UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Volatility Forecasting in Bull & Bear Markets

Karl Oskar Ekvall

Graduate School

Master of Science in Finance
Master Degree Project No. 2012:94
Supervisor: Mattias Sundén

Volatility Forecasting in Bull & Bear Markets*

Karl Oskar Ekvall†

This version: June 2012

Abstract

This thesis considers the performance of variance forecasting in bull and bear markets. Three asset indices, the DAX, the Standard & Poor's 500 and the CurrencyShares Euro Trust, are split into bull and bear periods whereby variance forecasting is evaluated in the two states. I employ a simple moving average, an EWMA, implied volatilities from official volatility indices and three GARCH specifications; a GARCH (1,1) and EGARCH(1,1) with Student's t errors and a GARCH (1,1) with Hansen's skewed t errors. I compute 30 days ahead variance forecasts using daily data and the true latent variance is approximated by the intra-month realized variance. Performance is measured by the R^2 from regressing the realized variance on the estimated variance, the QLIKE statistic and the MSE. I find that implied volatilities forecast best in bull markets and that the GARCH and EGARCH forecast best in bear markets. In general, the predictions' R^2 and QLIKE statistics suffer 30 % - 50 % in bear markets and the MSE is as much as 15 times higher compared to bull markets.

*Thesis for the Master of Science in Finance at University of Gothenburg

†Doktor Saléns Gata 15, 41322 Göteborg, Sweden. +46709700719. Email: k.o.ekvall@gmail.com

Acknowledgments

First, I would like to express my gratitude to Mattias Sundén for his comments and suggestions, which greatly improved this thesis, and for being a very helpful, accommodating supervisor throughout the whole process. A big thanks also to Lin Fickling for help with proof reading.

On a personal level I would like to thank Lin for being who she is, always by my side; it means everything to me. I am also very grateful for all the support from my parents, Hans and Barbro. Lastly, I also want to mention Johan Mellberg for being the best of companions throughout the master program leading up to this thesis.

Contents

1 Introduction **1**

2 Theoretical Framework **4**
2.1 Volatility Proxy 4
2.2 Moving Average 6
2.3 Exponentially Weighted Moving Average 6
2.4 Implied Volatility 7
2.5 Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Models . 8
2.5.1 The Distribution Assumption and Its Implications 11
2.6 Performance Measures 13
2.7 Bull and Bear Markets 15

3 Empirical Methodology **17**

4 Data and Model Fitting **23**

5 Results and Analysis **29**
5.1 Bull and Bear Market Results 29
5.2 Results without bull/bear split 32
5.3 Analysis 37

6 Conclusions **42**

A Appendix **48**
A.1 VIX Methodology 48
A.2 Software 48

1 Introduction

Much of economics is concerned with what may happen in the future and such future expectations are relevant in everything from microeconomics to asset pricing to corporate finance. In order to cope with the uncertainty of the future one can rely on a number of different techniques. Guessing what will happen tomorrow by using information of today of course lies at the core of this concept and is also practiced in a wide range of sciences outside economics. This type of forecasting could be applied to most anything observable over time, such as interest rates, the number of bacteria in a certain substance, default risks or even the number of customers in a store etc. One of the most studied phenomena in finance is the variability of some asset's returns, its volatility. Although an asset's volatility is interesting in itself it also prices derivatives connected to that asset and it has important implications for, among other things, risk management in general and hedging in particular. The importance of volatility through derivatives is underlined by the huge size of today's derivative markets. For example, in the fourth quarter of 2011 U.S Commercial banks alone held derivatives with a notational amount of \$248[1] trillion, to be compared with the US 2011 GDP of around \$15 trillion. By the big part volatility plays in financial derivatives it is apparent that the behavior of tomorrow's volatility is of great interest today, and accordingly a vast literature on volatility forecasting already exists.

The most influential models are the Autoregressive Conditional Heteroscedasticity (ARCH) model due to Engle (1982) and the Generalized ARCH (GARCH) due to Bollerslev (1986) which both deal with how to model time varying conditional variance. There exist many papers devoted to the application of these models and their extensions, e.g. Andersen et al. (2005) and Figlewski (1997) both offer practical advice on how to apply available volatility forecasting theory in different settings. Much of the existing literature is concerned with theoretical or empirical comparisons of different forecasting models (see Poon and Granger (2003) for an overview) in a certain isolated setting or market. Poon and Granger (2003) suggest that more work is needed to understand how models behave under different market conditions but I have only found two papers considering this issue in terms of bull and bear markets; Brownlees et al. (2011) in "A Practical Guide to Volatility Forecasting Through Calm and Storm" examine the effect of the 2008 financial crisis on forecasting performance and Chiang and Huang (2011) conduct a brief comparison of bull and bear market results when using GARCH models to forecast implied volatility. Although touching upon the issue of bull and bear markets, these papers focus more on other aspects and I have not found any paper dedicated to how these different market states affect the predictions. Bull and bear markets are common terms

[1] US Department of the Treasury, http://www.occ.gov

and have previously been examined scientifically, although rarely in relation to volatility forecasting. Lunde and Timmermann (2004) as well as Pagan and Sossounov (2003) offer ways of modeling bull and bear states. I apply a variant of the Pagan and Sossounov methodology in this thesis when attempting to answer the question 'How is volatility forecasting affected by bull and bear markets?'. This main question is approached by focusing on two sub-questions and then consolidating the findings:

1. How is the relative performance among forecasting techniques affected by the market state?
2. How is the absolute performance of volatility forecasting affected by the market state?

The answers to these questions will help decide which techniques should be employed in different scenarios and how to best correct for changes in the market conditions. To find the answers I apply seven volatility forecasting techniques in three different markets and measure the performance by three different measures, or loss functions. The relative performance is assessed by comparing the models' respective loss functions with the test proposed by Diebold and Mariano (1995). The forecast horizon is 30 days and I proxy the true latent variance with the intra-month realized variance, argued to be the most appropriate proxy for latent variance by Andersen et al. (2004) among others. Note here that I compare variance forecasts, not standard deviation forecasts which is common in the literature. I compare the predictions from three differently weighted moving average models, a GARCH(1,1), an EGARCH(1,1) due to Nelson (1991) and implied volatilities from official volatility indices. Both GARCH models are employed with Student's t errors and the regular GARCH is in addition employed with Hansen's t errors as described in Hansen (1994), allowing for skewness. I carry out the comparison in the German DAX index, Standard & Poor's 500 (S&P 500) and CurrencyShares Euro Trust ($US/Euro), tracking the $US/Euro exchange rate, after splitting each index into bull and bear periods.

I find that the implied volatility forecasts are superior in bull markets where the level of volatility as well as volatility of volatility is lower and the market more informationally efficient. The GARCH specifications give the best forecasts in bear periods although the implied volatilities are good (second best) also in this setting. All predictions' R^2 suffer approximately 30 % - 50 % , QLIKE 30 % - 40 % and the MSE is often around 15 times higher in bear markets than in bull markets, confirming the findings of Chiang and Huang (2011). In line with Figlewski and Wang (2000) I find leverage effects in the stock markets that could (and maybe should) be interpreted as "market down effects" and this benefits the EGARCH *vis-à-vis* the other models in scenarios where there are significant leverage effects. The EGARCH is the only model that sometimes performs better in bear markets than in bull markets and therefore handles the shift between states best in relative terms. I also find that deviations from Gaussian white noise in the return processes, such as fat tails, skewness and volatility clustering, are more apparent in stock returns than in the returns of the $US/Euro and this causes all models to be outperformed by a simple moving average. It is also found that empirical distributions changing over time punish forecasts based on more flexible theoretical distributions and thus makes it hard to improve predictions by accounting for skewness and excess kurtosis. An interesting finding not directly related to the main research question is that, according to the Kuiper statistic (Kuiper, 1962), allowing for skewness and excess kurtosis through Hansen's t distribution is not enough to approximate the stock returns' distributions with statistical significance.

The main caveat to the ranking of models through my results is that the ranking is not consistent over loss functions. A forecaster has to take heed when choosing which model to use, so as to match his own preferences rather than just looking at the overall performance. My ranking only serves as an overview of the performance and does not take the forecaster's preferences into account. Moreover, the differences in prediction errors are sometimes so small that the loss functions cannot differ between models at the 95 % confidence level. This leads me to conclude that, depending on the loss function of interest, one does not always have much to gain by using a more advanced model compared to a simple weighted moving average. A major problem in finding a superior technique is that market behavior changes over time, causing the shape of the error distribution to change significantly over time.

2 Theoretical Framework

This section covers the basic theory needed to understand the models employed and analyzed in this thesis. The word volatility is a bit vague and can refer to different things but it is closely related to the variability of a stochastic process. This thesis focuses on the variability in returns of different financial assets and indices tracking the development of such assets. Thus, I henceforth use volatility interchangeably with variability of returns. The literature is not uniform on whether volatility refers to standard deviation or variance, although the former is more common. Therefore, most techniques and methods are named in terms of volatility, whether they pertain to standard deviation or variance. The reader should bear in mind that this thesis considers variance forecasting and not standard deviation forecasting, although the results are generalizable to either case.

2.1 Volatility Proxy

I define the volatility measures by considering a standard setting in financial economics where the analyzed asset's (log-) price development over time is assumed to be governed by the following differential equation

$$dP_t = \mu_t dt + \sigma_t dW_t$$

where $P_t = \ln(\text{Price}_t)$, t is a time index, μ the drift of the process and W_t is a standard Brownian Motion, representing the stochastic part of asset prices, and thus $W_t \sim N(0, 1)$. Here and throughout, lower case letters are reserved for observed values while capital and Greek letters are used for random variables.

With the given setting the price of an asset at time t is given by:

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s.$$

With this definition σ scales the standard deviation of the process and σ is therefore one, and arguably the most common, measure of volatility. As such it is also one among many measures of uncertainty and risk. Furthermore, we recall that prices, P_t, are expressed in logarithmic form and thus the log-returns $R_t = \ln(\text{Price}_t/\text{Price}_{t-1}) = \ln(\text{Price}_t) - \ln(\text{Price}_{t-1})$[2] are given by

$$R_t = P_t - P_{t-1} = \int_{t-1}^t \mu_s ds + \int_{t-1}^t \sigma_s dW_s,$$

from which, under the assumptions that there are no jumps in the process and that σ_t

[2] $\ln(\text{Price}_t/\text{Price}_{t-1}) \approx (\text{Price}_t - \text{Price}_{t-1})/\text{Price}_{t-1} = \Delta\text{Price}_t/\text{Price}_{t-1}$ for small ΔPrice_t so that the log returns are approximately equal to the discrete returns

and W_t are independent, we can deduce

$$R_t \sim N\left(\int_{t-1}^{t} \mu_s ds, \int_{t-1}^{t} \sigma_s^2 ds\right)$$

Now, since volatility is related to the variability of returns it is natural to look at the variance of this distribution and label it for an arbitrary time period of length h such that

$$IV_t^{t+h} = \int_t^{t+h} \sigma_s^2 ds$$

which accordingly is called the 'integrated variance' or sometimes also the 'integrated volatility'. This is the technical definition on which I base the analysis in this thesis.

All price processes are discrete in reality, or at least discretely observed, and the instantaneous returns as well as the parameter σ cannot be directly observed (σ is often called the latent volatility), so they have to be approximated. One unbiased and consistent estimate of σ_t^2 is the square returns of the series in period t. However, this proxy is very noisy in that it itself often exhibits high volatility and Andersen et al. (2003), among others, argue that the so called Realized Volatility/Variance (RV) is a better proxy for evaluating volatility forecasts. This measure approximates the integrated variance by a sum of observed values of intra-period squared returns. Specifically, the integrated variance for one period, here measured in months, can be approximated as

$$IV_0^1 = \int_0^1 \sigma_s^2 ds \approx \sum_{i=1}^{30} R_i^2$$

where i is an index for the days in the examined month and R_i is the centralized daily return of day i. Notably, the approximation in theory[3] becomes better as the sample frequency of intra-period returns increases and we have that (Poon and Granger, 2003)[4]

$$\left(\sum_{i=0}^{m-1} R_{m-i}^2\right) \xrightarrow{p} \left(\int_0^1 \sigma_s^2 ds\right) \Leftrightarrow \mathbb{P}\left[\left|\left(\sum_{i=0}^{m-1} R_{m-i}^2\right) - \left(\int_0^1 \sigma_s^2 ds\right)\right| > \varepsilon\right] \to 0, \text{ as } m \to \infty$$

where ε is an arbitrarily small real number, m is the number of intra-period observations and 1/m thus the time between observations, measured in the period-unit. This approximation of integrated variance is henceforth used as a proxy for the true integrated variance to evaluate the accuracy of the computed variance forecasts. The computed variance forecasts/predictions are all different ways of finding the expected variance over next thirty day's by using present

[3] Andersen et al. (2011) points out that if too frequent observations are used in application, noise introduced by the market distorts the estimates rather than improves them, but that this doesn't occur until the frequency is 'ultra high'.

[4] See e.g Wooldridge (2001) for an explanation of convergence in probability.

information. In mathematical terms, for each day, t, of the sample the forecast/prediction is given by the conditional expectation

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right],$$

where t is now a daily time index, so that the period from t to $t+29$ is (roughly) the coming month, and where $\mathcal{F}_t$ is the information available at time t.

2.2 Moving Average

Moving average variance estimates future variance by its moving average value, equally weighted for a given number of past observations and scaled by time. Notably, models scaling variance by time implicitly assumes constant future variance and uses the property that variances are additive for independent increments. Using the centralized squared returns the expectation of the coming month's variance is given by

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right] = 30\frac{1}{T}\sum_{i=1}^{T} R_{t-i}^2$$

where R_t^2 denotes the centralized squared return in day t, and T is the number of historical observations used to predict IV_t^{t+29}.

While its simplicity makes the moving average appealing it has a number of important shortcomings; it says nothing about how variance evolves and why it takes on certain values. The model also puts equal weight on all observations, recent as well as older. This only makes sense if one indeed believes that the most recent observations of the process hold no more information about its future development than older observations. This potential shortcoming is what merits the inclusion of the Exponentially Weighted Moving Average (EWMA) model.

2.3 Exponentially Weighted Moving Average

The EWMA modifies the moving average model by putting more weight on recent observations than on older ones. Instead of weighting by $1/T$ the EWMA is defined in the following way

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right] = 30\left(\lambda R_{t-1}^2 + \lambda^2 R_{t-2}^2 + \lambda^3 R_{t-3}^2 + \dots\right)$$

which can be re-written on a simpler form using the recursive relation

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right] = \lambda\mathbb{E}\left[IV_{t-1}^{t+28}|\mathcal{F}_{t-2}\right] + 30\left(1-\lambda\right)R_{t-1}^2$$

where λ is a constant parameter between zero and one set by the researcher and R_t^2 is the centralized squared return at time t . The forecast is again scaled by time to predict the coming month's variance.

