, \theta_{2,1}, \theta_{3,1}), \ldots, (\theta_{1,N}, \theta_{2,N}, \theta_{3,N})$$

can be obtained. For sufficiently large $N$ and under some weak regularity conditions, it can be shown that $(\theta_{1,N}, \theta_{2,N}, \theta_{3,N})$ is approximatively a random
draw from the joint distribution \( f(\theta_1, \theta_2, \theta_3|X, M) \).

In practice, after a sufficiently large number of iterations has been run, a random sample

\[
(\theta_{1,N}, \theta_{2,N}, \theta_{3,N}), \ldots, (\theta_{1,N+n}, \theta_{2,N+n}, \theta_{3,N+n})
\]

from the joint distribution \( f(\theta_1, \theta_2, \theta_3|X, M) \) is obtained, which can be used to make inference. For instance, the normal point estimates of sample means and variances can be calculated as

\[
\hat{\theta}_i = \frac{1}{n+1} \sum_{j=N}^{N+n} \theta_{i,j}, \quad \hat{\sigma}_i^2 = \frac{1}{n} \sum_{j=N}^{N+n} (\theta_{i,j} - \hat{\theta}_i)^2.
\]

For some models the conditional density \( f(\theta_i|\theta_{-i}, X, M) \) is hard to derive or to sample from. In those cases the Griddy Gibbs sampler (Ritter and Tanner (1992)) can offer relief. The Griddy Gibbs sampler constructs an approximation to the conditional density numerically, by evaluating the (joint) posterior density on a grid of points. Formally, the Griddy Gibbs process goes as follows:

1. Evaluate \( f(\theta_i|\theta_{-i}, X, M) \) at \( \theta_{i,1}, \ldots, \theta_{i,n} \) to obtain \( \omega_1, \ldots, \omega_n \), where \( \omega_j = f(\theta_{i,j}|\theta_{-i}, X, M) \)

2. Use \( \omega_1, \ldots, \omega_n \) to obtain an approximation of the inverse cumulative density function of \( f(\theta_i|\theta_{-i}, X, M) \)

3. Draw a sample from the \( \text{Uniform}(0, 1) \) distribution and transform it to a random draw for \( \theta_i \) via the inverse CDF.

Ritter and Tanner (1992) provide some insightful remarks on the Griddy Gibbs sampling as well;

1. \( f(\theta_i|\theta_{-i}, X, M) \) needs to be known only up to a proportionality constant, since the normalisation can be obtained from \( \omega_1, \ldots, \omega_n \).

2. The grid \( \theta_{i,1}, \ldots, \theta_{i,n} \) is better if it is not uniformly distributed, but rather has more points in areas with high probability mass and less points in areas with low mass.

\(^{11}\) for an \( N \)-vector \( \theta, \theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_N) \)
3. The number of points in the grid needs not to be constant over all iterations and can be made finer when the algorithm has appeared to stabilize.

4. A simple approximation to the inverse CDF is piecewise constant \( F^{-1}(p) = \theta_{k,l} \text{ s.t. } \sum_{\theta_{i,j} < \theta_{k,l}} P(\theta_{i,j}) = p \) corresponding to the discrete distribution for \( \theta_{1,1}, \ldots, \theta_{1,n} \) where \( P(\theta_{i,j}) = \frac{\omega_j}{\sum_k \omega_k} \).

5. The interval \([\theta_{l,1}, \theta_{l,n}]\) must be checked carefully. In practice this can be done by observing the histogram of draws, and if concentration is observed close to the end points or only on a small intervall the original intervall may be expanded or shortened, respectively.

As Bayes’ law states that

\[
    f(\theta_i|\theta_{-i}, X, M) = \frac{f(\theta_i, X, M)}{\int_{\theta_{-i}} f(\theta_i, X, M)}
\]

the conditional density is proportional to the joint density, which thus can be used in Griddy Gibbs.

### 3.2.3 Gibbs Sampling and SV Model Estimation

Consider the mean and volatility equations (32) and (33). In a Gibbs Sampling setting, we would draw random samples from the conditional distributions

\[
    f(\beta|\alpha, \sigma), \quad f(\alpha|R, \beta, \sigma), \quad f(\sigma|R, \beta, \alpha).
\]

For a fixed \( \sigma \), the mean equation (32) can be divided by \( \sigma_{i|t-1} \) to obtain

\[
    \tilde{r}_t = \beta^T y_t + \epsilon_t,
\]

where \( y_t \) is the vector of lagged returns \( r_{t-l}, l = 1, \ldots, p \) and past innovations \( a_{t-l}, l = 1, \ldots, q \) used as explanatory variables in the mean equation (32), divided by \( \sigma_{i|t-1} \), \( y_t = (1, r_{t-1}, \ldots, r_{t-p}, a_{t-1}, \ldots, a_{t-q})^T/\sigma_{i|t-1} \). Taking a Bayesian approach and supposing that the prior distribution of \( \beta \) is multivariate normal with mean \( \mu_\beta \) and covariance matrix \( \Sigma_\beta \), then by the conjugate prior
properties of the multinormal distribution\(^{12}\) the posterior distribution of \(\mathbf{\beta}\) is also multinormal with covariance \(\tilde{\Sigma}_{\beta}\) and mean \(\tilde{\mathbf{\mu}}_{\beta}\). Denote the mean and covariance of \(\mathbf{y}_t\) by \(\mathbf{\mu}_y\) and \(\Sigma_y\). Then
\[
\tilde{\Sigma}_{\beta} = (\Sigma_{\beta}^{-1} + (p + q + 1)\Sigma_y^{-1})^{-1}, \quad \tilde{\mathbf{\mu}}_{\beta} = \tilde{\Sigma}_{\beta} (\Sigma_{\beta}^{-1} \mathbf{\mu}_{\beta} + (p + q + 1)\Sigma_y^{-1} \bar{y}),
\]
where \(\bar{y}\) is the sample mean 
\[
\frac{1}{N_{\max(p,q)}} \sum_{t=1}^{N_{\max(p,q)}} y_t.
\]

The volatility vector \(\mathbf{\sigma}\) is produced element by element, using the conditional posterior distributions \(f(h_t|\mathbf{R}, \mathbf{\sigma}_{-t}, \mathbf{\beta}, \mathbf{\alpha})\), which are produced by the normal distribution of \(a_t\) and the lognormal distribution of \(\sigma_{t|t-1}\). It can be shown that
\[
f(\sigma_{t|t-1}|\mathbf{R}, \mathbf{\sigma}_{-t}, \mathbf{\beta}, \mathbf{\alpha}) \propto \sigma_{t|t-1}^{-1.5} \exp \left[ -\frac{r_t - \mathbf{\beta}^T y_t}{2\sigma_{t|t-1}} - \frac{\ln \sigma_{t|t-1} - \zeta_t}{2\eta^2} \right]
\]
where \(\zeta_t = [\alpha_0(1-\alpha_1) + \alpha_1(\ln \sigma_{t-1} + \ln \sigma_{t-2})]/(1+\alpha_1^2)\) and \(\eta^2 = \zeta_t^2/(1+\alpha_1^2)\). Griddy Gibbs can be used for the sampling, and starting values for \(\sigma_{t|t-1}\) can be obtained by constructing a GARCH model and using the fitted values.

