Forecasting stock index volatility

Walenkamp, M.

Marieke Walenkamp

Forecasting stock index volatility

Master's thesis, defended on July 17, 2008

Thesis advisor: E. van Zwet

Mathematisch Instituut, Universiteit Leiden

Contents

CHAPTER 1

Introduction

For the last 25 years modelling and forecasting asset return volatility has been a very active area of research in finance. When referring to a return process, most of the time volatility is defined as the standard deviation of the process. It measures the magnitude of the random component of the return. For this reason most people interpret volatility as uncertainty, since the larger the volatility, the larger the random, unpredictable component of the return and so the less certain one can be about the expected return. Whereas it has turned out to be very difficult, if not impossible, to predict future asset returns from historical returns, it has in fact been concluded in numerous studies that there is predictability in the volatility of asset returns. Accurately modelling and forecasting volatility is important since volatility is an important variable in many areas of finance, like risk management, option pricing and also asset management: the volatility linked product market is growing rapidly. Over the last few years a large number of volatility products has been introduced, like variance options, variance corridors and volatility and variance swaps. The latter is the most widely traded and interest is still growing: compared to three years ago, total traded volume increased tenfold. It is because of this rapid growth that the Investment Strategy team of Aegon Asset Management, for which this research has been carried out, is interested in volatility forecasting. It is interested in particular in the predictability of "long" horizon volatility, that is beyond a few weeks, since its focus is on long horizon strategies.

A first question when modelling and forecasting asset return volatility, is what *exactly* it is that we want to model, that is how to define volatility exactly, in mathematical terms. Nowadays asset returns are available at a tick-by-tick basis. However, if we want to quantify the unpredictability of a certain asset's return over the last month, considering these one second interval returns, will only add noise. On the other hand, we do not want to throw away too much information either. All of the above mentioned volatility products define the volatility over an n-days interval as square root of the sum of the n squared daily returns in the interval. This definition is used to determine their payoff and therefore it is this what we will try to forecast.

In the existing literature, there are three main classes of asset return volatility models, which are the so called stochastic volatility models, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) class of models and the realized volatility models. However, almost all studies focuse on short horizon volatility modelling and forecasting, that is up to one month maximum but often just one day ahead. Therefore there is little known about the forecastability of volatility at longer horizons, even though it is of considerable interest: financial instruments

may have short holding periods, but holding periods of several weeks or even months are equally common.

This thesis investigates the predictability of the volatility of the Standard and Poor's (S&P) 500 stock index returns for forecasting horizons from 10 to 120 days. The S&P 500 is a stock market index containing the stocks of 500 large, mostly American, companies. We consider the volatility of only this stock index, since for most of the volatility linked products this is the most common underlier. That is, the volatility of this particular index is traded most often. First of all, we will generate volatility forecasts using two of the existing forecasting models: the GARCH model and the HAR model, a linear regression model. Forecasting performances of these two models are assessed using both statistical criteria and an economic criterion, which focuses on how well the direction of change in volatility is predicted. For the various forecasting horizons we compare the performance of these two forecasting models to a simple benchmark model. For each model we test whether adding macroeconomic variables increases its forecasting power. In addition, we test how changing sampling frequency influences forecasting performance. At the end we test whether the HAR model's forecasting results can be improved when using two modern regression techniques to cleverly select regressors: boosting and forward stepwise regression.

The remainder of the thesis is organised as follows. In Chapter 2 some general concepts of volatility are discussed. We explain why in practice volatility is often defined as square root of the sum of squared returns. In addition, the three main types of volatility forecasting models are discussed. We motivate our choice of models. In Chapter 3 some additional properties of the HAR and GARCH model are discussed. Testing procedures are outlined. Chapter 4 describes our testing data. In Chapter 5 empirical results for the two forecasting models are analyzed using statistical criteria. Chapter 6 concentrates on the economic interpretation of the empirical results. In chapter 7 we investigate whether the HAR model's results presented in the chapters 5 and 6 can be improved when applying the modern regression techniques. Chapter 8 concludes.

CHAPTER 2

Theoretical background

In this chapter, we give an overview of what has been written in the literature on asset return volatility. We create understanding of why in practice, volatility is often defined as square root of the sum of squared daily returns. We select two models out of the existing volatility forecasting models.