A common value of the weight parameter λ used in the financial economics literature is 0.94. This is much due to its use in the MSCI software 'RiskMetrics' and I follow this example. The EWMA model is, like the moving average, unconcerned with the data generating process of variance and makes no attempt to explain the 'why' and 'how' of the process. We again assume that past observations of variance say something about the future realizations but we have no explanation as to why this might be. The only difference compared to the simple moving average is that we now believe more recent observations have a higher relevance for the future than older observations. Although very old observations are still allowed to influence they are practically negligible due to the decreasing weight.

2.4 Implied Volatility

There are several ways to infer the market's expectation of volatility. When talking about implied volatility one usually refers to the volatility for which the observed market prices are "fair", or in other words no arbitrage, equilibrium prices. A common way to find these volatilities is to back them out from some model that one assumes the option prices to satisfy. The obvious example is the Black-Scholes (B-S) (Black and Scholes, 1973) model which assumes (for example) efficient, frictionless markets with no arbitrage possibilities as well as stock prices following a geometric Brownian motion with constant drift and variance. It is well known that although B-S is an elegant and easy-to-handle formula it is inconsistent with observed market prices; the B-S implied volatility varies over both strike price and 'moneyness', creating the so called volatility smiles and smirks.

There is vast literature with suggestions on how to improve and adjust the B-S implied volatilities but instead of doing this myself I make use of some of the official volatility indices available, each tied to an underlying stock or currency exchange rate index. All of the indices are calculated using the method developed by the Chicago Board Options Exchange (CBOE) for their 'VIX'-indices. The indices are model-free, in the sense that they do not impose restrictive assumptions on how options are priced in the market. The implied volatility is instead found via applying the no-arbitrage argument to the prices of (replicated) variance swaps, which are priced by the market and thus gives an expectation of the variance under the risk-neutral measure. The formula used when computing VIX and when finding the predictions in this thesis is[5]

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right] = \left(\frac{VIX_{t-1}}{100}\right)^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2} e^{rT} Q(K_i) - \frac{1}{T}\left(\frac{F}{K_0} - 1\right)^2$$

where VIX_t denotes the observed volatility index value at time t, T_i is the time to expiration for option i (in minutes divided by the number of minutes in a year) F a forward index level (on the underlying), K_0 the first strike below the forward index level, K_i the strike of i^{th} out of the money option, ΔK_i the interval between strike prices given by $0.5\,(K_{i+1} - K_{i-1})$, r the risk free interest rate and $Q(K_i)$ the midpoint of the bid-ask spread for each option with strike K_i. Some of the parameters merit further explanation and this can be found in the Appendix.

In short, VIX gives a measure of the volatility implied by the market in the sense that, given the observed market prices, the volatility given by VIX ensures that there are no arbitrage possibilities in option portfolios or equivalently in the variance swap rates (Carr and Wu, 2005).

2.5 Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Models

The GARCH model is an extension of the ARCH model of Engle (1982) and shares the same basis. The ARCH model was proposed as a way of modeling the variance of a process in addition to its mean. As the name indicates it allows for conditional heteroscedasticity, i.e. conditional non-constant variance. Looking at a process, R_t, the conditional mean μ_t and variance σ_t^2 is defined as

$$\mu_t = \mathbb{E}\left[R_t|\mathcal{F}_{t-1}\right] \text{ and } \sigma_t^2 = \mathrm{Var}\left(R_t|\mathcal{F}_{t-1}\right) = \mathbb{E}\left[\left(R_t - \mu_t\right)^2|\mathcal{F}_{t-1}\right]$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time t-1. The ARCH model is a simultaneous explanation of the mean and variance. In order to model the variance one must also model the mean μ_t so as to yield a series satisfying

$$R_t = \mu_t + \sigma_t\varepsilon_t \Leftrightarrow R_t - \mu_t = \sigma_t\varepsilon_t \tag{1}$$

where ε_t is a sequence of independent and identically distributed (*iid*), mean zero and unit variance random variables.

[5] For more details see CBOE's 'VIX White Paper' and Demeterfi et al. (1999)

The mean μ can be modeled in a number of different ways, including exogenous parameters or simply past values of the series itself, as long as any linear dependence over time is removed. Since the linear dependence over time is to be removed and since financial return series often exhibit weak dependence in the first moment a low order autoregressive (AR) model is often satisfactory (Tsay, 2005). In this thesis I only model constant means (in effect an AR(0)/ARMA(0,0) model) since this specification gives the lowest Bayesian Information Criterion (BIC)[6] value for all return series when compared to ARMA(R,M)[7] models for all R and M considered. Since the ARMA specification is never implemented I skip explaining it for brevity.

With the mean accounted for the shock ($\xi_t \equiv \sigma_t \varepsilon_t$) of the return series is assumed uncorrelated but dependent (in the second moment) in the ARCH(p)-model such that

$$\xi_t = \sigma_t \varepsilon_t, \text{ where } \sigma_t^2 = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \alpha_2 \xi_{t-2}^2 + \cdots + \alpha_p \xi_{t-p}^2 \tag{2}$$

where $\alpha_0, .., \alpha_{t-p}$ are coefficients to be estimated. Conditional independence of ξ_t and ξ_{t-p} , for arbitrary $p \geq 1$, can be shown by noting that

$$\begin{aligned} \mathbb{P}\left[\xi_t < x, \xi_{t-p} < y | \mathcal{F}_{t-1}\right] &= \mathbb{P}\left[\varepsilon_t \sigma_t < x, \varepsilon_{t-p} \sigma_{t-p} < y | \mathcal{F}_{t-1}\right] \\ &= \mathbb{P}\left[\varepsilon_t < \frac{x}{\sigma_t}, \varepsilon_{t-p} < \frac{y}{\sigma_{t-p}} | \mathcal{F}_{t-1}\right] \\ &= \mathbb{P}\left[\varepsilon_t < \frac{x}{\sigma_t} | \mathcal{F}_{t-1}\right] \times \mathbb{P}\left[\varepsilon_{t-p} < \frac{y}{\sigma_{t-p}} | \mathcal{F}_{t-1}\right] \\ &= \mathbb{P}\left[\xi_t < x | \mathcal{F}_{t-1}\right] \times \mathbb{P}\left[\xi_{t-p} < y | \mathcal{F}_{t-1}\right], \end{aligned} \tag{3}$$

where the equality in line two is due to that the values constituting σ_t and σ_{t-p} are known when conditioning on $\mathcal{F}_{t-1}$ (and can thus be treated as constants). The equality in line three, establishing independence, is clear from that ε_t is assumed *iid*, i.e. independent for different t.[8] Now, since ξ_t is a function of past values this structure is able to explain the so called volatility clustering empirically observed in asset returns; variance is allowed to vary over time and big shocks are likely to be followed by more big shocks. In summary, the ARCH model implies time varying conditional expected variance (conditional heteroscedasticity), constant

[6] All employed tests are described in the Empirical Methodology.

[7] ARMA(R,M): $r_t = \alpha + \varepsilon_t + \sum_{i=1}^{R} \beta_i r_{t-1} + \sum_{j=1}^{M} \gamma_j \varepsilon_{t-1}$, where α is a constant, $\varepsilon_t \sim iid(0, \sigma^2)$, and β_i and γ_j are parameters to be estimated $\forall i, j$.

[8] To be exact, Equation (3) holds 'almost surely', but not 'surely', since its validity rests on expectations conditional on the information generated up to time t-1, $\mathcal{F}_{t-1}$. See Williams (1991) for details on conditional expectations and thereto related properties. In essence, we cannot say that Equation 3 is always true but well that $\mathbb{P}$ [Equation (3) is True] $= 1$

expected unconditional variance and an unconditional mean of zero for ξ_t. The unconditional mean of zero can be seen by the law of total expectations

$$\mathbb{E}\left[\xi_t\right] = \mathbb{E}\left\{\mathbb{E}\left[\xi_t|\mathcal{F}_{t-1}\right]\right\} = \mathbb{E}\left\{\sigma_t\mathbb{E}\left[\varepsilon_t\right]\right\} = 0$$

where the last equality is due to the fact that ε_t is assumed to have zero mean $\forall t$.

The GARCH(p,q) model proposed by Bollerslev (1986) builds on the ARCH by generalizing it to allow inclusion of past values of σ to write the error term ξ_t as

$$\xi_t = \sigma_t\varepsilon_t, \text{ and } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i\xi_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \tag{4}$$

where the constant α_0 and the parameters α_i and β_j are to be estimated so that the imposed model fits the data at hand as well as possible. For the GARCH(1,1) employed in this thesis the 30 days ahead variance prediction based on Equation (4), regardless of the assumed distribution, is given by[9]:

$$\mathbb{E}\left[IV_t^{t+29}|\mathcal{F}_{t-1}\right] = \sum_{i=0}^{29}\sigma_{t+i}^2 = \sum_{i=0}^{29}\frac{\alpha_0\left[1-(\alpha_1+\beta_1)^i\right]}{1-\alpha_1-\beta_1} + (\alpha_1+\beta_1)^i\left(\alpha_0+\alpha_1\xi_{t-1}+\beta_1\sigma_{t-1}^2\right),$$

where i gives the 1+i step (day) ahead forecast and other notation is as before.

As an extension of the GARCH-model Nelson (1991) proposes the exponential GARCH (EGARCH). This model allows for asymmetric effects in the return-series. More specifically, it allows for different effects of positive and negative return on variance, something which is often observed in financial time series, why I include the EGARCH model in this thesis. The EGARCH(m,s) model can be written on the following form:

$$\ln\left(\sigma_t^2\right) = \alpha_0 + \sum_{i=1}^{s}\left[\alpha_i\left(|\varepsilon_{t-i}| - \mathbb{E}\left[|\varepsilon_{t-i}|\right]\right) + \gamma_i\varepsilon_{t-i}\right] + \sum_{j=1}^{m}\beta_j\ln\left(\sigma_{t-j}^2\right) \tag{5}$$

where α_0 is a constant, α_i and γ_i parameters tied to the i:th ARCH-effect, $\varepsilon_t = \xi_t/\sigma_t$ as before, β_i a parameter tied to the i:th GARCH-effect. Note that γ_i is here capturing the so called leverage effect, or "sign effect", while α_i captures the "magnitude effect". If negative returns contribute more to variance than positive, γ will be negative so that a negative ε_{t-1} increases the log-variance more than a positive ε_{t-1}, and vice versa. A significant advantage of the EGARCH compared to the regular GARCH is that the former allows for negative pa-

[9] See e.g. Tsay, 2005 p. 115 for a complete derivation

rameters since $\ln\left(\sigma_t^2\right)$ (in contrast to σ_t^2) can be negative and still well-defined, i.e. σ_t^2 will always be positive.

Note that the EGARCH forecast, in contrast to the GARCH forecast, is depending on the distribution assumption for ε_t. Due to this and the fact that the model is defined in log-variance rather than variance, it is more involved, and often impossible, to obtain an analytical expression for the forecast. For example, when using the Student's t distribution we have (Tsay, 2005, p.124):

$$\mathbb{E}\left[|\varepsilon_{t-i}|\right] = \frac{2\sqrt{\nu-2}\Gamma\left(\nu/2+1/2\right)}{\left(\nu-1\right)\Gamma\left(\nu/2\right)\sqrt{\pi}},$$

where $\nu \in]2,\infty[$, $\lambda \in]-1,1[$ and $\Gamma\left(x\right)$ denotes the gamma function given by

$$\Gamma\left(x\right) = \int_0^\infty z^{x-1}e^{-z}dz$$

With this expectation we get the one day ahead log-variance prediction $\mathbb{E}\left[\ln(\sigma_t^2)|\mathcal{F}_{t-1}\right]$ from Equation (5). The motivation for using the Student's t and some characteristics of different distributions are discussed in section 2.5.1 below.

The EGARCH(1,1) *1+i*-day ($i \geq 1$) ahead log-variance prediction can be written as as[10]

$$\mathbb{E}\left[\ln(\sigma_{t+i}^2)|\mathcal{F}_{t-1}\right] = \alpha_0\sum_{j=0}^{i-1}\beta^j + \beta^i\mathbb{E}\left[\ln(\sigma_t^2)|\mathcal{F}_{t-1}\right]$$

Thus, we have an analytical expression for the daily log-variances and from this the prediction for the coming month's integrated variance is obtained numerically[11]; no general closed form exist for this forecast using EGARCH models (Andersen et al., 2005).

2.5.1 The Distribution Assumption and Its Implications

In basic stock return models it is common, yet well known erroneous, to assume $\varepsilon_t \sim N(0,1)$, where ε_t is the error term in Equation (1). There is an abundance of literature showing that stock returns generally have a higher peak and fatter tails, i.e. excess kurtosis, than what is implied by the normal distribution (see for example Karlin and Taylor (1998) or Hull (2005)). It is common to correct for this by using the Student t-distribution and Brownlees et al. (2011) further argue that *"The Student t down-weights extremes with respect to the Gaussian, thus it can provide a more robust estimate of the long run variance"* (They find, however, that

[10] See Ederington and Guan (2005) for details.

[11] I employ the MatLab function *garchpred* from the Econometrics Toolbox, for more info see http://www.mathworks.se/help/toolbox/econ/garchpred.html

using the t-distribution did not on average improve forecasts relative to using the Gaussian). I use the t-distribution because of its theoretical advantage of being fatter tailed than the Gaussian. In addition, I also consider the skewed t-distribution due to Hansen (1994). The skewed t-distribution nests the regular Student's t and merits an explanation since it is not as commonly used as the Student's t distribution. A random variable is Hansen's skewed t, or Hansen's t for short, distributed if its density is given by

$$g\left(z|\nu,\lambda\right)=\begin{cases}\beta\gamma\left(1+\frac{1}{\nu-2}\left[\frac{\beta z+\alpha}{1-\lambda}\right]^2\right)^{-(\nu+1)/2} & z<-\alpha/\beta\\ \beta\gamma\left(1+\frac{1}{\nu-2}\left[\frac{\beta z+\alpha}{1+\lambda}\right]^2\right)^{-(\nu+1)/2} & z\geq-\alpha/\beta\end{cases},$$

$$\alpha=4\lambda\gamma\left(\frac{\nu-2}{\nu-1}\right),\ \beta^2=1+3\lambda^2-\alpha^2,\ \gamma=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\left(\nu-1\right)}\Gamma\left(\nu/2\right)}$$

Hansen shows that this is indeed a density and that it reduces to the Student's t distribution when $\lambda=0$. λ is then the skewness parameter of the distribution and ν the degree of freedom. It should also be noted that this distribution is normalized to unit variance, in line with what we want in the GARCH model. Figure 1 shows the shape of Hansen's skewed t distribution's probability density function for different parameter values. When using a degree of freedom (ν) of 300 and 0 skewness (λ) we see in the figure that the distribution is very close to a Gaussian, in line with what is wanted since 0 skew reduces the distribution to Student's t and the Student's t converges to the Gaussian when the degree of freedom is "large". The plotted density with degree of freedom of 13 and skew of -0.1, as we will see in the results, approximates the return distributions of the herein analyzed stock indices.