The sampling of \(\mathbf{\alpha}\) is done by partitioning the parameters as \(\mathbf{a} = (\alpha_0, \alpha_1)^T\) and \(\zeta^2\). First, the conditional posterior distribution \(f(\mathbf{a}|\mathbf{R}, \mathbf{\beta}, \mathbf{\sigma}, \zeta^2) = f(\mathbf{a}|\mathbf{\sigma}, \zeta^2)\) is needed. If the prior distribution of \(\mathbf{a}\) is multivariate normal with mean \(\mathbf{\mu}_a\) and covariance matrix \(\Sigma_a\), then the posterior is also multivariate normal with mean \(\tilde{\mathbf{\mu}}_a\) and covariance matrix \(\tilde{\Sigma}_a\) where
\[
\tilde{\Sigma}_a^{-1} = \frac{1}{\zeta^2} \sum_{t=2}^{N} \mathbf{z}_t \mathbf{z}_t^T + \Sigma_a^{-1}, \quad \tilde{\mathbf{\mu}}_a = \tilde{\Sigma}_a \left( \frac{1}{\zeta^2} \sum_{t=2}^{N} \mathbf{z}_t \ln \sigma_{t|t-1} + \Sigma_a^{-1} \mathbf{\mu}_a \right),
\]
where \(\mathbf{z}_t = (1, \ln \sigma_{t|t-1})^T\). Next, \(f(\zeta^2|\mathbf{R}, \mathbf{\beta}, \mathbf{\sigma}, \mathbf{a}) = f(\zeta^2|\mathbf{\sigma}, \mathbf{a})\) is to be considered. Given \(\mathbf{\sigma}\) and \(\mathbf{a}\), \(\nu_t\) can be calculated from (33). If the prior distribution of \(\zeta^2\) is \(\chi^2_m \sim \chi^2_m\), then
\[
\frac{m \lambda + \sum_{t=2}^{N} \nu_t^2}{\zeta^2_t} \sim \chi^2_{m+N-1}.
\]

\(^{12}\) In general, obtaining prior distributions is not easy, however there are cases when posterior and prior distributions belong to the same family of distributions. In this case the prior distribution is called conjugate prior.
3.3 Creating Options Implied Volatility Forecasts

Under the assumptions of Section 2.2.4, the price of a European call option on a non dividend-paying stock and an European put option on an non dividend-paying stock are

\[
C(S,T) = S\Phi(d_1) - Ke^{-rT}\Phi(d_2), \quad \text{and} \quad (34) \\
P(S,T) = Ke^{-rT}\Phi(-d_2) - S\Phi(-d_1), \quad (35)
\]

where \( \Phi \) denotes the standard normal cumulative density function \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du \) and

\[
d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \\
d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}.
\]

In principle, the calculation of the implied volatility \( \sigma \), or the option implied standard deviation (ISD) from the above equalities given an observed market price \( C \) or \( P \) is straightforward and can be done by simple numerical methods such as the Newton-Raphson method\(^{13}\). Yet, the computation of ISD’s for volatility forecasting involves several issues.

First, the underlying assumption of no dividend payments in equations (34) and (35) must be corrected for in some way. One approach is to treat the current stock price as a sum of two components: a riskless component consisting of dividends payable during the life of the option, and a risky component. This division is convenient, because the Black-Scholes pricing formula is correct if values for the risky component of the stock price \( S_r \) and the volatility of this component \( \sigma_r \) are used. Because of the different tax treatment on dividends and capital gains, it is widely acknowledged that stock prices do not drop by the full amount of the dividend on the ex-dividend date. The magnitude of the riskless component therefore has to be

\(^{13}\)In practice, a more difficult task is to evaluate \( \Phi(x) \). However, routines for calculating \( \Phi(x) \) exist in almost every software package. Ready made routines for calculating ISD also exist in many software packages and as freeware on the internet.
treated as a multiple of the present value of all the dividends during the life of the option discounted from the ex-dividend date to the present, or

\[ S_r(t) = S(t) - \delta \sum_{t_i > t} d_{t_i} e^{-r(t_i - t)}, \]  

where \( S(t) \) denotes the observed price at time \( t \), \( d_{t_i} \) denotes the dividend to be paid at time \( t_i \), \( \delta \leq 1 \) denotes the proportion of the dividend payments by which the share price drops on the ex-dividend date and \( r \) is the risk free interest rate. In practice, the dividend adjustment fraction \( \delta \) could be estimated from historical data for each individual stock, or then an average rate could be used. Kalay (1982) finds the dividend adjustment fraction to be 88% for a large sample of stocks between 1966 and 1967, and in a study by Beckers (1981), a dividend adjustment fraction of 85% is used referring to previous studies.

The value of \( \sigma \) implied by the pricing formulas (34) or (35) is the annual standard deviation of the stock. From the underlying assumption on the stock price dynamic (28), it follows that the logarithmic return, \( r_t \), of an asset is normally distributed with mean \( (\mu - \frac{\sigma^2}{2}) \Delta t \) and variance \( \sigma^2 \Delta t \) where \( \Delta t \) denotes the fraction of one year that the return interval constitutes. Since there are about 252 trading days a year, a measure of the daily standard deviation could be inferred from the obtained value of \( \sigma \) as \( \sigma \sqrt{\frac{1}{252}} \). This can however be a poor estimate of any specific daily standard deviation, since the obtained \( \sigma \) reflects the annualized equivalent of the volatility over the remaining lifespan of the option, where any daily differences have been averaged out.