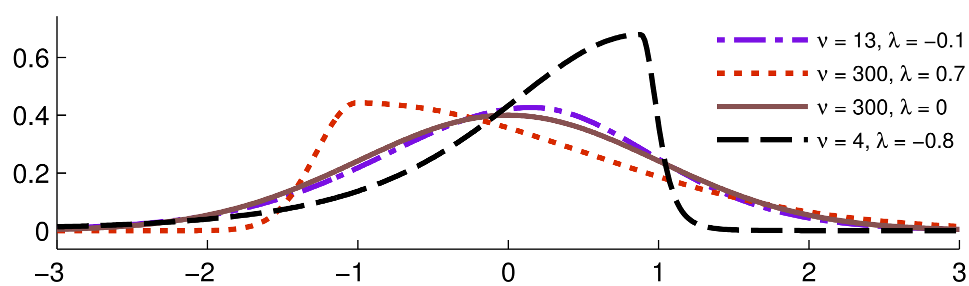


Figure 1: Probability Density Function for Hansen's Skewed t Distribution
Notes: The figure shows the probability density function for Hansen's skewed t distribution for different values of the degree of freedom (ν) and the skewness parameter (λ). A skewness of zero reduces the distribution to a Student's t.

If one uses a non-normalized density the variance needs to be corrected to ensure unity. The variance of a Student's t distributed random variable, denote $\theta_t \sim t(\nu)$, is given by $\nu/(\nu-2)$ where ν is the degree of freedom. The lower ν the fatter the tails and when $\nu \to \infty$, the t-distribution approaches the normal distribution. Thus, in order to fit the fatter tails in stock-returns a low ν is appropriate, usually somewhere between 2-7 (see for example Wilhelmsson (2006) or Andersen et al. (2005)). In this thesis I use likelihood functions that estimate ν and other parameters simultaneously. To ensure that the error term, ε_t in Equation (1), is still of unit variance I set $\varepsilon_t = \theta_t/\sqrt{\nu/(\nu-2)}$.

For the regular GARCH, the conditional log-likelihood function to be maximized is ,due to the *iid* assumption, the log of the product of all conditional densities. The conditional independence is shown in Equation (3) and for the Student's t distribution this product, also fitting ν, is (Tsay, 2005)[12]:

$$\ell\left(\xi_{m+1},..,\xi_T|\vec{\xi}_M,\vec{\alpha},\vec{\beta},\nu\right) = -\sum_{t=m+1}^{T}\left[\frac{\nu+1}{2}\ln\left(1+\frac{\xi_t^2}{(\nu-2)\,\sigma_t^2}\right)+\frac{1}{2}\ln\left(\sigma_t^2\right)\right]+$$

$$+(T-m)\left[\ln\left(\Gamma\left(\frac{\nu+1}{2}\right)\right)-0.5\ln\left((\nu-2)\,\pi\right)\right]$$

where T is the horizon, $\vec{\alpha} = \{\alpha_0,..,\alpha_p\}$, $\vec{\beta} = \{\beta_0,..,\beta_q\}$, $\vec{\xi}_M = \{\xi_1,..,\xi_m\}$.

Since we observe $\{\xi_t\}_0^m$, the likelihood is maximized over the parameters ν and σ_t. These estimates can then be used in Equation (4) to form expectations on future variance. If the estimates of σ_t, denoted $\hat{\sigma}_t$ are correct so that $\hat{\sigma}_t \approx \sigma_t$, we see from Equation (1) that $(\sigma_t\varepsilon_t)/\hat{\sigma}_t = \xi_t/\hat{\sigma}_t \approx \varepsilon_t$. And since the error terms ε_t are assumed *iid* we can check the validity of the estimated mean by testing if $\xi_t/\hat{\sigma}_t$, called the standardized residuals, are uncorrelated over time and check the validity of the estimated variance equation by testing if $(\xi_t/\hat{\sigma}_t)^2$ are uncorrelated over time Tsay (2005).

2.6 Performance Measures

I compare the accuracy of the variance forecasts by three different measures, also called loss functions. Making use of the results of Patton (2011) and Meddahi (2001) I employ what they term robust loss functions. Here, the robustness of a loss function only relates to if it ranks different forecasts correctly, that is, if it ranks forecasts in the same way that they would have been ranked if the true integrated variance was used and not a proxy. When it comes to the

[12] The likelihood for Hansen's t is obtained in the same way, using the given density. Details are found in Hansen (1994). In this thesis the likelihood estimation with Hansen's t is based on MatLab code found in the Oxford MFE Toolbox, http://www.kevinsheppard.com/wiki/MFE_Toolbox

absolute performance Patton (2011) points out that the actual difference between the forecast and the proxy can vary with the noise in the proxy. The three loss functions are:

1. The R^2 from an Ordinary Least Squares (OLS) regression using the following model:

$$RV_t = \alpha + \beta EV_t + \epsilon_t$$

where RV_t denotes the realized variance in period t, α is a constant to be estimated, β the regression coefficient to be estimated, EV_t the estimated variance in period t and ϵ_t an error term capturing the measurement error and all variation in RV_t not explained by the explanatory variable. The R^2 should be interpreted as 'the variation around the mean in the explanatory variables (here only the estimated variance) and a constant explains $100\times R^2$ % of the variation around the mean in the dependent variable (here the realized variance)'.

Since all models are estimated with the same number of parameters as well as on a similar dataset it is more appropriate to compare the R^2 among the different models and indices than otherwise.

2. The average QLIKE loss function defined as

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}(QLIKE_t) = \frac{1}{T}\sum_{t=1}^{T}\left(\ln(EV_t) + \frac{RV_t}{EV_t}\right),$$

where T is the number of observations. The QLIKE loss function is proven by Patton (2011) to be the only robust loss function based on the standardized forecast error RV_t/EV_t. The interpretation of the QLIKE loss function is clear by noting that, if we minimize it, we get the first order condition for an extreme point

$$\nabla QLIKE = 0 \Leftrightarrow \frac{d}{dEV_t}QLIKE_t = \frac{1}{EV_t} - \frac{RV_t}{EV_t^2} = 0,\ \forall t,$$

which is fulfilled *iff* the estimated variance, EV_t, is equal to the realized variance, RV_t. Thus, the lower QLIKE score, the better forecast[13]. We also see from the first derivative with respect to EV_t that the QLIKE is characterized by punishing negative deviations from the correct forecast harder than positive ($EV_t, RV_t > 0$).

[13]The solution indeed a minimum; the second order condition for a minimum is always fulfilled if $EV_t = RV_t \neq 0$ since we then have $d^2QLIKE_t/dEV_t^2 = -(EV_t)^{-2} + 2\left(RV_t/EV_t^3\right) = 1/EV_t^2 > 0$

3. The mean squared error defined as

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(RV_t - EV_t)^2$$

The MSE is characterized by punishing outliers harder than loss functions based on absolute values and is clearly minimized when $EV_t = RV_t$. Moreover, Patton(2011, p.6) states that "...[the MSE] *is the only robust loss function [...] that depends solely on the forecast error,* $RV_t - EV_t$[14].

2.7 Bull and Bear Markets

My definition of bull and bear markets is inspired by Pagan and Sossounov (2003). It may deviate from common notions in several ways since bull and bear markets are used in a colloquial manner and not strictly defined. A common ground is that a bull market is a state of expected capital gains and a bear market the reverse. I define the two market states by looking separately at the price levels of each of my analyzed price processes. Thus, my definition refers to the state of a specific process rather than some overall global state.

Looking at a finite sequence (p_t), or n-tuple, where n is the number of observations, of a price process I define a new tuple

$$(p_{tj}) = P \cup T$$

$$P = (p_t : p_{t-150}, ..p_{t-1} < p_t > p_{t+1}, ..p_{t+150})$$

$$T = (p_t : p_{t-150}, ..p_{t-1} > p_t < p_{t+1}, ..p_{t+150})$$

where t here denotes a daily time index by which the tuples are ordered. I call the tuple P peaks and the tuple T troughs.

From the tuple (p_{tj}) I take out and order elements in the following way:

1. If $p_{t1} \in P$, take the first $p_t \in P$ fulfilling the requirement that there are no other $p_t \in P \cup T$ in the interval $t-100, .., t, .., t+100$ and take this element to a new finite sub-sequence and define it $p_{tj1} \in (p_{tjm})$. To find p_{tj2} take the first $p \in T$ after p_{tj2} in the tuple (p_t) that fulfills the requirement that there are no other $p \in P \cup T$ in the interval $t-100, .., t, .., t+100$. The algorithm continues pick elements, switching between P and T until all (p_{tj}) are examined.

[14] Patton's notation is $\hat{\sigma}^2 - h$.

2. If $p_{t1} \in T$, start with the first $p_t \in T$ fulfilling the requirement that there are no other $p_t \in P \cup T$ in the interval $t - 100, .., t, .., t + 100$ and continue in an analogue way to 1.

By looking at the final tuple (p_{tjm}) as turning points of the market bull and bear market observations are defined as the observations between the turning points in the original tuple (p_t). If the price process had a lower value at the last turning point than at the upcoming, all observations in between are considered bull observations. Accordingly, the time in between these observations is called a bull market or bull state, and vice versa for bear markets. This algorithm ensures that the market is always in a bull or bear state, that it goes from one to the other and that the duration of a market state is at least one hundred days. Figure 2 uses the DAX index to illustrate which points belong to the different tuples in the definition.

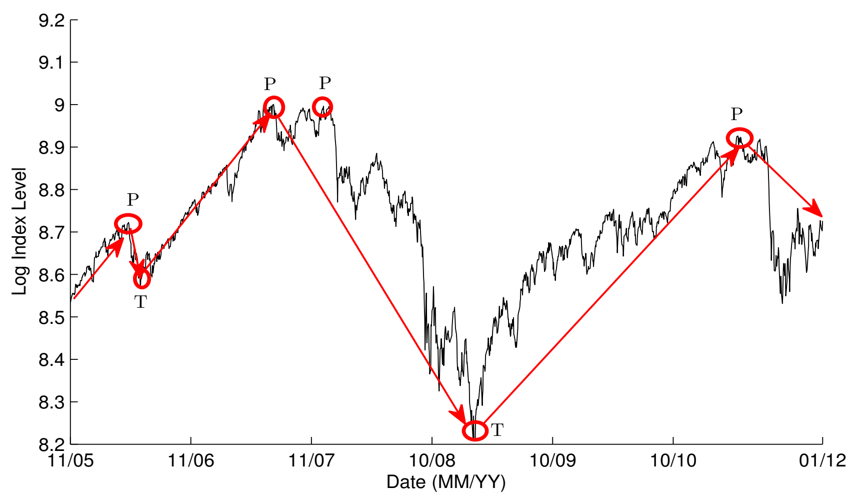


Figure 2: Demonstration of bull and bear market algorithm on the DAX Index.
Notes: The red circles highlight points where the value is either higher or lower than any other observation within 150 days, defined as the tuple $(p_{tj}) = P \cup T$*, where* $P = (p_t : p_{t-150}, ..p_{t-1} < p_t > p_{t+1}, ..p_{t+150})$ *are the peaks in this definition and* $T = (p_t : p_{t-150}, ..p_{t-1} > p_t < p_{t+1}, ..p_{t+150})$ *the troughs, where* p_t *denotes an observation in the tuple of all observations. The arrows show the bull and bear markets resulting from making sure that every extreme point of the market is a turning point and that no turning point is inside 100 observations of another turning point. The third P is "eliminated" because it is following another P and thus points 1,2,3,5 & 6 are market turning points, which tuple is denoted* (p_{tjm}).

3 Empirical Methodology

For the analysis I collect data on three indices[15] and their respective volatility index. The indices (Underlying Index Ticker, Volatility Index Ticker) are Deutscher Aktien Index (GDAXI, VDAX)[16], consisting of the top 30 German stocks on the Frankfurt Stock Exchange, the Standard & Poor's 500 (GSPC, VIX)[17], including 500 leading companies of the US economy and the CurrencyShares Euro Trust(FXE, EVZ)[18], designed to track the $US/Euro exchange rate. Using three different indices gives more observations to work with and lowers the risk of sample/index-specific results. This also facilitates generalizable inferences. Including the non-stock index allows me to examine whether my results are applicable to volatility forecasting in different sorts of series or if they are restricted to stock indices. The volatility indices are calculated by (Volatility Index Ticker) Chicago Board of Options Exchange[19] (VIX & EVZ) and Deutsche Börse (VDAX). All data is available for free at the internet; VIX and EVZ are obtained from CBOE and DAX and VDAX from Yahoo Finance[20].

I collect daily data as far back as is possible, where the limit in all cases is imposed by the (non-) availability of the respective volatility index. The indices do not cover the same period and it is important to bear in mind when analyzing them that the purpose is not a comparison of the indices themselves at a certain point in time. They serve to test statistical techniques in different empirical settings, independent of each other and each others properties. In addition to the dataset used for forecasting I collect 800 pre-sample observations for indices to facilitate calibration of the models. To allow for analysis of the performance differences between bull and bear market conditions I divide the dataset into two sub-sets in much of the work below by applying the algorithm described in the theory section. Once the bull and bear market periods are found in each index they are converted to continuous compound returns in order to analyze the variance of these returns over time. It should also be pointed out that the bull and bear market algorithm is only available *ex-post* since an extreme point can be identified at the earliest 150 days after it occurred. As such I apply the algorithm only as a way of looking at if forecasts *would* have benefited from being updated when switching market condition. The algorithm does not offer a way of implementing such a switch in real time.

It should be noted that there are some weaknesses associated with the dataset. Since the

[15] This work originally included six indices but I leave three out for brevity. Results where similar over the other three indices, indicating that the results are generalizable, while focusing on fewer indices allows a more thorough analysis. The left out indices where STOXX Euro 50, Nikkei 225 and U.S Oil Fund.

[16] http://www.boerse-frankfurt.de/en/equities/dax+DE0008469008

[17] http://www.standardandpoors.com/indices/sp-500/en/us/?indexId=spusa-500-usduf–p-us-l–

[18] http://www.currencyshares.com/products/overview.rails?symbol=FXE

[19] http://www.cboe.com/default.aspx

[20] http://finance.yahoo.com/

indices only report values on trading days there are gaps in the observations every weekend and holiday. The biggest gap in each series, where either the underlying index or the volatility index does not have a reported value, are between 4 days for the $US/Euro index and 7 days for S&P 500. Moreover, I use daily observations since this is what I have access to for all indices although using intra-day observations could bring the empirically observed process closer a continuous process and improve discrete approximations of continuous phenomena, such as the integrated variance.

I start by inspecting the pre-sample data to get a view of what types of processes it is reasonable to believe the indices to follow and how to best fit the models to these processes. The moving average and EWMA models are initiated on the pre-sample data directly without tweaking since these models are employed in (almost) the same way regardless of the underlying process. I plot the autocorrelation function for all return series and all squared return series. The plot of the normal returns is used to see if there are indications of a dependence structure in the mean over time, which has implications for how to model the mean in the GARCH specifications, and the squared returns autocorrelation plot to see if the variance shows any clear dependence over time, which has implications for how to model the variance in the GARCH specifications. I strengthen the inferences from the squared return autocorrelation plots by computing Engle's Lagrange Multiplier test for conditional heteroscedasticity. Engle's test is obtained by simply regressing past values of the squared error term on itself, i.e. considering equation $a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_p a_{t-p}^2$, where a_t is the residual obtained by accounting for the mean in return series. $T \times R^2$ from this regression is according to Engle (1982) is χ^2- distributed with p degrees of freedom under the null hypothesis that $\alpha_i = 0$ for $i = 1, .., p$. The autocorrelation plot together with Engle's test gives information on with how many parameters the GARCH should be implemented.

The GARCH parameters are estimated by maximum likelihood and thus I need to assume a distribution for the error term. I plot histograms and compute the Jarque-Bera test Jarque and Bera (1987) to test the data for normality. The null hypothesis is that the data has a skewness of 0 and a kurtosis of 3, which is what a characterizes the Gaussian distribution, and the alternative hypothesis is that the data is either skewed or of excess kurtosis. The skewness (S) and kurtosis (K) are defined as the third and fourth standardized central moment of a distribution respectively. For a random variable X with mean μ and variance σ^2 this is $S = \mathbb{E}\left[\left(\left(X-\mu\right)/\sigma\right)^3\right]$ and $K = \mathbb{E}\left[\left(\left(X-\mu\right)/\sigma\right)^4\right]$. An unbiased estimator of the skewness, S, is given by

$$\hat{S} = S \times \frac{\sqrt{T^2 - T}}{T - 2},$$

where the second term is the bias correcting factor so that S is the sample skewness, or the biased skewness estimator, given by

$$S = \frac{\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \bar{X}\right)^3}{\left(\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \bar{X}\right)^2\right)^{3/2}}.$$

An unbiased estimator of the kurtosis, K, is given by

$$\hat{K} = \frac{(T-1)\left[(T+1)K - 3(T-1)\right]}{(T-2)(T-3)} + 3,$$

where K is the sample kurtosis given by

$$K = \frac{\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \bar{X}\right)^4}{\left(\frac{1}{T}\sum_{i=1}^{T}\left(X_i - \bar{X}\right)^2\right)^{2}},$$

and where $\bar{X} = (1/T)\sum_{i=1}^{T} X_i$ is the mean of the variable X and T is the number of observations in both formulas.

Jarque and Bera (1987) shows that the statistic given by

$$\text{J-B} = \frac{T}{6}\left[S^2 + 0.25\,(K-3)^2\right]$$

is asymptotically χ^2-distributed with 2 degrees of freedom under the null hypothesis and I use this to examine the dataset for excess kurtosis and skewness commonly found in stock returns. Since it is known that the convergence to a χ^2-distribution is slow for the J-B statistic, numerically generated p-value tables specifically for this purpose are available.

The GARCH specifications are fit to the pre-sample data after inspecting the above statistics. To arrive at the correct specification for the mean, so as to remove the linear dependence described in relation to Equation (1), I use the Bayesian Information Criterion (BIC). This criterion is defined as

$$BIC = -2\ln(\text{likelihood}) + \ln(T) \times k$$

where T is the number of observations and k the number of estimated parameters, that is, $R+M$ in the ARMA model. The information criterion measures the fit of a model in relation to its complexity. A good fit, indicated by a high likelihood, is rewarded while extra parameters are "punished" so as to avoid overfitting and selection of as small a model as possible, still with good explanatory power. Brooks (2008) states that the BIC has an advantage over many other available information criteria in that it is consistent and asymptotically gives the "correct" model. Its drawback is that it is inefficient compared to for example the Akaike Information Criterion, but the sample is relatively large so I prefer the consistency.

To decide on how many ARCH- and GARCH-effects to include in the forecasting, that is to decide i and j in Equations (4) & (5), I first note that the general GARCH-specification is versatile in that it allows for influence of all previous ARCH-effects through the GARCH-effect. Also, in almost all herein referenced papers the employed specifications are the simple GARCH (1,1) and extensions of it. I therefore also apply the specifications with $i = 1$ and $j = 1$. The validity of this is controlled by examining the standardized residuals of the GARCH(1,1) specification. The standardized residuals are examined to make sure that the conditional heteroscedasticity has been accounted for. Ideally, the standardized residual should show no ARCH effects, again analyzed with Engle's test. Furthermore, Kuiper's test (Kuiper, 1962) is computed to examine the goodness of fit for the assumed distribution of the error term in the GARCH models. The Kuiper test statistic, for discrete observations, is defined as

$$v = D^+ + D^- = \max\left(\frac{i}{T} - F(x_i)\right) + \max\left(F(x_i) - \frac{i-1}{T}\right),$$

where x_i are the observed values (assumed independent realizations) of a random variable X, $F(\cdot)$ the theoretical cumulative distribution function we want to examine the fit of and T is the number of observations. If observations are ordered from smaller to larger such that $x_1 \leq x_2 \leq ..x_i.. \leq x_T$, i/T is an unbiased estimate of the cumulative distribution function, denote $H(x) = \mathbb{P}[X \leq x]$, from which the sample was drawn. This is clear since i/T is the fraction of observations smaller than observation x_i. It follows that D^+ and D^- are the maximum distances between the empirical distribution of the observed values and an assumed theoretical distribution function, on the upside and downside respectively (See Figure 7 for estimates of $H(x)$ plotted versus different theoretical distributions). Kuiper shows that the p-value of this statistic is given by

$$\mathbb{P}\{V > v\} = 2\sum_{j=1}^{\infty}\left(4j^2\xi^2 - 1\right) exp\left(-2j\xi^2\right),$$

where $\xi = v\left(\sqrt{T} + 0.155 + 0.24T^{-0.5}\right)$, v is the observed statistic and T the number of observations. The statistic is a test of how "close" the assumed and empirical distribution functions are and therefore a lower value is better. We reject the null that the empirical observations $\{x_i\}_1^T$ are drawn from the theoretical distribution $F(\cdot)$ if the p-value is below the desired significance level. In addition to using the Kuiper statistic for model selection I also compute it over time and use it as a diagnostic tool, so as to examine if differences in predictive accuracy can be (in part) explained by a difference in fit of the assumed distributions over time.

All forecasting is conducted out-of-sample and the models are allowed to make use of the most recent information when estimating the 30-day ahead variance by re-fitting the parameters every day as the forecast moves forward through the sample. For the moving average model this means a rolling window of 30 or 60 squared returns, for the EWMA it means adding a recent observation without taking out the oldest one, and for the implied volatility it has no effect at all since these forecasts are already computed using the latest information for the relevant period. For the GARCH models it means adding an observations to the pre-sample period on which the parameters are estimated. For the $US/Euro index, MatLab's numerical maximization algorithm[21] was not able to find a solution for the first period when using this procedure. To get around this I use a rolling window of 800 observations, instead of just adding the most recent to increase the sample.

I compare the 30-day ahead variance predictions with the realized variance approximation of integrated variance described in Section 2. The differences between the estimated variances and the realized variance are the prediction errors and I compute the three loss functions, described in section 2, for all six predictions in all three return series. Relative performance of the techniques is compared in bull and bear markets by computing the loss functions for each model's predictions in the two respective market conditions. This gives an overview of each model's performance in the two market conditions and of how the performance differs between the market states. I compare all forecasts head to head with the test proposed by Diebold and Mariano (1995), which is a test for statistical significance of difference in prediction accuracy. The test statistic is obtained as follows:

Let e_{t1} and e_{t2} be the forecast errors from two models' predictions for period t, that is $e_{t1} = \sigma_t^2 - \hat{\sigma}_{t1}^2$ and $e_{t2} = \sigma_t^2 - \hat{\sigma}_{t2}^2$, where σ_t^2 is the (approximated) true value in period t and $\hat{\sigma}_{ti}^2$ is model i:s prediction in period t. Then take a loss function $\ell\,(e_{ti})$ and denote the loss differential d_t=$\ell\,(e_{t1})$-$\ell\,(e_{t2})$. The loss differential is assumed to have short memory and to be covariance stationary, that is, the first and second moments of the distribution for d_t are the same for all t. The D-M statistic is given by D-M = $\bar{d} \times T^{0.5} \times LRV^{-0.5}$, where $\bar{d} = (1/T)\sum_{t=1}^{T} d_t$ and LRV is the estimated long-run variance in $\sqrt{T}\bar{d}$. Since my daily forecasts are 30-day ahead, errors are theoretically autocorrelated for 30 lags. I estimate the long run variance using the method of Newey and West (1987) in order to correct for this autocorrelation (and possible heteroscedasticity). Diebold and Mariano show that under the

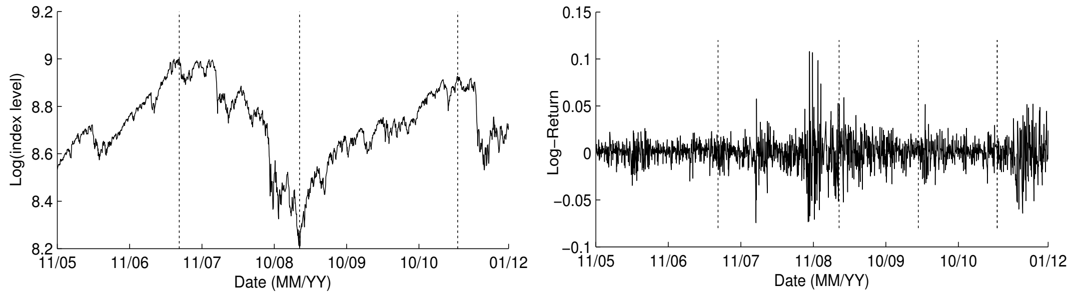
[21] I employ the MatLab Optimization Toolbox's function 'fmincon' with option 'active-set'. See http://www.mathworks.se/help/toolbox/optim/ug/fmincon.html for details on this function. I also experimented with other options for this function but this did not solve the problem

null hypothesis $\mathbb{E}[d_t] = 0$ we have approximately D-M $\sim N(0,1)$ and I use this to test the null against the alternative that $\mathbb{E}[d_t] \neq 0$. This is done for both the MSE and the QLIKE loss function.

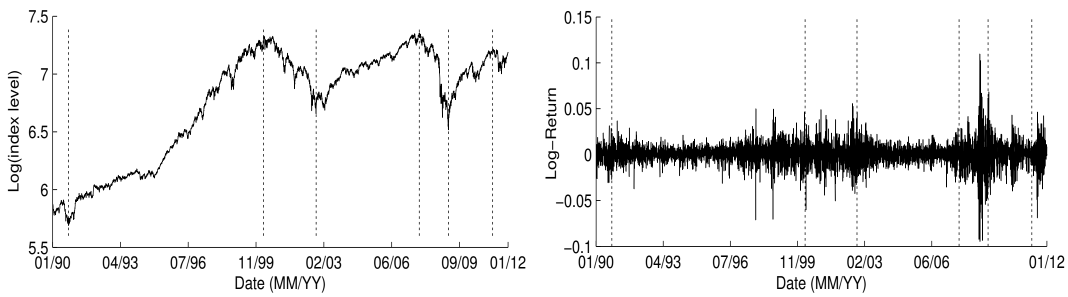
I contrast the bull and bear market results with general results obtained without splitting the dataset into different market conditions. Except from the split, general results are produced in the exact same way as the bull and bear market results.

4 Data and Model Fitting

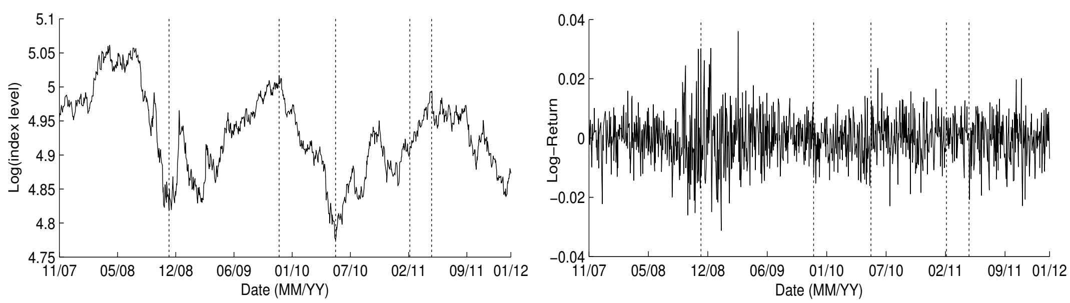
The development of the log-indices over time are shown in the upper graphs in Figure 3 where the dashed lines indicate the turning points found by the bull and bear market algorithm.



(a) DAX



(b) Standard & Poor's 500



(c) CurrencyShares Euro Trust

Figure 3: Log-index and log-return development over time

Notes: The left hand graphs display the development of the log of the indices over the whole observation period. The right hand graphs show the continuous compound returns. The dashed lines mark the turning points of bull & bear markets.

In the development figures we see that all markets exhibit periods of positive and negative trends. Due to the difference in data availability the observation period differs for each index; the $US/Euro observation period is the shortest with around four years from November 2007 to January 2012 while the S&P 500 observation period is longest with its 22 years, covering January 1990 to January 2012. The first difference of the indices are plotted in the lower graphs of Figure 3. Because of the logarithmic form this constitutes the continuously compounded returns. From the first differences it is clear that the stock market returns exhibit volatility clustering. In the $US/Euro returns bursts in volatility occur more randomly and the dependence structure is less clear than in the stock returns. All indices display high volatility around years of financial turmoil, as seen around 2008 and also around 2000 for the S&P 500. Volatility is also higher in bear markets than in bull markets for for the stock indices while the $US/Euro shows a more unclear pattern. All of the return-series appear to have a constant mean around zero throughout their respective observation period. Descriptive statistics are presented in Table 1.