Another issue to be considered is how to treat the observable volatility smiles discussed in Section 2.2.4. The most common strategy is to choose the ISD from an ATM option, or the NTM\(^{14}\) option if an ATM option is unavailable. The argument used is that ATM options tend to be most traded and hence least susceptible to measurement errors. Furthermore ATM ISD is theoretically most sound in a stochastic volatility setting (Poon and Granger (2003)). Another possibility is to create a weighted average of several ISD’s

---

\(^{14}\)Nearest-to-the-money
Table 2: Overview of Loss Functions Applicable to Forecast Estimation.

<table>
<thead>
<tr>
<th>Loss Function</th>
<th>MSE1</th>
<th>MSE2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Square Error</td>
<td>( \frac{1}{N} \sum_{t=1}^{N} (\sigma_t - \hat{\sigma}_t)^2 )</td>
<td>( \frac{1}{N} \sum_{t=1}^{N} (\sigma_t^2 - \hat{\sigma}_t^2)^2 )</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>( \frac{1}{N} \sum_{t=1}^{N}</td>
<td>\sigma_t - \hat{\sigma}_t</td>
</tr>
<tr>
<td>Mean Absolute Percent Error</td>
<td>( \frac{1}{N} \sum_{t=1}^{N} \left</td>
<td>\frac{\sigma_t - \hat{\sigma}_t}{\sigma_t} \right</td>
</tr>
</tbody>
</table>

obtained from options with different strikes\(^\text{15}\). One such weighting scheme used by Beckers (1981) would be to use the ISD that minimizes the square sum of errors between market prices and theoretical prices for different options

\[
MISD = \text{arg min} \sum_i w_i \left[ P_i - BS_i(ISD) \right]^2, \quad (37)
\]

where \( P_i \) denotes the observed market price of the \( i \):th option, \( BS_i \) denotes the theoretical Black-Scholes price for the \( i \):th option as a function of the used ISD, and \( w_i \) denote weights\(^\text{16}\).

4 Volatility Forecast Evaluation

Ideally, forecasting evaluation should reflect the intended use of the forecasts obtained. In a more general setting, where the use of the forecasts is not specified, general measures have to be used. There are a multitude of such general measures, the most popular ones being the mean square error, the mean absolute error and the mean absolute percent error. Mean absolute error criteria are interesting because they are more robust to outliers than normal mean square error measures and mean absolute percentage error measures can be considered more justified than other measures when forecasts differ widely in magnitude, since errors are measured relative to the outcome.

This is valuable in forecasting volatility which might change radically from day to day. Alternative loss functions is given in Table 2.

\(^{15}\)The options should have the same maturity however, since the ISD is a measure of the markets view of the volatility of the stock over the remaining life of the option and is hence time and maturity dependent.

\(^{16}\)Usually, \( w_i = \partial BS_i(ISD)/\partial ISD, \) or \( S\varphi(d_1)\sqrt{T} \) where \( \varphi \) denotes the standard normal probability density function.
Another popular forecast evaluation method is the regression based method where the ex-post realisations are regressed on the ex-ante forecasts

\[ Y_i = \alpha + \beta \hat{Y}_i + \eta_i, \]

where the forecast can be considered unbiased if \( \alpha = 0 \) and \( \beta = 1 \). This regression based evaluation is often critisized since it may yield higher \( R^2 \) for a biased estimate than for an unbiased estimate. It is however to be remem-bered that a biased forecast can be of great forecasting value if the bias can be corrected. Hence, \( \tilde{Y}_i = Y_i - \eta_i \) can be considered a good forecast if the \( R^2 \) of the above regression tends to unity.

Volatility forecasts have special characteristics which – unless they are accounted for – may lead to erroneous conclusions about forecast quality. Let the return innovation \( a_t \) be written as in (17), where the latent volatility \( \sigma_t \) evolves through time according to an identified model. A straightforward evaluation of the forecasting performance of the model is to compare the model forecast to subsequent realizations. For volatility, which is not ob-servable, this cannot be done directly. Fortunately, if the model is correctly specified, then the squared return innovations \( a_t^2 \) provide an unbiased esti-mate of the squared volatility \( \sigma_t^2 \); if

\[ r_t = \mu_{t|t-1} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \text{where } \epsilon_t \sim N(0, 1) \text{ i.i.d.,} \]

then

\[ E [a_t^2 \mid F_{t-1}] = \sigma_t^2 E[\epsilon_t^2 \mid F_{t-1}] = \sigma_t^2. \]

However, the squared innovations provide a very noisy measurement due to the idiosyncratic error term \( \epsilon_t^2 \). If this component is large compared to \( \sigma_t \) with considerable variation from observation to observation, the fraction of the squared return attributable to the volatility process is low. As a conse-quence, when judged by standard forecasting criteria using \( a_t^2 \) as a measure of ex-post volatility, poor predictive power is a direct consequence of the

Consider the enlightening illustration by Poon and Granger (2003): Since \( \epsilon_t^2 \sim \chi_1^2 \), \( \Pr[a_t^2 \in \left[ \frac{1}{2} \sigma_t^2, \frac{3}{2} \sigma_t^2 \right]] = \Pr[\epsilon_t^2 \in \left[ \frac{1}{2}, \frac{3}{2} \right]] = 0.2588 \), i.e. \( a_t \) is 50% greater or smaller than \( \sigma_t \) almost 75% of the time!
noise in the return generating process. Therefore, interpreting the resulting $R^2$ from the regression

$$a_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t$$

(38)
is not very informative unless it is compared to a benchmark for the value expected under the null hypothesis of correct model specification. Andersen and Bollerslev (1998) show that under the null hypothesis of a GARCH(1,1) process generating the volatility process, the $R^2$ of the regression (38) cannot exceed the reciprocal of the kurtosis of the standardized innovations, $[E[\epsilon_t^4]]^{-1}$. In the specific setting of a correctly specified GARCH(1,1) model with Gaussian errors, the $R^2$ is bounded from above by $\frac{1}{3}$, the bound being even lower with fat-tailed errors. The poor quality of squared daily returns as a volatility proxy has lead to the investigation of using other measures for this purpose. One better proxy is the use of high-frequency intraday squared returns, with which $R^2$’s close to the theoretical ideal values were obtained by Andersen and Bollerslev (1998), leading to the conclusion that “Yes, ARCH and stochastic volatility models do provide good volatility forecasts!” Thus, if high-frequency data is unavailable, $R^2$ measures obtained from regression based forecast evaluation with squared returns as volatility proxies have to be considered at most as comparable between each other, not as absolute measures of the forecast quality in a normal sense of the measure.