	DAX	S&P 500	C.S Euro Trust
Daily Observations	1555	5559	1060
Mean $\left(10^{-4}\right)$	1.2	2.3	-0.8
Standard Deviation	0.016	0.012	0.0079
Max Value	0.18	0.11	0.036
Min Value	-0.074	-0.095	-0.031
Skewness	0.077	-0.23	0.017
Excess Kurtosis	6.0	8.4	1.3
Vol. of Vol.	0.0007	0.0005	0.0001
Jarque-Bera p-value	<0.001	<0.001	<0.001
Engle's Test p-value	<0.001	<0.001	0.0065

Table 1: Descriptive Statistics

Notes: This table shows descriptive statistics for all continuous compound return series of the analyzed indices. Excess kurtosis is defined relative to a Gaussian distribution's kurtosis of 3. Volatility of volatility is defined as the estimated standard deviation of the series of realized volatility observations. The Jarque-Bera test's null hypothesis is that there is no skewness or excess kurtosis in the data and Engle's test's null hypothesis is that there are no ARCH-effects.

From Table 1 we see that the number of observations ranges from a low 1060 for the $US/Euro index to a high 5559 for the S&P 500 index. The mean log-returns are indeed all close to zero, S&P 500 the highest around 0.0002 and the $US/Euro index the lowest around -0.00008. The dispersion-related statistics indicate that the DAX returns are on average the most volatile with the highest estimated standard deviation of 0.016 and the highest maximum absolute return of 18 %. The exchange rate returns have the lowest standard deviation of 0.0079 and also the lowest maximum absolute return of 3.6 %. The distribution of the returns are plotted as histograms in Figure 4.

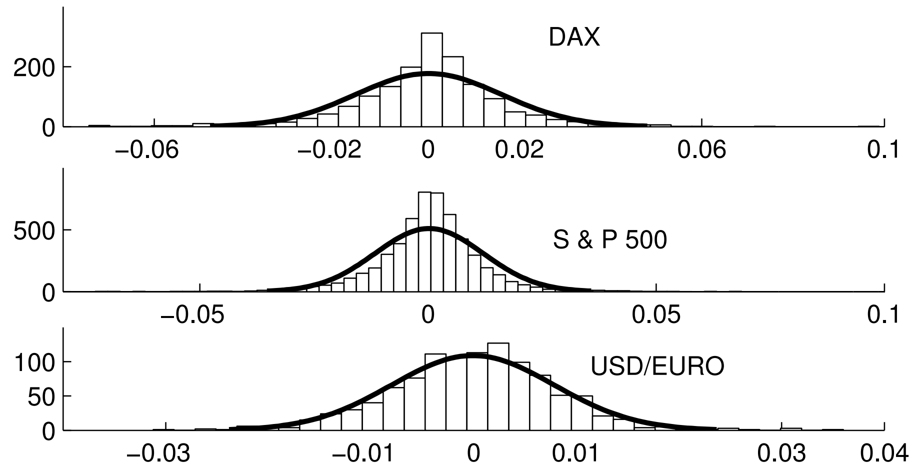


Figure 4: Histograms for the return distributions of all indices
Notes: The thick line represents a superimposed normal distribution

The histograms in Figure 4 show that the stock returns are bell-shaped and *leptokurtically* distributed. There is indication of a peak also in the returns of the $US/Euro index but it is not as "high" and this series merits further testing before normality is rejected. From the values on excess kurtosis in Table 1 we see that all distributions have fatter tails than the normal distribution and a Jarque-Bera test of normality rejects the null hypothesis of normally distributed data at the 0.1% level of significance for all indices, including the $US/Euro.

Table 2 shows selected descriptive statistics for the return series when using the split into bull and bear markets. We see that the exchange rate index has spent more time in bear markets than in bull markets while the reverse is true for the stock indices. The difference between the number of days spent in the different market states is biggest in the S&P 500 index, where the bull observations, 4190, are more than three times as many as the bear observations, 1375. The standard deviation of returns is higher in bear markets for the stock indices but not for the $US/Euro. The increased variability in bear market returns is also reflected in the kurtosis, which is higher in bear markets, again except for the exchange rate index. The sign of the skewness varies among market conditions and indices but there is a clear pattern that the magnitude of the skewness increases in bear markets for all indices. It increases with a factor from around 2 at the lowest for the $US/Euro to a factor of around 6 for the S&P 500. All scenarios reject the null hypothesis of no skewness and excess kurtosis of the Jarque-Bera test on the 1% critical level and below.

	DAX		S&P 500		$US/Euro	
	Bull	Bear	Bull	Bear	Bull	Bear
Obs	966	589	4190	1375	490	570
Mean	0.0012	-0.0017	0.0007	-0.0013	0.0008	-0.0009
Std. Dev.	0.012	0.021	0.0094	0.017	0.0082	0.0076
Skewness	-0.084	0.29	-0.025	-0.15	0.13	0.28
Vol. of Vol.	0.0003	0.0011	0.0002	0.0008	1.6	0.73
Excess Kurtosis	2.1	4.1	4.7	5.2	0.0001	0.0001
Jarque-Bera p-value	<0.001	<0.001	<0.001	<0.001	<0.001	0.0034

Table 2: Descriptive Statistics Bull and Bear Markets

Notes: This table shows descriptive statistics for all continuous compound return series of the analyzed indices when dividing the data into bull and bear markets. Excess kurtosis is defined relative to a Gaussian distribution's kurtosis of 3. Volatility of volatility is defined as the estimated standard deviation of the series of realized volatility observations. The Jarque-Bera test's null hypothesis is that there is no skewness or excess kurtosis in the data.

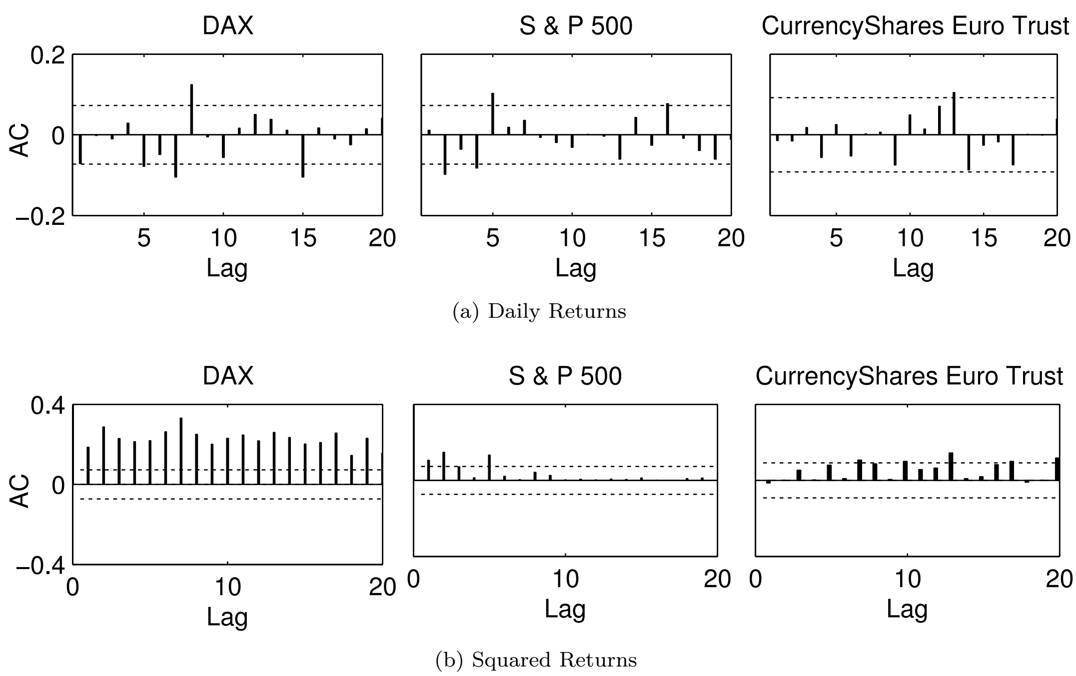


Figure 5: Autocorrelation Function for the Mean Daily Returns

Note: For a times series, the autocorrelation/ serial correlation is the correlation between observations and their corresponding lags. The dotted lines mark 95% confidence bands.

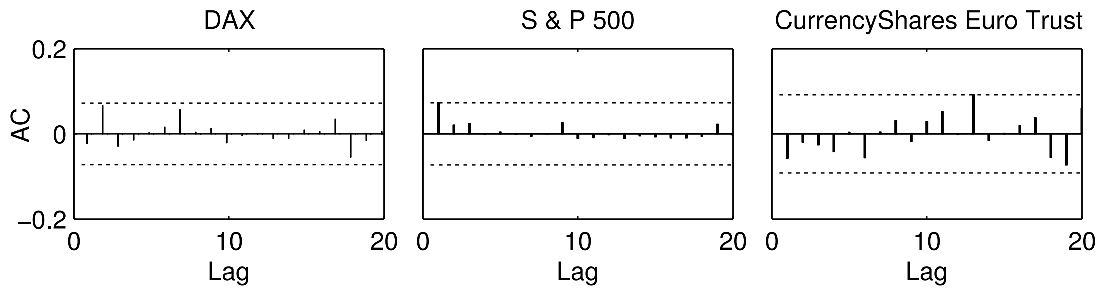


Figure 6: Autocorrelation Function for the Squared Standardized Residuals.
Notes: The standardized residuals are computed as $(\xi_t/\hat{\sigma_t})^2$ *where* ξ_t *is an uncorrelated but conditionally heteroscedastic error term and* $\hat{\sigma}_t$ *the estimated conditional standard deviation in period* t. *If* $(\xi_t/\hat{\sigma_t})^2$ *is uncorrelated over time the employed volatility equation has modeled the correlation in the variance over time in a sufficient way., i.e. the ARCH or GARCH specification is satisfactory. The dotted lines mark 95% confidence bands.*

Looking at the pre-sample data I find in Figure 5a that the return series exhibit no significant serial correlation in the mean for the first lags. There are however some significant correlations in higher lag orders, indicating that some dependence not explained by the a constant mean or low order ARMA specification exists. I compute the BIC for ARMA(R,M) models of orders (0,0) to (10,10)[22] and a constant mean gives the lowest value in all cases. So in order to satisfy Equation (1), based on the pre-sample data, I let the equation for the mean return be simply a constant in all series. Figure 5b shows that the squared returns have significant serial correlation in the stock indices at the 95% confidence level. Again the result for the exchange rate index is not as clear. There is close to no significant autocorrelation in the squared returns of this series except one at the 13th lag.

	DAX	S&P500	$US/Euro
Jarque-Bera	<0.001	<0.001	0.25
Engle's Test	<0.001	0.0040	0.68

Table 3: Test Statistics for Pre-sample Data.
Notes: This table shows statistics for tests on the pre-sample data only. The Jarque-Bera test's null hypothesis is that there is no skewness or excess kurtosis in the data and Engle's test's null hypothesis is that there are no ARCH-effects.

Engle's test presented in Table 3 accordingly indicates that the the $US/Euro do not exhibit significant ARCH effects with p-value 0.68 while the other indices show significant ARCH-effects at even the 0.1% critical level. As for the shape of the distribution, the Jarque-Bera test reported in Table 3 rejects the null hypothesis of no skewness and no excess kurtosis

[22] I leave the tables with all BIC results out for space considerations and they are available upon request

with $p < 0.001$ for the stock indices, but fails to reject the null for the exchange rate index with p-value 0.25. This suggests the pre-sample returns are potentially normally distributed in the exchange rate index. The autocorrelation plots in Figure 6 show the autocorrelation in the standardized residuals after fitting a GARCH(1,1) model to the pre-sample data. The plots show that the squared standardized residuals are uncorrelated over time, indicating that the dependence structure in the volatility is explained by the employed GARCH(1,1) model. Since the non-stock index neither exhibits significant conditional heteroscedasticity, nor indicates non-normality in the pre-sample data, one could argue that GARCH-modeling with assumed Student's/Hansen's t distributed errors is unmerited. However, since the purpose is to compare forecast techniques in different scenarios I still apply the GARCH model with these distributions. Moreover, the employed algorithm estimates the degree of freedom ν for the t-distribution and we have that for large ν, the t-distribution goes to a normal. So if the true distribution is normal, the degree of freedom will be estimated to fit this. The same is true for Hansen's t since it nests the Student's t and the skewness λ can be estimated to fit a symmetric distribution.[23]

[23] I here ignore the fact that estimating extra parameters is computationally inefficient. In a smaller sample with fewer degrees of freedom or in cases where the speed of the computation is an issue, one could impose the normally distributed returns for the $US/Euro index.

5 Results and Analysis

5.1 Bull and Bear Market Results

Forecasting results are presented in 4-6 for bull and bear markets, and further down in Table 7 for the aggregate.

DAX	OLS R^2		QLIKE		MSE	
	Bull	Bear	Bull $(\times -1)$	Bear $(\times -1)$	Bull $(\times 10^{-5})$	Bear $(\times 10^{-3})$
MA Vol. 30	0.47	0.23	4.50	3.21^{2}	0.92	0.21
			(0.028)	(0.072)	(0.063)	(0.022)
MA Vol. 60	0.45	0.10	4.55	3.07	0.87	0.24
			(0.024)	(0.082)	(0.057)	(0.024)
EWMA	0.47	0.26	4.52	$3.23^{2,6}$	0.80^{1}	0.18
			(0.027)	(0.073)	(0.058)	(0.021)
GARCH	0.47	0.33	$4.58^{1,2,3,6}$	$3.31^{1,2,3,6}$	$0.62^{1,5}$	0.15
			(0.023)	(0.071)	(0.043)	(0.019)
GARCH	0.47	0.33	$4.59^{1,2,3,6}$	$3.31^{1,2,3,6}$	0.63^{1}	0.15
(Hansen's t)			(0.023)	(0.070)	(0.045)	(0.019)
EGARCH	0.32	0.38	4.56	2.93	0.56	0.14
			(0.026)	(0.088)	(0.034)	(0.019)
Implied Vol.	0.53	0.34	$4.59^{1,2,3,6}$	3.13^{6}	$0.36^{1,2,3,4,5,6}$	0.16
			(0.026)	(0.086)	(0.030)	(0.021)

Table 4: Performance of 30-day Variance Forecasts in Bull and Bear Markets in the DAX Index .

Notes: Standard errors are in parenthesis. GARCH models are (1,1) specifications and fitted with Student's t errors unless stated otherwise. Superscript, i*, denotes better performance than model* i *according to the Diebold Mariano test at the 95% confidence level. Models are labeled 1=MA Vol., 2=MA Vol. 60,...,7=Implied Vol. OLS R^2 is the R^2 when regressing the realized variance on the estimated variance. The QLIKE and MSE (Mean Squared Error) are defined in Section 2.*

Table 4 shows that the implied volatility forecasts are the most accurate in bull markets for the DAX index. The forecast based on the VDAX earns the highest R^2 with 0.53 and also, with statistical significance, outperforms all the other models according to the MSE and four of the other models according to the QLIKE loss function. In the bull markets, the GARCH(1,1) specification, EWMA and Moving Average 30 perform equivalently according to the R^2 but the GARCHs stand out with more statistically significant wins. Notably, the EGARCH is the worst model in bull markets with an R^2 of 0.32 and although it is not worst by any other statistic, neither does it outperform any other model with statistical significance. The results are more mixed in bear markets. The EGARCH has the highest R^2 but is still unable to outperform the other models in terms of the MSE and QLIKE loss functions. Specifically, it shows the highest (worst) QLIKE value of all models. Instead, the GARCH specifications perform best on average in bear markets. They have the third highest R^2, the lowest QLIKE and second lowest MSE. While there is never a significant difference between them, the GARCH with assumed Student's t distribution outperforms one more model than its equivalent with assumed Hansen's t errors in the bull market scores. It is clear that all models except maybe the EGARCH perform worse in bear markets than in bull markets.

S&P 500	OLS R^2		QLIKE		MSE	
	Bull	Bear	Bull $(-1\times)$	Bear $(-1\times)$	Bull $(\times 10^{-5})$	Bear $(\times 10^{-3})$
MA Vol. 30	0.38	0.36	5.10	3.78	0.66	0.12
			(0.016)	(0.038)	(0.037)	(0.012)
MA Vol. 60	0.35	0.22	5.13	3.74	0.65	0.15
			(0.014)	(0.043)	(0.033)	(0.015)
EWMA	0.40	0.40	5.12	3.79	0.58	0.11
			(0.015)	(0.039)	(0.032)	(0.012)
GARCH	0.40	0.42	5.14	$3.83^{1,2,3}$	$0.48^{1,2,3}$	0.10
			(0.013)	(0.037)	(0.027)	(0.011)
GARCH	0.40	0.43	5.14	$3.83^{1,2,3}$	$0.47^{1,2,3,4}$	0.09
(Hansen's t)			(0.013)	(0.037)	(0.027)	(0.011)
EGARCH	0.44	0.47	$5.28^{1,2,3,4,5,7}$	$3.88^{2,7}$	$0.33^{1,2,3,4,5}$	0.09
			(0.011)	(0.045)	(0.021)	(0.012)
Implied Vol.	0.49	0.41	$5.16^{1,3}$	3.81	$0.39^{1,2}$	0.09
			(0.012)	(0.043)	(0.018)	(0.012)

Table 5: Performance of 30-day Variance Forecasts in Bull and Bear Markets in the Standard & Poor's 500 Index.

Notes: Standard errors are in parenthesis. GARCH models are (1,1) specifications and fitted with Student's t errors unless stated otherwise. Superscript, i*, denotes better performance than model* i *according to the Diebold Mariano test at the 95% confidence level. Models are labeled 1=MA Vol., 2=MA Vol. 60,...,7=Implied Vol. OLS R^2 is the R^2 when regressing the realized variance on the estimated variance. The QLIKE and MSE (Mean Squared Error) are defined in Section 2.*

The Standard and Poor's 500 index results differ on some points from those of the DAX index. Table 5 shows that the implied volatilities are still good predictions with the highest R^2 in bull markets but now the EGARCH stands out as the best performer in bull markets. It outperforms all other models by the QLIKE measure and all but the implied volatilities according to the MSE. In addition, it has the highest R^2 in bear markets and also the lowest QLIKE and MSE (shared with GARCH using Hansen's t errors and the implied volatilities). The difference in predictive accuracy between the two market states is smaller in the S&P 500 than in the DAX. This is indicated by the fact that only three out of seven models have a lower R^2 in bear markets. On the other hand, the differences in QLIKE and MSE are of approximately the same magnitude as in the DAX index.

\$US/Euro	OLS R^2		QLIKE		MSE	
	Bull	Bear	Bull ($\times -1$)	Bear ($\times -1$)	Bull ($10^{-5}\times$)	Bear ($10^{-5}\times$)
MA Vol. 30	0.73	0.59	5.38^{6}	5.35	0.09	0.07
			(0.023)	(0.025)	(0.009)	(0.006)
MA Vol. 60	0.70	0.48	5.37^{6}	5.33	0.10	0.11
			(0.023)	(0.029)	(0.008)	(0.010)
EWMA	0.68	0.58	5.38^{6}	5.35	0.08	0.08
			(0.023)	(0.026)	(0.009)	(0.006)
GARCH	0.71	0.56	$5.38^{2,6}$	5.34	0.08	0.08
			(0.023)	(0.027)	(0.007)	(0.008)
GARCH	0.71	0.56	$5.38^{2,6}$	5.34	0.08^{4}	0.08
(Hansen's t)			(0.023)	(0.027)	(0.007)	(0.008)
EGARCH	0.72	0.50	5.26	5.28	0.11	0.09
			(0.027)	(0.028)	(0.001)	(0.009)
Implied Vol.	0.71	0.47	$5.41^{1,2,3,4,5,6}$	5.31	0.04^{3}	0.10
			(0.024)	(0.029)	(0.006)	(0.009)

Table 6: Performance of 30-day Variance Forecasts in Bull and Bear Markets in the Currency Shares Euro Trust (\$US/Euro)

Notes: Standard errors are in parenthesis. GARCH models are (1,1) specifications and fitted with Student's t errors unless stated otherwise. Superscript, i, *denotes better performance than model* i *according to the Diebold Mariano test at the 95% confidence level. Models are labeled 1=MA Vol., 2=MA Vol. 60,...,7=Implied Vol. OLS R^2 is the R^2 when regressing the realized variance on the estimated variance. The QLIKE and MSE (Mean Squared Error) are defined in Section 2.*

Table 6 presents forecasting results in the \$US/Euro exchange rate. Like for the stock indices the implied volatilities provide good forecasts, at least in bull markets. Although the R^2 only ranks it as the third best model the QLIKE and MSE both rank implied volatilities as number one in bull markets. In bear markets it is instead the 30 days moving average that provides the best forecasts with highest R^2, lowest QLIKE and lowest MSE. Although the results are not statistically significant it is clear that the simple models perform relatively better in this market compared to the stock markets. We also see that the forecasts are in general better in absolute values compared to the stock markets. For example, the maximum R^2 in any other scenario is 0.53 (Implied volatility in DAX bull markets) but here all models deliver R^2 around 0.7 in bull markets. Moreover, the difference between market states is smaller than in the DAX and S&P 500.

5.2 Results without bull/bear split

I contrast the bull and bear market results by computing the same statistics for the whole dataset. Table 7 displays the error statistics over all markets and observations.

	DAX Index			S&P 500			\$US/Euro		
Model	OLS R^2	QLIKE $(\times -1)$	MSE $(\times 10^{-4})$	OLS R^2	QLIKE $(\times -1)$	MSE $(\times 10^{-4})$	OLS R^2	QLIKE $(\times -1)$	MSE $(\times 10^{-5})$
MA Vol. 30	0.34	4.03^{6}	0.84	0.44	4.78	0.35	0.57	5.37^{6}	0.08
		(0.036)	(0.084)		(0.017)	(0.031)		(0.017)	(0.005)
MA Vol. 60	0.22	4.01^{6}	0.92	0.32	4.79	0.40	0.42	5.35^{6}	0.10
		(0.039)	(0.091)		(0.017)	(0.036)		(0.019)	(0.007)
EWMA	0.38	4.04^{6}	0.72	0.47	4.80	0.30^{1}	0.55	5.36^{6}	0.79
		(0.036)	(0.078)		(0.017)	(0.029)		(0.018)	(0.005)
GARCH	0.43	$4.12^{1,2,3,6}$	0.60	0.49	4.82^{1}	$0.27^{1,3}$	0.51	5.36^{6}	0.83
		(0.034)	(0.073)		(0.016)	(0.027)		(0.018)	(0.005)
GARCH	0.43	$4.12^{1,2,3,6}$	0.60	0.49	4.82^{1}	$0.27^{1,3}$	0.51	5.36^{6}	0.83
(Hansen's t)		(0.034)	(0.072)		(0.016)	(0.027)		(0.018)	(0.005)
EGARCH	0.48	3.89	0.55	0.54	$4.94^{1,2,3,4,5,7}$	0.25	0.51	5.27	0.10
		(0.041)	(0.073)		(0.016)	(0.029)		(0.020)	(0.007)
Implied Vol.	0.45	4.06^{6}	0.61	0.48	$4.83^{1,2}$	0.26	0.55	5.36^{6}	0.72
		(0.040)	(0.079)		(0.016)	(0.030)		(0.019)	(0.006)

Table 7: Performance of 30-day Variance Forecasts.

Notes: Standard errors are in parenthesis. GARCH models are (1,1) specifications and fitted with Student's t errors unless stated otherwise. Superscript, i, *denotes better performance than model* i *according to the Diebold Mariano test at the 95% confidence level. Models are labeled 1=MA Vol., 2=MA Vol. 60,...,7=Implied Vol. OLS R^2 is the R^2 when regressing the realized variance on the estimated variance. The QLIKE and MSE (Mean Squared Error) are defined in Section 2.*

From the first section in Table 7 we see that the DAX index return variance is on average best forecast by a GARCH specification. The regular GARCH(1,1), with either of the two employed distribution assumptions, is the best performer according to the QLIKE. However, the EGARCH is superior in terms of R^2 and MSE. The implied volatility also shows low error statistics but the three moving average specifications trail by all measures, although in most cases with no statistical significance.

The general S&P 500 results rank models with some difference in comparison to the DAX index. We see from the second section in Table 7 that the EGARCH performs best by all measures. Notably, its R^2 is 0.06 above the second highest and it outperforms all models with statistical significance by the QLIKE measure. The two other GARCH specifications and the implied volatility offer good alternatives, with similar performance among the three. Overall,

all loss functions indicate lower prediction errors in the S&P 500 compared to the DAX.

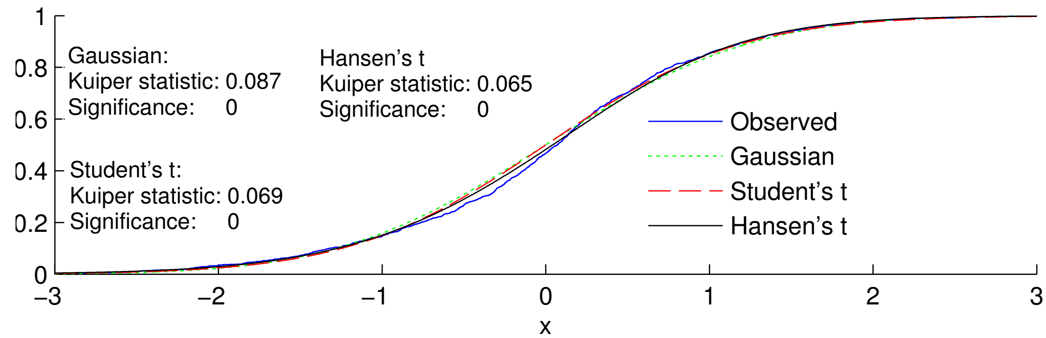
The aggregate $US/Euro results in the third section in Table 7 turns the S&P 500 ranking of models close to upside down. The EGARCH is the worst performer by both the QLIKE and MSE and only ranked 4th by the R^2. It is instead the 30 days moving average model that performs best by all measures. The EWMA and implied volatility offer good alternatives with 0.02 lower R^2 and 0.01 higher QLIKE. It is notable that the only statistically significant result is that the EGARCH is outperformed by all models in the QLIKE measure. Regarding absolute performance it is notable that overall predictions are better for $US/Euro than for both DAX and S&P 500.

GARCH	DAX	S&P 500	$US/Euro	EGARCH	DAX	S&P 500	$US/Euro
α_0	~0	~0	~0	α_0	-0.15	-0.12	-0.045
	(0.0002)	(0.0001)	(0.0001)		(0.0002)	(0.0001)	(0.0001)
α_1	0.087	0.068	0.040	α_1	0.12	0.12	0.085
	(0.011)	(0.0061)	(0.0089)		(0.020)	(0.011)	(0.017)
β_1	0.91	0.93	0.96	β_1	0.98	0.99	0.99
	(0.012)	(0.0060)	(0.0093)		(0.0030)	(0.0019)	(0.0026)
ν	9	6	17	ν	11	7	16
	(1.5)	(0.41)	(3.4)		(2.4)	(0.46)	(5.5)
λ	-0.11	-0.084	-0.083	γ	-0.13	-0.095	-0.0094
	(0.03)	(0.02)	(0.04)		(0.014)	(0.0079)	(0.011)

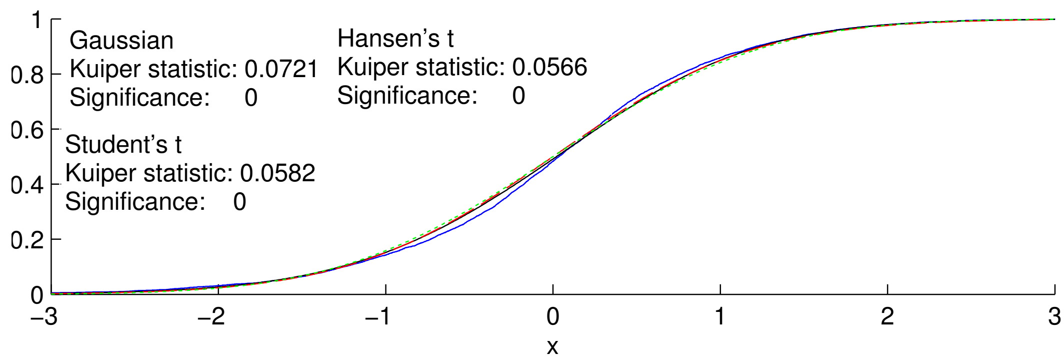
Table 8: Fitted GARCH Parameters.

Notes: The GARCH is estimated with assumed Hansen's t and the EGARCH with assumed Student's t. α_0 is a constant in the variance equation, α_1 the weight for the effect of last periods squared return, β the weight of the last periods estimated variance, ν is the degree of freedom of a t-distribution fitted to the return series in question and γ denotes the asymmetric effect on volatility as in the EGARCH model. If $\gamma < 0$, negative returns affects volatility more than positive and vice versa. λ is the asymmetry parameter for Hansen's t distribution and is < 0 if the estimated distribution is negatively skewed.