5 Model Estimation for Four Finnish Equities

In this section, different GARCH models and options-implied volatility forecasts will be constructed for four Finnish equities; Neste Oil, Nokia, Sampo and UPM Kymmene. These will henceforth be referred to as NES, NOK, SAM and UPM, respectively. The choice of these equities has been based on the need for liquid options with the stocks as underlying, since liquid options should produce better volatility forecasts. This is needed if the forecasting performance of option-ISD is to be compared to forecasts by other volatil-
ity models. Based on Helsinki Stock Exchange warrant\textsuperscript{18} trading statistics\textsuperscript{19} from November 2006 to February 2007, the most liquid equity warrants were written on these four stocks.

GARCH forecasts are based only on historical price information. Since stock returns are affected by non-historical information, any forecast based only on historical information cannot fully predict return characteristics. Option ISD's in efficient markets reflect all past information as well as rational expectations of the future. The option-implied forecasts should therefore, since they are based on a larger information set, be more reliable than GARCH forecasts.

5.1 The Data

Daily closing prices for the equities under consideration have been used for both model estimation and evaluation. The data is downloadable free of cost at OMX’s webpages\textsuperscript{20} at the time of writing. For GARCH models, the model estimation sample consists of closing prices from 1 January 2002 to 1 December 2006 for NOK, SAM and UPM. For NES, which was listed only in April 2005, the estimation period spans from 1 January 2006 to 1 December 2006. An evaluation period of 70 days, 2 December 2006 to 14 May 2007 has been used for all equities.

The price series of UPM can be seen in Figure 1, whereas the price series for the other equities can be found in Appendix B. The return series corresponding to the price series are shown in Appendix C.

The return series (NES, NOK, SAM and UPM) show no autocorrelation measured by the Ljung-Box-Pierce Q-test or Engles ARCH-test at lags of 10, 20 or 30 days. The mean component of the return series of these three

\textsuperscript{18}A warrant is an option issued by a company or a financial institution. In most literature however, warrants are referred to as call options issued by a company on forthcoming new share issues. This concept matches that of the “Options” traded on the Helsinki Exchange, which are call options issued by institutions on their own stock. Warrants on the Helsinki Exchange are simply options issued by financial institutions.

\textsuperscript{19}Available from the OMX website http://www.omxgroup.com/nordicexchange/newsandstatistics/statisticsanalysis/Equities/marketshareswarrants/ at the time of writing.

\textsuperscript{20}http://www.omxgroup.com/nordicexchange/newsandstatistics/statisticsanalysis/Equities/dailyhelsinki/dailyclosehistory/
Table 3: Constant Means of Return Series.

| Equity | \( \mu_t | \mu_{t-1} = \mu \) | t Value |
|--------|----------------|--------|
| NES    | -0.00028979 | -0.1921 |
| NOK    | 1.6891e-005 | 0.0289 |
| SAM    | 0.0011958  | 2.6066 |
| UPM    | 0.00028695 | 0.6681 |

equities can thus be modelled by a simple constant, \( \mu_t | \mu_{t-1} = \mu \). The constant means will be used throughout our analysis for the return series. An overview of the mean components is given in Table 3.

The de-meaned squared return series (i.e. squared innovations \( a_t^2 \)) of all four equities show autocorrelation, which justifies the fitting of GARCH or SV models to the data of these equities. This autocorrelation is however weak for all of the equities, raising doubt on the possible predictive performance of any standard autoregressive model. As an example, the ACF and PACF of UPM squared returns is showed in Figure 2. As opposed to our results, autocorrelation is found to be strong for squared returns of financial assets.
in most similar studies.

Figure 2: ACF and PACF of Squared Returns of UPM Kymmene. The autocorrelation is barely significant at just above 0.05. This raises doubt about the possible predictive performance of autoregressive models.

The data for creating option implied volatility forecasts is not readily available. Because the general practice of most data providers is only to provide price-history data for active warrants, a very limited data sample has been available to us\textsuperscript{21}. The same OMX data service used for the equity data also provides price data for a limited amount of warrants maturing in end 2007 and beginning 2008. Unfortunately, the same practice of removing price data for non-active warrants is followed by the OMX data service as well. We have therefore gained access only to one highly liquid warrant per stock, the NTM warrant from fall 2007. The one warrant for each stock that has been selected is the warrant with the corresponding stock as an underlyer that was most actively traded in September\textsuperscript{22}. For all equities but NOK, a warrant that at the time of writing was still active was chosen. Table 4

\textsuperscript{21}The services of Thomson Data Stream and Reuters do not provide price data of non-active warrants on the Helsinki exchange.

\textsuperscript{22}Based on the statistics mentioned in Footnote 19. Because of the unfortunate practise of removing non-active warrants, the warrant used for NOK is not the one that was most actively traded in September measured by $\#$ of trades since we were unable to get data for that warrant. It is however the NTM warrant.
Table 4: Overview of Warrants Used For Implied Volatility Forecast Creation. Turnover and trade number statistics are from September 2007. Strike and Turnover are reported in EUR. The # of Trades are reported in single units. Factor denotes how many warrants are needed to buy one share.

<table>
<thead>
<tr>
<th>Name</th>
<th>Strike</th>
<th>Start</th>
<th>Maturity</th>
<th>Factor</th>
<th>Issuer</th>
<th>Turnover</th>
<th># Trades</th>
</tr>
</thead>
<tbody>
<tr>
<td>7KNESEW250</td>
<td>25.0</td>
<td>20.08.07</td>
<td>16.11.07</td>
<td>5:1</td>
<td>SHB</td>
<td>1 686 932</td>
<td>599</td>
</tr>
<tr>
<td>7JNOKEW240</td>
<td>24.0</td>
<td>09.07.07</td>
<td>19.10.07</td>
<td>5:1</td>
<td>Nordea</td>
<td>5 285 939</td>
<td>102</td>
</tr>
<tr>
<td>7KSAMEW215</td>
<td>21.5</td>
<td>20.08.07</td>
<td>16.11.07</td>
<td>5:1</td>
<td>SHB</td>
<td>557 440</td>
<td>274</td>
</tr>
<tr>
<td>7LUPMEW160</td>
<td>16.0</td>
<td>20.08.07</td>
<td>21.12.07</td>
<td>5:1</td>
<td>SHB</td>
<td>220 472</td>
<td>100</td>
</tr>
</tbody>
</table>

gives an overview of the warrants used for creating volatility forecasts.

5.2 GARCH Volatility Forecasts

The estimation of GARCH models was done with the GARCH toolbox in Matlab. A ready-made implementation increases confidence in the estimation results as the optimization of the log-likelihood functions involved is done with a ready-made and tested tool. We begin by estimating a standard GARCH(1,1) model for all four equities. The estimated models are displayed in Table 5 along with $t$-statistics for the estimated parameters.

A large number of observations ($N > 1000$ for NOK, SAM and UPM, $N > 200$ for NES) has been used in the evaluation of the GARCH model. Hence, we approximate the degrees of freedom for the $t$ distribution of the parameters to be infinite. All parameters of the models for NOK, SAM and UPM are significant at a 5% confidence level. For NES, $\alpha_0$ and $\alpha_1$ are not significant, however. We omit the formulation of new model for NES since more models are built below.

The most interesting results on forecasting evaluation are in Table 6. Regular regression-based forecast evaluation statistics and loss functions are displayed for the 70 one-day forecasts obtained with the estimated GARCH(1,1) models. Since the observed squared innovations serve as measures of realisations in our calculations, only corresponding loss functions calculating
Table 5: GARCH(1,1) Model Specifications

\[
\sigma_{t|t-1}^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1|t-2}^2
\]

<table>
<thead>
<tr>
<th></th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \beta_1 )</th>
<th>t Value</th>
<th>t Value</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>5.5856e-005</td>
<td>1.2495</td>
<td>0.76167</td>
<td>6.4983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOK</td>
<td>6.4108e-007</td>
<td>0.012536</td>
<td>0.98539</td>
<td>639.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAM</td>
<td>2.7648e-005</td>
<td>0.17707</td>
<td>0.75136</td>
<td>36.276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPM</td>
<td>1.5821e-006</td>
<td>0.039464</td>
<td>0.95544</td>
<td>144.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: GARCH(1,1) Forecasting Results; Statistics from the regression \( a_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t \) and standard loss-function values.

\[
a_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t
\]

<table>
<thead>
<tr>
<th></th>
<th>( R^2 )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( MSE_2 )</th>
<th>( MAE_2 )</th>
<th>( MAPE_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>0.010841</td>
<td>0.0001185</td>
<td>0.47471</td>
<td>2.6498e-007</td>
<td>0.00037537</td>
<td>8451.6</td>
</tr>
<tr>
<td>NOK</td>
<td>0.030035</td>
<td>0.0014685</td>
<td>-5.7541</td>
<td>1.5836e-007</td>
<td>0.00024496</td>
<td>85157</td>
</tr>
<tr>
<td>SAM</td>
<td>0.0029357</td>
<td>8.7219e-005</td>
<td>0.19345</td>
<td>4.2087e-008</td>
<td>0.00017096</td>
<td>260.74</td>
</tr>
<tr>
<td>UPM</td>
<td>0.024583</td>
<td>0.0004858</td>
<td>-2.5639</td>
<td>5.2637e-008</td>
<td>0.00013886</td>
<td>41.305</td>
</tr>
</tbody>
</table>

deviations in squared variables can be obtained (see Table 2). A plot of the daily squared returns and the GARCH(1,1) forecasts is presented in Figure 3. The figure shows that the GARCH forecasts move in the same direction as the squared returns. The latter are a lot more volatile than the estimated GARCH model forecasts, however. This is probably due to the fact that the shocks \( \epsilon_t \) in the return innovation component \( a_t = \sigma_t \epsilon_t \) are large in comparison to the volatility component \( \sigma_t \), as discussed in Section 4. As a consequence, our volatility forecasts perform poorly measured by \( R^2 \) and loss functions when compared to squared returns.

To give an overview of the quality of the evaluated models we have furthermore included the same statistics for the in-sample data as used in evaluating the forecasts. These quantities correspond to normal model evaluation.
criteria. The results are displayed in Table 7.

In addition to the very common GARCH(1,1) model, we evaluated two other commonly used GARCH-models: GARCH(2,1) and GARCH(1,2) for all equities. These did however not perform significantly better than the GARCH(1,1) model, and the results from these two models are therefore left out.

We also evaluate EGARCH models for the equities. The Matlab implementation of an EGARCH model however differs slightly from Nelson’s original model (23), and has the specification

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left( \frac{\epsilon_{t-i}}{\sigma_{t-i}} - E \left[ \frac{\epsilon_{t-i}}{\sigma_{t-i}} \right] \right) + \sum_{k=1}^{r} \gamma_k \left( \frac{|\epsilon_{t-k}|}{\sigma_{t-k}} \right) + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2),$$

(39)

where the conditioning sub indices of \( \sigma \) has been left out for readability. The difference in the two EGARCH models (39) and (23) is that the former uses standardized past errors as arguments, and adds this quantity as an extra explanatory variable, whereas the latter uses the non-manipulated pure past errors as arguments. The general idea behind the two models remains the same.
Table 7: GARCH(1,1) Model In-Sample Statistics; Statistics from the regression $\sigma_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t$ and standard loss-function values.

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$MSE_2$</th>
<th>$MAE_2$</th>
<th>$MAPE_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>0.04372</td>
<td>6.5998e-005</td>
<td>0.85463</td>
<td>5.6836e-007</td>
<td>0.00049174</td>
<td>45.819</td>
</tr>
<tr>
<td>NOK</td>
<td>0.056915</td>
<td>-2.3401e-005</td>
<td>1.0776</td>
<td>3.3324e-006</td>
<td>0.00071227</td>
<td>44.316</td>
</tr>
<tr>
<td>SAM</td>
<td>0.011079</td>
<td>0.00018786</td>
<td>0.3461</td>
<td>1.3086e-006</td>
<td>0.00035449</td>
<td>24.005</td>
</tr>
<tr>
<td>UPM</td>
<td>0.069977</td>
<td>2.7638e-005</td>
<td>0.91449</td>
<td>5.8627e-007</td>
<td>0.00036125</td>
<td>37.872</td>
</tr>
</tbody>
</table>

Table 8: EGARCH(1,1) Model Specifications

\[
\ln \sigma_t^2 = \alpha_0 + \alpha_1 \frac{|\epsilon_{t-1}|}{\sigma_{t-1}} \left[ \frac{|\epsilon_{t-1}|}{\sigma_{t-1}} - E \left[ \frac{|\epsilon_{t-1}|}{\sigma_{t-1}} \right] \right] + \beta_1 \sigma_{t-1}^2 + \gamma_1 \frac{|\epsilon_{t-1}|}{\sigma_{t-1}}
\]