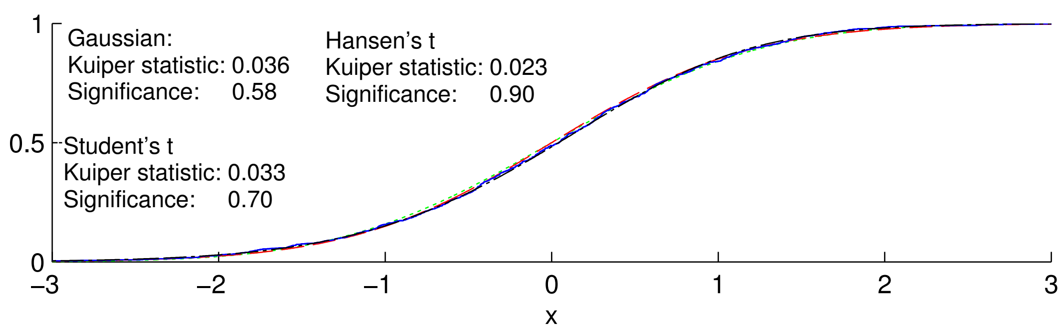
I report the estimated GARCH parameters in Table 8 and note that all indices have returns where $\alpha_1 + \beta_1$ is close to one. This indicates a strong dependence structure where the monthly variance is almost completely determined by previous values. There are negative leverage effects in all series as indicated by the EGARCH parameter γ. They are around -0.1 in all stock returns but only -0.01 in the exchange rate index. The estimated Hansen's t skewness parameter λ is also negative in all indices. The estimated degree of freedom ν for the fitted distribution is lower in the stock indices, where the lowest degree of freedom is that of the S&P 500, indicating fatter tails in the stock market returns. All of the estimated parameters agree with skewed and fat tailed distributions as indicated by the descriptive statistics and histograms in Section 4.



(a) DAX Index



(b) Standard & Poor's 500



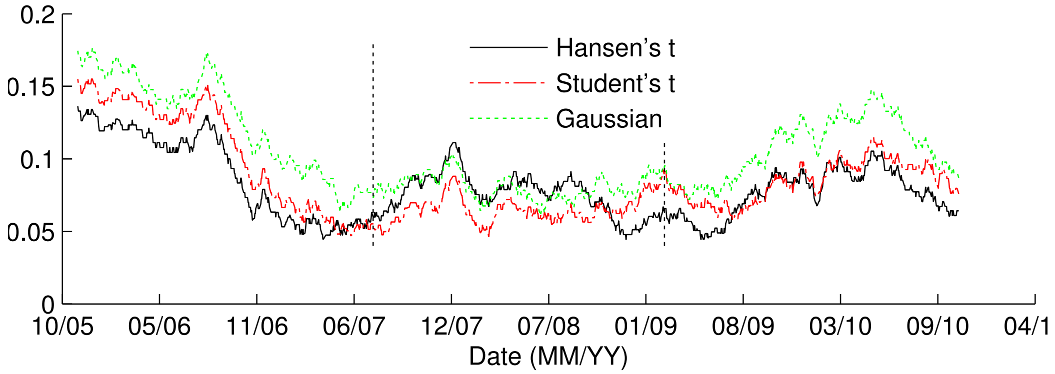
(c) CurrencyShares Euro Trust

Figure 7: Kuiper Statistics and Plots

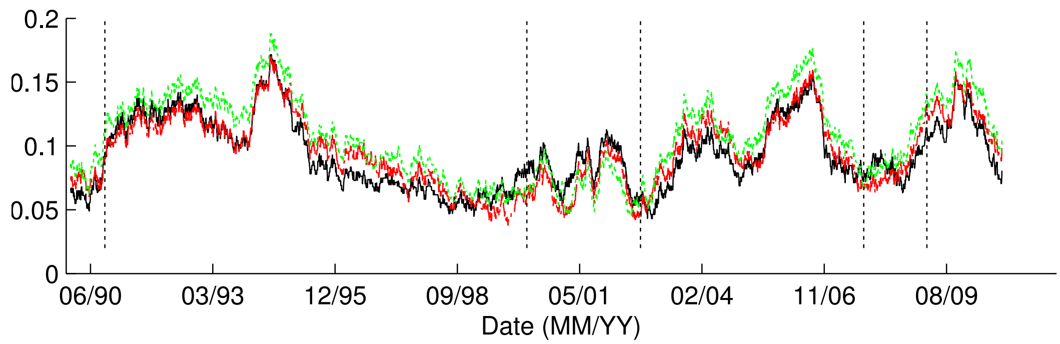
Notes: The Kuiper statistic is computed for the whole forecasting sample, using the parameters obtained when fitting the GARCH and EGARCH models (See Table 8). The statistic is the sum of the maximum distance on the upside and downside between the observed values and the theoretical distribution, graphically illustrated by the cdf plots. We reject the null hypothesis that the observed values are drawn from the theoretical distribution in question if the significance value is lower than the desired critical level. Legends are the same for all figures.

Figure 7 presents the Kuiper statistic for the employed distributions, computed on the full sample. Note that we should always have that the Kuiper statistic is ordered as *Hansen's t* $\leq$ *Student's t* $\leq$ *Gaussian* since the more versatile distributions nest the simpler. Hansen's t indeed shows the lowest Kuiper statistic for all three indices, followed by Student's t. However, in no index do we at the 1% critical level fail to reject the null that the returns are Hansen's t distributed and at the same time reject that they are Student's t distributed. Indeed, the more versatile distributions seem to offer no significant advantage over the Gaussian since all distributions are rejected in the stock indices while no distribution is rejected in the $US/Euro index.

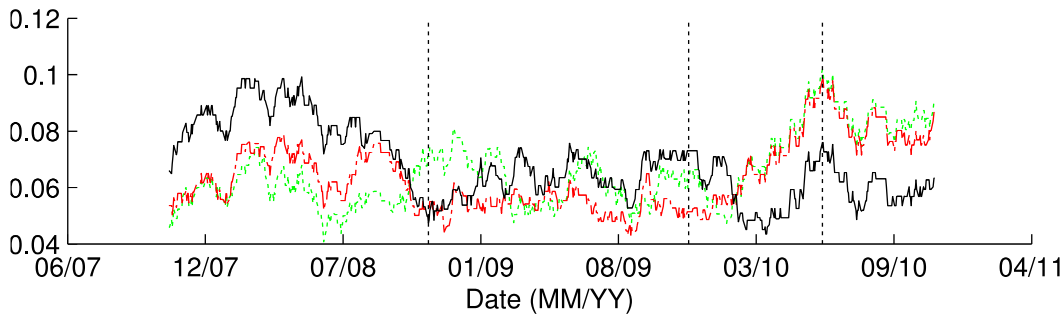
The graphs in Figure 8 show the development of the Kuiper statistic over time for a rolling window of 300 observations in each index. The statistic is not continuously updated but instead shows how the estimated distributions, as indicated by the parameters in Table 8, fit the data over time. The vertical dotted lines indicate the market turning points as before. What we see is that the DAX index is relatively stable over time in the sense that the Gaussian is worst fitting during almost all observation periods. Also, the Hansen's t fits best during most periods and the simpler Student's t is a better fit only between ,roughly, June 2007 and October 2008. Notably, the mid-sample bear market is where Hansen's t fits worse than both of the other distributions. The differences are smaller in the S&P 500 and we see that the Student's t is the best fit for many observations, especially in the beginning of the sample. Though, again, the Gaussian is the worst fit for most of the observations. Like in the DAX index Hansen's t distribution shows the worst fit during bear markets, e.g. around 1999-2001 and around 2008. In the exchange rate index we have much larger variation over time in the ranking of distributions. Hansen's t goes from being worse by a substantial marginal in the beginning to being the best with an equally big marginal in the end of the sample. We even see that the Gaussian is the best fit in some periods in the beginning of the sample and that the Student's t fits best in the middle of the sample. The bull/bear market effect on the fit is not evident in the $US/Euro returns.



(a) DAX Index



(b) Standard & Poor's 500



(c) CurrencyShares Euro Trust

Figure 8: Kuiper Statistic over Time

Notes: The Kuiper statistic is the sum of the maximum distance between the observed values' and the theoretical distributions' cdf plots on the upside and on the downside . The figures show this distance computed for a rolling window of 300 observations, always using the theoretical distributions given by the parameters obtained when fitting the GARCH and EGARCH models to the full forecasting sample (see Table 8).

5.3 Analysis

In this section I discuss results and attempt to answer the questions posed in the introduction. The pattern throughout the results is that the variance forecasting is more accurate in bull markets than in bear markets. This is true for all models although some are more affected than others, and results also vary among indices. The prediction results are similar in the two stock indices while the exchange rate index displays notable differences.

In the stock indices my findings by and large agree with those of Chiang and Huang (2011) who also find that predictions are on average more accurate in bull markets. The EGARCH model is an exception and performs better in DAX bear markets (and also best of all models in this scenario) than in DAX bull markets, in terms of R^2 and MSE. Otherwise, the DAX index is informationally efficient in bull market in the sense that the implied volatilities give the best forecasts. The bear markets are less efficient and GARCH models outperform the market's implied volatility. Notable deviations from the overall pattern are that the moving average models perform well in DAX bull markets and that the EGARCH performs much worse by the QLIKE measure than by the MSE and R^2.

The performance of the simple moving average together with the standard deviation of realized variance presented in the descriptive statistics lead me to conclude that volatility is more stable in the DAX bull markets than in bear markets and it is logical that these points benefit the moving average relative to the more advanced models. It is harder to find a convincing explanation for the EGARCH's sub-par performance in the QLIKE measure in DAX bear states. What differentiates the QLIKE measure from the other loss functions is that it is based on the relative forecast error and that it punishes negative errors harder than positive. A possible explanation for the EGARCH performance is then that the EGARCH forecasts are downward biased, which they are often known to be[24], and thus punished harder in the QLIKE measure than in the other measures.

The overall worse performance of all models in bear markets can be explained in multiple ways. Wilhelmsson (2006) mentions higher skewness and/or kurtosis as reasons for worsened performance of GARCH models in general. Other possible explanations include higher volatility of volatility (Poon and Granger, 2003) and higher levels of volatility (Diebold et al., 1997). In fact, there are signs of all these in the DAX bear market when compared to the DAX bull market. We see this in the return plots (higher level of volatility) and the descriptive

[24]The downward bias of the EGARCH is mentioned by practitioners in several sources but I have found no scientific paper examining this. An explanation is given in for example the MatLab Econometrics toolbox- http://www.mathworks.se/help/toolbox/econ/. Andersen et al. (2005) discusses the EGARCH bias in general.

statistics (higher kurtosis, skewness and volatility of volatility as well as level of volatility). Furthermore, Kirchler (2009), in an experimental study, finds markets to be more informationally efficient in bull states than in bear states and this agrees with my finding that the implied volatilities deliver, in both relative and absolute terms, better forecasts in bull markets.

Results in the Standard & Poor's 500 are similar to those of the DAX index in that the implied volatilities and the GARCH models outperform the simpler average models and in that bear market predictions are worse than bull market predictions. However, some notable differences compared to the DAX do exist. The S&P 500 is less informationally efficient in that the EGARCH model stands out as the best performer followed by the regular GARCH, both in bull and bear markets, rather than the implied volatilities. The R^2 in bull markets are an exception to this rule. It is also notable that the GARCH with Hansen's t deliver one more statistically significant "win" in this market compared to the Student's t GARCH. When looking for an explanation in the descriptive statistics we see that the skewness is negative in both market states in the S&P 500 in contrast to the DAX index. This stable skewness over time is a likely explanation for the better performance of Hansen's t. It could also explain the EGARCH performance through a more stable leverage effect. The difference in excess kurtosis between bull and bear states is also smaller than in the DAX index, further facilitating the GARCH performance through better parameter estimates.

Moreover, the EGARCH stands out in the S&P 500 by handling bear markets best, in relative terms but also on in absolute terms according to some measures. Figlewski and Wang (2000) offer an explanation when they argue the leverage effect to be a "market down effect". Bear markets are per definition going more down than up and thus models capturing the leverage effect are relatively more viable than in bull markets. This indicates that the leverage effect is important to account for in stock market returns but also that the regular GARCH is in general a better choice when the effect is 'small' , or in practice when we don't know it to be significant and stable over time. Arguably, when comparing EGARCH and GARCH forecasts, the leverage effect needs to be significant, not only statistically but also in economical meaning, in order to outweigh the EGARCH's bias in multiple step ahead forecasts.

The inferences from the bull and bear split in the stock indices are strengthened by looking at the overall results without split. The GARCH performs well in both the DAX and the S&P 500, as do the EGARCH with the exception of the QLIKE measure in the DAX index. The implied volatilities perform well in both stock indices although not clearly best in either of them. Altogether, three notable observations from the results in stock indices without bull and bear market splits are that the implied volatilities are best in neither index even though it emerged as the best bull market performer, the EGARCH's performance is better in the S&P 500 and prediction errors are on lower levels overall in the S&P 500 than in the DAX. I find possible explanations for these observations to be:

1. The assumed distributions fit the S&P 500 returns better (Figure 7),
2. Volatility of volatility as well as level of volatility is lower in the S&P 500 (Table 1),
3. Skewness is more stable over time in the S&P 500 (Table 2),
4. The last period's realized variance is more important for the next in the S&P 500 (Table 8),
5. We have many more observations available in the S&P 500 (Table 1).

Points 1. - 3. as already mentioned help explain why GARCH performance is worsened. These points are also valid when explaining the performance differences of bull and bear markets. Point 2. in addition explain why simple averages perform better since they are sensitive to volatility of volatility, again in line with what was observed in the bull and bear split. Points 3. and 4. offer an explanation as to why the EGARCH performs better in the S&P 500; the leverage effect is likely more stable over time as indicated by the skewness and since this leverage effect is captured only by the latest observation in the EGARCH(1,1) specification, the fact that this observation is given more weight (Table 8) may allow the EGARCH to capture it better. Point 5) facilitates better predictions since a larger sample with which parameters are estimated gives better estimations of the "true" parameters. Here it should be noted that also this point is valid for the bull and bear split, and that it in fact serves to weaken the conclusion that bull and bear markets are fundamentally different since observed forecast differences might be attributed to a difference in sample size. The difference in performance of implied volatilities between the two stock indices is best explained by a difference in informational efficiency. It is well known that larger markets are in general more efficient than smaller and this may well explain the difference in accuracy between the DAX and S&P 500.