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_0$</th>
<th>$t$ Value</th>
<th>$\alpha_1$</th>
<th>$t$ Value</th>
<th>$\beta_1$</th>
<th>$t$ Value</th>
<th>$\gamma_1$</th>
<th>$t$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>-0.63853</td>
<td>-1.998</td>
<td>0.066205</td>
<td>0.92185</td>
<td>0.91665</td>
<td>21.869</td>
<td>-0.19393</td>
<td>-2.8559</td>
</tr>
<tr>
<td>NOK</td>
<td>-0.02062</td>
<td>-2.7167</td>
<td>0.04923</td>
<td>7.7309</td>
<td>0.9969</td>
<td>1032.4</td>
<td>0.0015363</td>
<td>0.31235</td>
</tr>
<tr>
<td>SAM</td>
<td>-0.18367</td>
<td>-4.2781</td>
<td>0.10153</td>
<td>5.0863</td>
<td>0.97671</td>
<td>186.61</td>
<td>-0.025923</td>
<td>-1.7337</td>
</tr>
<tr>
<td>UPM</td>
<td>-0.06005</td>
<td>-2.8204</td>
<td>0.10432</td>
<td>6.7951</td>
<td>0.99207</td>
<td>377.46</td>
<td>-0.010934</td>
<td>-1.2516</td>
</tr>
</tbody>
</table>

The estimated EGARCH(1,1) models are displayed in Table 8 and the corresponding forecast statistics are displayed in Table 10. It can be seen from both in-sample and forecasting evaluation statistics that the EGARCH model outperforms the normal GARCH model, although not significantly. The standardized error parameter, $\gamma_1$, is not significant at a 5% confidence level except for in the model for NES. Since the argument of the parameter $\alpha_1$ also reflects the asymmetric reaction to past positive and negative shocks, this does not mean that no leverage effect is detectable in these return series.

In the above mentioned models, the conditional distribution of innovations was assumed to be Gaussian. The same models were evaluated with $t$-innovations, but the results improved marginally if at all.
Table 9: EGARCH(1,1) Model In-Sample Statistics; Statistics from the regression $a_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t$ and standard loss-function values.

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$MSE_2$</th>
<th>$MAE_2$</th>
<th>$MAPE_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>0.14802</td>
<td>-0.00015784</td>
<td>1.3221</td>
<td>5.1088e-007</td>
<td>0.00047196</td>
<td>46.188</td>
</tr>
<tr>
<td>NOK</td>
<td>0.06122</td>
<td>3.9356e-005</td>
<td>0.95071</td>
<td>3.316e-006</td>
<td>0.00070922</td>
<td>46.103</td>
</tr>
<tr>
<td>SAM</td>
<td>0.03300</td>
<td>-5.0484e-005</td>
<td>1.1915</td>
<td>1.2309e-006</td>
<td>0.00032808</td>
<td>23.095</td>
</tr>
<tr>
<td>UPM</td>
<td>0.07085</td>
<td>1.7934e-006</td>
<td>1.0008</td>
<td>5.8533e-007</td>
<td>0.00036008</td>
<td>38.016</td>
</tr>
</tbody>
</table>

Table 10: EGARCH(1,1) Forecasting Results; Statistics from the regression $a_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t$ and standard loss-function values.

<table>
<thead>
<tr>
<th></th>
<th>$R^2$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$MSE_2$</th>
<th>$MAE_2$</th>
<th>$MAPE_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>0.01844</td>
<td>0.00055538</td>
<td>-0.62246</td>
<td>2.9553e-007</td>
<td>0.00039928</td>
<td>192.51</td>
</tr>
<tr>
<td>NOK</td>
<td>0.02619</td>
<td>0.00097972</td>
<td>-3.3379</td>
<td>1.5991e-007</td>
<td>0.00025312</td>
<td>2.6813e+011</td>
</tr>
<tr>
<td>SAM</td>
<td>0.00254</td>
<td>6.7795e-005</td>
<td>0.3034</td>
<td>4.0035e-008</td>
<td>0.00016636</td>
<td>815.71</td>
</tr>
<tr>
<td>UPM</td>
<td>0.03127</td>
<td>0.00059656</td>
<td>-3.1071</td>
<td>5.2974e-008</td>
<td>0.00014542</td>
<td>35.454</td>
</tr>
</tbody>
</table>
Despite the rather poor forecasting performance of the estimated models, our results are in line with most other similar studies. The overview of volatility forecasting studies by Poon and Granger (2003) shows that results in terms of $R^2$ statistics below 5% are normal and that results better than 10% are hardly ever reached with GARCH models estimated and evaluated with closing-price data only. One reason for the poor accuracy of the forecasts is probably due to using the noisy squared returns as proxies for the volatility component of the return series. Generally speaking, the weak autocorrelation shown in the squared return series (see eg. Figure 2) already reveals that any autoregressive model will not succeed in describing the return series.

5.3 Forecasts Based on Warrant Implied Volatility

The implied volatilities from warrants has been calculated for every day with observations from the corresponding warrant. As the risk-free interest rate, the coupon-rate of the newest Finnish government bond\textsuperscript{23} has been used, 4.1%. Each daily implied volatility is calculated using the closing prices of both the underlyer and warrant on the corresponding day. The graph of daily ISD’s from UPM is shown in Figure 4, ISD graphs of for the other equities can be found in Appendix D.

Figure 4: Yearly Standard Deviations of the UPM Kymmenene Stock Implied by daily Closing Prices

\textsuperscript{23}Finnish Government 3Y bond, coupon 4.1% last issue date 14.9.2007, validity 27.8.2007-27.8.2010

35
Table 11: Warrant Implied one-day Forecasting Results; Statistics from the regression \( \hat{a}_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \eta_t \) and standard loss-function values.

<table>
<thead>
<tr>
<th></th>
<th>( R^2 )</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( MSE_2 )</th>
<th>( MAE_2 )</th>
<th>( MAPE_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>0.0023834</td>
<td>0.00020133</td>
<td>0.29137</td>
<td>2.0125e-007</td>
<td>0.00034033</td>
<td>104.38</td>
</tr>
<tr>
<td>NOK</td>
<td>0.0024403</td>
<td>0.00050952</td>
<td>-0.024205</td>
<td>2.5831e-006</td>
<td>0.0010406</td>
<td>83.138</td>
</tr>
<tr>
<td>SAM</td>
<td>0.0068519</td>
<td>2.617e-005</td>
<td>0.63786</td>
<td>4.8097e-007</td>
<td>0.00044604</td>
<td>101.01</td>
</tr>
<tr>
<td>UPM</td>
<td>0.0067032</td>
<td>0.0012637</td>
<td>-2.0874</td>
<td>1.7099e-006</td>
<td>0.00057739</td>
<td>46.028</td>
</tr>
</tbody>
</table>

The standard deviation obtained is technically speaking the yearly standard deviation for the stock. However, since the warrants do not have one year of time to maturity, the obtained measure has to be seen as an annualized equivalent to the volatility expected from the remaining lifetime of the warrant. We use the daily obtained (yearly equivalent) implied standard deviations to create one-day forecasts. We simply multiply the obtained yearly standard deviation by \( \sqrt{\frac{1}{252}} \) to obtain the corresponding one-day volatility, which we use as an ex-ante one-day volatility forecasts. These forecasts are squared and regressed on the following day squared returns in the same way as the regression-based forecast evaluation in Section 5.2. The results are presented in Table 11, and a graph showing the daily squared returns and corresponding forecast for Fortum is shown in Figure 5.

The very poor predictive power in terms of the \( R^2 \) statistic from the regression of the squared returns on squared volatility forecasts does not mean that the warrant ISD is a poor volatility forecast, however. Since the warrant ISD has to be thought of as a measure of volatility over the whole remainder of the life of the option, all that in fact can be expected from a thus calculated daily volatility measure is to reflect the average daily volatility over the remaining life of the warrant. The results however do suggest that the daily ISD calculated from a warrant price seems to be a poor estimate of future daily volatility.
5.4 Comparison of Obtained Volatility Forecasts

The correct interpretation of implied volatility forecasts is a measure of the volatility of the underlyer for the remainder of the life of the derivative. In order to thus be able to compare GARCH and implied volatility forecasts in a “fair” way, a somewhat ad-hoc test will be made. For all equities but UPM\textsuperscript{24}, we will compare the implied volatility for the last two weeks (i.e. 10 trading days) of maturity derived from the respective warrants to a GARCH 10 day forecast for the same period. Both forecast will be compared to the sample standard deviation calculated from the respective stock prices from the same period. This test has obvious flaws, but should illustrate how well or poorly the two different forecasting methods perform in comparison to a well established benchmark. It also gives us the opportunity to use another evaluation criteria than the noisy squared returns. Generally speaking, if more historical price data for warrants would be available this same test could be done for each equity at several instances when a warrant is expiring in order to get a more reliable picture of the relative predictive power of ISD ang GARCH models for each stock.

\textsuperscript{24}The warrant chosen to study UPM matures only in December 2007
Table 12: EGARCH(1,1) Model Specifications Used in Creating 10 Day Forecasts

\[
\ln \sigma_t^2 = \alpha_0 + \alpha_1 \left[ \frac{|e_{t-1}|}{\sigma_{t-1}} - E \left[ \frac{|e_{t-1}|}{\sigma_{t-1}} \right] \right] + \beta_1 \sigma_{t-1}^2 + \gamma_1 \left[ \frac{|e_{t-1}|}{\sigma_{t-1}} \right]
\]

<table>
<thead>
<tr>
<th></th>
<th>( \alpha_0 )</th>
<th>t Value</th>
<th>( \alpha_1 )</th>
<th>t Value</th>
<th>( \beta_1 )</th>
<th>t Value</th>
<th>( \gamma_1 )</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NES</td>
<td>-0.88301</td>
<td>-1.9854</td>
<td>0.17868</td>
<td>2.4388</td>
<td>0.88773</td>
<td>15.708</td>
<td>-0.098566</td>
<td>-2.4018</td>
</tr>
<tr>
<td>NOK</td>
<td>-0.030092</td>
<td>-4.1345</td>
<td>0.056697</td>
<td>8.7629</td>
<td>0.99559</td>
<td>1046.6</td>
<td>-0.0080224</td>
<td>-1.3127</td>
</tr>
<tr>
<td>SAM</td>
<td>-0.17124</td>
<td>-4.453</td>
<td>0.095597</td>
<td>5.4903</td>
<td>0.97847</td>
<td>210.33</td>
<td>-0.028192</td>
<td>-2.2863</td>
</tr>
</tbody>
</table>

The 10 day GARCH forecast is calculated using an EGARCH model where the estimation period used spans from 1 January 2002 until 2 November 2007 for NES and SAM and 1 January 2002 until 5 October 2007 for NOK. Since

\[
r[10] = r_{t+1} + r_{t+2} + \ldots + r_{t+10},
\]

we have

\[
Var[r[10]] = Var[r_{t+1}] + Var[r_{t+1}] + \ldots + Var[r_{t+10}] = \sigma^2_{r_{t+1} | t} + \sigma^2_{r_{t+2} | t} + \ldots + \sigma^2_{r_{t+10} | t},
\]

and thus can compare the square sum of the recursively obtained forecasts for \( \sigma \) to the sample variance calculated from the corresponding closing prices as in (4). The estimated models are presented in Table 12 and the corresponding forecasts are presented in Table 13.

Warrant ISD’s are calculated using the warrant and stock closing prices of the Friday before the two-trading-week period under consideration, i.e. 5 October for NOK and 2 November for NES and SAM. This way the same historical information set is available for both EGARCH and warrant ISD forecasts. The sample standard deviations, EGARCH-forecasted standard deviations and warrant implied standard deviations are presented in Table 14.

The EGARCH forecasts and option ISD’s are clearly larger than the sample standard deviation. For NES and SAM, the EGARCH forecasts are of similar magnitude as the option ISD’s, whereas for NOK, the option ISD is
Table 13: EGARCH(1,1) Ten day volatility forecasts for the period 5-16 November 2007 for NES and SAM, for the period 8-19 October 2007 for NOK

<table>
<thead>
<tr>
<th>Day</th>
<th>NES</th>
<th>NOK</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.018643</td>
<td>0.019891</td>
<td>0.015671</td>
</tr>
<tr>
<td>2</td>
<td>0.018747</td>
<td>0.019936</td>
<td>0.015732</td>
</tr>
<tr>
<td>3</td>
<td>0.018841</td>
<td>0.01998</td>
<td>0.015791</td>
</tr>
<tr>
<td>4</td>
<td>0.018924</td>
<td>0.020024</td>
<td>0.015849</td>
</tr>
<tr>
<td>5</td>
<td>0.018998</td>
<td>0.020068</td>
<td>0.015907</td>
</tr>
<tr>
<td>6</td>
<td>0.019064</td>
<td>0.020111</td>
<td>0.015963</td>
</tr>
<tr>
<td>7</td>
<td>0.019123</td>
<td>0.020155</td>
<td>0.016018</td>
</tr>
<tr>
<td>8</td>
<td>0.019175</td>
<td>0.020199</td>
<td>0.016072</td>
</tr>
<tr>
<td>9</td>
<td>0.019222</td>
<td>0.020242</td>
<td>0.016126</td>
</tr>
<tr>
<td>10</td>
<td>0.019263</td>
<td>0.020286</td>
<td>0.016178</td>
</tr>
</tbody>
</table>

Table 14: EGARCH(1,1) volatility forecasts, warrant ISD’s and real sample standard deviation for 10-day comparison period. Both forecasts overestimate the volatility for the period.