I contrast the stock index results with those of the $US/Euro since the returns of this index display different statistical properties. Accordingly we also see that the results are different. The difference between market states in terms of predictive accuracy is smaller in the exchange rate index than in the stock indices. The simple average models perform well relative to the other models and in comparison to the other indices, while the EGARCH and GARCH models fail to consistently outperform the simpler techniques. The implied volatilities are good in bull markets but do not on average outperform the moving averages in bear markets. Theory tells us that the low order of autocorrelation in the second moment (Figure 5b) of returns together with the stable level of variance (Table 2) benefit the moving average models relative to the GARCH models and these two properties partly explain the results. The changing Kuiper statistics in Figure 8 offer yet another explanation as to why the GARCH models are trailing in the exchange rate index since they rely on good parameter estimates, in turn relying on a stable distribution, to produce good predictions. Moreover, it is in accordance with expectations that the implied volatilities still predict well since they are model-independent. And although the GARCH models perform worse relative to the other models, all models deliver better predictions in absolute values than in the stock indices. This fact indicates a more predictable variance overall, which is logical seeing as the variance is more stable over time (Figure 3).

In general it is clear from the shifting results in bull and bear markets, and also among indices, that in order to outperform implied volatility forecasts by using GARCH variants, one needs to tailor the GARCH to the specific case by considering the shape of the distribution, the market state etc. This information is hardly ever available with certainty *ex ante* and this makes it hard to consistently outperform the implied volatilities. Moreover, if more data were available, potentially even better forecasts could be obtained from the implied volatilities by correcting any bias in the volatility indices, see e.g. Poon and Granger (2003) for a method of doing this. I experimented with this but was unable to achieve reasonable forecasts, possibly due to too few pre-sample observations.

The inconsistency in ranking and performance among measures is the main problem with my results. It does not invalidate the ranking of models but it implies that caution should be practiced when selecting models. If a model has to be chosen over another based on my results one would have to be careful to choose the model according to purpose. In other words, one has to be aware of what the loss functions actually measure before relying on them since they are not equivalent. No one model is superior on all accounts although the implied volatility stands out as the most stable performer over all measures and settings, with an emphasis on bull markets. In the overall results the EGARCH stands out and in the bear markets it also seems wise to employ a GARCH specification. However, if the analyzed market is characterized by low dependence in the second moment and little volatility clustering, a simple moving average predicts the variance better than all of the more advanced models tested in this thesis. The main problem with employing more advanced models is that they rely on estimated parameters that need not be stable over time. I experimented with a rolling window for the GARCH models, so as to allow them to better adjust to a changing distribution, but this only served to worsen the forecasts. I find that one in many cases has nothing to gain, and maybe even something to loose, from allowing the model more freedom, exemplified by the inclusion of a skew parameter. In order to make use of the apparent differences between market states one would need a way to identify these states in close to real time.

Questions open for future research are for example if the presented results hold up with higher frequency data, allowing for better approximations of the realized variance, and for other forecast horizons. 30 periods is a relatively long horizon for variance forecasting and GARCH type models usually perform much better at shorter horizons relative to other models. Some of the findings presented are not directly related to this thesis' research question but are still interesting in a wider perspective and also offer suggestions for future research. Specifically, the stock returns of the S&P 500 and DAX index cannot be modeled as Hansen's t or Student's t, let alone Gaussian, and pass statistical tests. Moreover, the shape of the distributions of these returns change significantly over time and I deduce that more flexible models are needed to fully capture the dynamics of stock returns; allowing for conditional heteroscedasticity, skewness and kurtosis based on historical observations is not enough.

6 Conclusions

I conclude that variance forecasting is in general more accurate in bull markets than in bear markets and that errors increase between 30 % and 50 % in terms of R^2 and QLIKE in bear markets. The MSE increases as much as with a factor of 15. This is in contrast to Brownlees et al. (2011) who find small differences in forecast accuracy between what they term 'calm and storm'. It is however in line with the findings of Chiang and Huang (2011) who find that GARCH models forecasting S&P 500 implied volatilities display higher mean absolute errors and root mean squared errors in bear markets than in bull markets. My results indicate that the difference in prediction accuracy between the two states in my sample is due to the following reasons:

1. The level of variance is around 1.5 to 2 times higher in bear markets for all indices where bear market predictions are worse than bull market predictions.
2. The volatility of volatility is around 3 to 4 times higher in bear markets for all indices where bear market predictions are worse than bull market predictions
3. The shape of the distribution of returns often changes over time and this causes parameters estimated on available information to be erroneous in predicting the future
4. The bear market states are shorter and thus models based on historical information (both simple mean models but also those using previous information to estimate other parameters) will suffer more from bull observations in the rolling window during bear markets than from bear observations in the rolling window during bull markets
5. Markets are less informationally efficient in bear states, affecting predictions based on the markets' beliefs

I find that the best forecasts are given by the market implied volatilities in bull markets and by an EGARCH(1,1) or GARCH(1,1) in bear markets, the model choice depending on whether the index shows a stable skewness over time or not. I also find that bear states affect some models more than others and change the ranking through this. The EGARCH is the only model sometimes showing indications of being better in bear states than in bull states and it is the model handling the change of state best in relative terms. The simple average models are affected heavily by bear markets and it is in general harder to distinguish between the performance of models in bear markets. A corollary of my findings is that the choice of model has to be tailored for specific purposes since different loss functions give different rankings. If the forecaster for some reason has a preference which character is given by one of the loss functions in particular, or closely related to it, the correct model choice can differ from the overall best forecast model. For example, a forecaster sensitive to relative errors may want to employ an implied volatility technique in a situation where a forecaster sensitive to absolute errors may want to employ an EGARCH model. My general conclusions are based on the aggregate results over all loss functions and do not account for the forecaster's preference.

References

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, March 2003. ISSN 1468-0262. doi: 10.1111/1468-0262.00418. URL http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00418/abstract.

Torben G Andersen, Tim Bollerslev, Francis X Diebold, Torben G. Andersen A, Tim Bollerslev B, and Francis X. Diebold C. Parametric and nonparametric volatility measurement. *IN L.P. HANSEN AND Y. AIT-SAHALIA (EDS.), HANDBOOK OF FINANCIAL ECONOMETRICS*, 2004. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.130.6479.

Torben G. Andersen, Tim Bollerslev, Peter Christoffersen, and Francis X. Diebold. Volatility forecasting. *SSRN eLibrary*, February 2005. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=673405.

Torben G. Andersen, Tim Bollerslev, and Nour Meddahi. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234, January 2011. ISSN 0304-4076. doi: 10.1016/j.jeconom.2010.03.032. URL http://www.sciencedirect.com/science/article/pii/S0304407610000746.

Fischer Black and Myron S Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–54, 1973. URL http://ideas.repec.org/a/ucp/jpolec/v81y1973i3p637-54.html.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986. ISSN 0304-4076. URL http://econpapers.repec.org/article/eeeeconom/v_3a31_3ay_3a1986_3ai_3a3_3ap_3a307-327.htm.

Chris Brooks. *Introductory Econometrics for Finance*. Cambridge University Press, 2 edition, June 2008. ISBN 052169468X.

Christian T. Brownlees, Robert F. Engle, and Bryan T. Kelly. A practical guide to volatility forecasting through calm and storm. *SSRN eLibrary*, August 2011. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1502915.

Peter Carr and Liuren Wu. A tale of two indices. *SSRN eLibrary*, 2005. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=871729.

Min-Hsien Chiang and Hsin-Yi Huang. Stock market momentum, business conditions, and GARCH option pricing models. *Journal of Empirical Finance*, 18(3):488–505, June 2011.

ISSN 0927-5398. doi: 10.1016/j.jempfin.2011.01.004. URL http://www.sciencedirect.com/science/article/pii/S0927539811000053.

K Demeterfi, E Derman, M Kamal, and J Zou. More than you ever wanted to know about volatility swaps. *Goldman Sachs Quantitative Strategies Research Notes*, 1999.

Francis X. Diebold and Roberto S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, July 1995. ISSN 0735-0015. doi: 10.2307/1392185. URL http://www.jstor.org/stable/1392185.

Francis X. Diebold, Andrew Hickman, Atsushi Inoue, and Til Schuermann. Converting 1-Day volatility to h-Day volatitlity: Scaling by root-h is worse than you think. Center for Financial Institutions Working Paper 97-34, Wharton School Center for Financial Institutions, University of Pennsylvania, 1997. URL http://ideas.repec.org/p/wop/pennin/97-34.html.

Louis H Ederington and Wei Guan. Forecasting volatility. *Journal of Futures Markets*, 25(5): 465–490, May 2005. ISSN 1096-9934. doi: 10.1002/fut.20146. URL http://onlinelibrary.wiley.com/doi/10.1002/fut.20146/abstract.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. URL http://econpapers.repec.org/article/ecmemetrp/v_3a50_3ay_3a1982_3ai_3a4_3ap_3a987-1007.htm.

Stephen Figlewski. Forecasting volatility. *Financial Markets, Institutions & Instruments*, 6 (1):1–88, February 1997. ISSN 1468-0416. doi: 10.1111/1468-0416.00009. URL http://onlinelibrary.wiley.com/doi/10.1111/1468-0416.00009/abstract.

Stephen Figlewski and Xiaozu Wang. Is the 'Leverage effect' a leverage effect? *SSRN eLibrary*, November 2000. URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=256109.

Bruce E Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–30, 1994. URL http://ideas.repec.org/a/ier/iecrev/v35y1994i3p705-30.html.

John C. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, 6 edition, June 2005. ISBN 0131499084.

Carlos M. Jarque and Anil K. Bera. A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique*, 55(2):163–172, 1987. ISSN 0306-7734. doi: 10.2307/1403192. URL http://www.jstor.org/stable/1403192.

Samuel Karlin and Howard M. Taylor. *An Introduction to Stochastic Modeling, Third Edition*. Academic Press, 3 edition, February 1998. ISBN 0126848874.

Michael Kirchler. Underreaction to fundamental information and asymmetry in mispricing between bullish and bearish markets. an experimental study. *Journal of Economic Dynamics and Control*, 33(2):491–506, 2009. URL http://ideas.repec.org/a/eee/dyncon/v33y2009i2p491-506.html.

N H Kuiper. Tests concerning random points on a circle. *Proceedings Of The Koninklijke Nederlandse Akademie Van Wetenschappen*, 63:38–47, 1962.

Asger Lunde and Allan Timmermann. Duration dependence in stock prices. *Journal of Business and Economic Statistics*, 22(3):253–273, July 2004. ISSN 0735-0015, 1537-2707. doi: 10.1198/073500104000000136. URL http://amstat.tandfonline.com/doi/abs/10.1198/073500104000000136?journalCode=ubes20.

Nour Meddahi. A theoretical comparison between integrated andRealized volatilities / a theoretical comparison between integrated and realized volatilities. CIRANO Working Paper 2001s-71, CIRANO, 2001. URL http://ideas.repec.org/p/cir/cirwor/2001s-71.html.

Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, Econometric Society, 1991. URL http://ideas.repec.org/a/ecm/emetrp/v59y1991i2p347-70.html.

Whitney K. Newey and Kenneth D. West. A simple, positive Semi-Definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, May 1987. ISSN 0012-9682. doi: 10.2307/1913610. URL http://www.jstor.org/stable/1913610.

Adrian R. Pagan and Kirill A. Sossounov. A simple framework for analysing bull and bear markets. Journal of applied econometrics, John Wiley & Sons, Ltd., 2003. URL http://ideas.repec.org/a/jae/japmet/v18y2003i1p23-46.html.

Andew J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, January 2011. ISSN 0304-4076. doi: 10.1016/j.jeconom.2010.03.034. URL http://www.sciencedirect.com/science/article/pii/S030440761000076X.

Ser-Huang Poon and Clive W. J. Granger. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539, 2003. URL http://ideas.repec.org/a/aea/jeclit/v41y2003i2p478-539.html.

Ruey S. Tsay. *Analysis of financial time series*. John Wiley and Sons, August 2005. ISBN 9780471690740.

Anders Wilhelmsson. Garch forecasting performance under different distribution assumptions. *Journal of Forecasting*, 25(8):561–578, December 2006. ISSN 1099-131X. doi: 10.1002/for.1009. URL http://onlinelibrary.wiley.com/doi/10.1002/for.1009/abstract.

David Williams. *Probability with Martingales*. Cambridge University Press, February 1991. ISBN 0521406056.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 1 edition, October 2001. ISBN 0262232197.

A Appendix

A.1 VIX Methodology

I follow the example given in CBOE's VIX white paper[25]. First, non-zero bid price, out of the money calls and puts are selected and centered around the at the money price. The forward (underlying) index level F is found by identifying the strike price for the option with the smallest call-put spread and using this strike in the formula:

$$F = strike + e^{rT}\left[\text{Call Price} - \text{Put Price}\right]$$

where Call Price and Put Price refers to the price of the Call and Put option associated with the given strike price.

With F defined, K_0 is found by taking the strike price of the option closest below F. And with K_0 found we start to include out of the money put options with successively lower strikes $(K_{-1}, .., K_{-n})$ until we reach bid prices of zero, i.e. no buyers. Then, the same is done for out of the money calls with successively higher strikes $(K_1, .., K_m)$. Lastly, the price of the option with strike K_0 is computed as the average of the put and call price for this strike. This procedure is done for so called near-term and next-term options, i.e. options that are to expire on the next and second from now settlement date. With these values, σ^2 is computed as the weighted 30 days moving average and VIX is thus given by

$$VIX = 100 \times \sqrt{\left(\left[T_1\sigma_1^2\left(\frac{N_{T_2} - N_{30}}{N_{T_2} - N_{T_1}}\right)\right] + \left[T_2\sigma_2^2\left(\frac{N_{T_2} - N_{30}}{N_{T_2} - N_{T_1}}\right) + \right]\right) \times \frac{N_{365}}{N_{30}}}$$

where the subscripts 1 and 2 (and accordingly subscripts T_1 and T_1) indicate near-term and next-term respectively, N_{T_i} is the time in minutes to settlement and N_n is the number of minutes in an n-day period.

A.2 Software

In the making of this thesis I make use of the following software and related resources:

* MatLab with additional code from Optimization Toolbox[26], Econometric Toolbox[27], Oxford MFE Toolbox[28], Eric Jondeau and Michael Rockinger[29] and Andrew Patton[30]

[25] http://www.cboe.com/micro/vix/vixwhite.pdf
[26] http://www.mathworks.se/products/optimization/
[27] http://www.mathworks.se/products/econometrics/
[28] http://www.kevinsheppard.com/wiki/MFE_Toolbox
[29] http://www.hec.unil.ch/matlabcodes/index.html
[30] http://public.econ.duke.edu/~ap172/code.html

* LyX[31]
* Zotero[32]

[31] http://www.lyx.org/
[32] http://www.zotero.org/