<table>
<thead>
<tr>
<th></th>
<th>NES</th>
<th>NOK</th>
<th>SAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGARCH forecast</td>
<td>0.0601</td>
<td>0.0632</td>
<td>0.0504</td>
</tr>
<tr>
<td>Warrant ISD</td>
<td>0.0474</td>
<td>0.1082</td>
<td>0.0439</td>
</tr>
<tr>
<td>Sample Standard Deviation</td>
<td>0.0157</td>
<td>0.0250</td>
<td>0.0115</td>
</tr>
</tbody>
</table>
significantly larger than the EGARCH forecast. Because of the small sample studied, it is difficult to draw conclusions.

The better ISD estimates than the EGARCH estimates for NES and SAM gives some indication of option ISD outperforming the EGARCH model. For NOK, the EGARCH estimate is however closer to the sample standard deviation than the option ISD.

Since both EGARCH and warrant ISD forecasts for NES and SAM are of similar magnitudes, it is possible that the latter arises because traders use some GARCH model in their estimation of volatility. If this is the case, the difference may be due to a different volatility model being used or the incorporation of some additional information, or simply because a different option pricing model or parameters are used. The warrant ISD’s are all pretty close to four times the sample standard deviation.

6 Conclusions

An overview of different volatility forecasting models and their estimation and evaluation has been discussed in the first four sections of this study. In the subsequent sections GARCH models and warrant ISD’s have been estimated for Finnish equities.

Our results from comparing obtained GARCH forecasts to out of sample data is in line with the results from similar studies; the forecasts perform poorly in terms of an $R^2$ statistic where the volatility forecasts are regressed on squared returns. This does however not necessarily mean that the obtained forecasts are of poor quality, since the squared returns used as volatility proxies are noisy. The squared return series of the equities studied showed only weak autocorrelation, whereas previous studies of US equities and exchange rates have documented significant autocorrelations.

Warrant ISD’s were calculated from NTM warrants traded on the Helsinki exchange. Because of an unfortunate practise of removing price data for expired warrants followed by data providers, a limited sample of warrants was studied. Obtained warrant ISD’s were in line with GARCH forecasts. Based on our results neither of the two forecasting methodologies can reliably be said to outperform the other.
Two interesting fields for further study come to mind. First, the combination of different volatility models from different model classes. This subject has received little attention in the literature, although some promising results have been obtained in studies combining historical volatility based forecasts with options implied forecasts. This approach would be especially interesting in a small market as the Finnish one, where it is reasonable to believe that pricing anomalies mainly due to relatively low liquidity in the warrants markets may corrupt implied volatility. Another equally interesting and practically relevant subject of study is the forecasting the warrants implied volatility itself. The author has not come across studies on this subject. The benefits from correctly forecasting implied volatilities, or even the direction in which this will move are obvious. By constructing different warrant combinations, or portfolios of warrants and the underlying stock, the information could be used to create a trading strategy based on expected developments in the options implied volatility. Models of this kind would be especially valuable to institutional or professional investors for whom active trading and the usage of leverage is possible.
A Tests for Heteroscedasticity

The Ljung-Box-Pierce Q test statistic is calculated as

\[ Q = N(N + 2) \sum_{k=1}^{l} \frac{\rho_k^2}{N - k} \]  

(40)

where \( N \) is the number of observations, \( l \) is the number of autocorrelation lags included in the test and \( \rho_k^2 \) is the squared sample autocorrelation at lag \( k \) calculated as

\[ \rho_k = \frac{\sum_{t=k+1}^{N} (a_t^2 - \bar{a}^2)(a_{t-k}^2 - \bar{a}^2)}{\sum_{t=1}^{N} (a_t^2 - \bar{a}^2)}, \]

(41)

where \( \bar{a}^2 \) denotes the sample mean of the squared residual series. If the sample value of \( Q \) exceeds the critical value of a chi-square distribution with \( l \) degrees of freedom, then at least one value of \( \rho \) is statistically different from zero at the specified significance level. Engle’s ARCH test\(^25\) is equivalent to the \( F \) statistic for testing \( \alpha_i = 0 \ i = 1 \ldots p \) in the regression

\[ a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_p a_{t-p}^2 + e_t. \]

Denoting the square sums \( TSS = \sum_{t=p+1}^{N} (a_t^2 - \bar{a}^2)^2 \) and \( RSS = \sum_{t=p+1}^{N} \hat{e}_t^2 \), where \( \hat{e}_t \) are the errors from the previous least squares regression, the test statistic becomes

\[ LM = \frac{N - 2p - 1}{p} \frac{TSS - RSS}{RSS}, \]

(42)

which conditioned to the null hypothesis asymptotically follows a \( \chi^2 \)-distribution with \( p \) degrees of freedom.

\(^{25}\)The test is a Lagrange multiplier test, but in the form presented below is popularly known as an ARCH test
B Price Graphs

Figure 6: Daily Closing Prices of Neste Oil

Figure 7: Daily Closing Prices of Nokia
C Return Graphs

Figure 8: Daily Closing Prices of Sampo

Figure 9: Daily Log Returns of Neste Oil
Figure 10: Daily Log Returns of Nokia

Figure 11: Daily Log Returns of Sampo
Figure 12: Daily Log Returns of UPM

D Graphs of Daily Implied Standard Deviation

Figure 13: Yearly Standard Deviations of the Neste Oil Stock Implied by daily Closing Prices
Figure 14: Yearly Standard Deviations of the Nokia Stock Implied by daily Closing Prices

Figure 15: Yearly Standard Deviations of the Sampo Stock Implied by daily Closing Prices
References