2.1 Concepts of volatility

Consider the discretely sampled series $\{p_t\}_{t=1}^T$, where p_t denotes the logarithmic price of an asset at time t and where the unit interval corresponds to one day. We define r_t as the continuously compounded return on the asset over the interval $[t-1, t]$, so $r_t = p_t - p_{t-1}$. Conditional on the information set $F_{t-1} = \{p_1, p_2, ...p_{t-1}\}$, we can write this one period return r_t as the sum of a deterministic value μ_t and a random variable ϵ_t:

$$(1) \qquad r_t | F_{t-1} = \mu_t + \epsilon_t$$

μ_t is the expected conditional mean return, i.e. the expected return conditional on F_{t-1}: $\mu_t = \mathrm{E}\{r_t | F_{t-1}\}$. ϵ_t is the return shock, the stochastic part of the return. It can be expressed as a mean zero, variance one, serially uncorrelated (white noise) process, z_t, scaled by time-varying conditional standard deviation:

$$(2) \qquad r_t | F_{t-1} = \mu_t + \epsilon_t = \mu_t + \sigma_t z_t$$

It is this decomposition that underlies the GARCH model, which will be discussed in the next section.

These discrete returns are often interpreted as discrete samplings from an underlying, continuous time diffusion model:

$$(3) \qquad dp(t) = \mu(t)dt + \sigma(t)dW(t)$$

where $p(t)$ again denotes the logarithmic price process of the asset, $\mu(t)$ is called the drift, $\sigma(t)$ the instantaneous volatility and $W(t)$ is a Wiener process. The return over the $[t-1, t]$ time interval equals

$$(4) \qquad r(t) = p(t) - p(t-1) = \int_{t-1}^t \mu(s)ds + \int_{t-1}^t \sigma(s)dW(s)$$

Note the resemblance between this expression and the one in (2).

In this continuous time setting the variance over the interval $[t-1, t]$ equals

$$(4) \qquad \int_{t-1}^t \sigma^2(s)ds$$

This expression is often referred to as the *integrated variance* IV(t). It is the square root of this IV(t) which is usually meant by the volatility over the interval $[t-1, t]$. The problem is that this IV is an unobserved variable, since we only observe discrete samplings from the continuous proces (3). However, it can be approximated by an observable variable called realized variance.

Assuming that in the interval $[t-1, t]$ we have m intraday return observations, we define the following daily variance estimator:

$$(6) \qquad \widehat{\sigma}^2_{(m),t} = \sum_{i=1}^{m} r^2_{m, t-1+\frac{i}{m}}$$

where $r^2_{m, t-1+\frac{i}{m}}$ is the return calculated over the interval $[t-1+\frac{i-1}{m}, t-1+\frac{i}{m}]$. From the theory of quadratic variation (see Karatzas and Shreve(1988)), it follows that the estimate in (6) converges in probability to the integrated variance over the period $[t-1, t]$:

$$(7) \qquad \lim_{m \to \infty} \widehat{\sigma}^2_{(m),t} \xrightarrow{p} \int_{t-1}^{t} \sigma^2(s)ds$$

So the integrated variance is theoretically observable from the sample path of the return process, as long as the sampling process is frequent enough. The measure $\widehat{\sigma}^2_{(m),t}$ is indicated by the term *realized variance* and its square root by *realized volatility* (RV).

Most studies concentrate on modelling and forecasting daily volatility and assess the quality of their forecasts by comparing it to this sum of squared *intraday* returns. Using square root of the sum of squared *daily* returns as a proxy for for example monthly volatility, comes down to the same idea, but applied on a larger scale.

It must be stressed that we do not claim the theory in this section to be "true" or "false". We just discussed it to create understanding of where this sum of squared daily returns comes from.

2.2 The three main types of volatility forecasting models

In the previous section we explained some basic concepts of volatility. In this section we briefly discuss the three main types of volatility modeling and forecasting models, which are the GARCH, stochastic volatility and realized volatility (RV) based models. For now, just their basic concepts are discussed. A comparison of their predictive power is made in the next section.

GARCH

With the introduction of the ARCH(q) model, Engle (1982) set out the idea of modeling volatility as a time-varying function of current information. Note that this was long before the concept of RV was introduced, which was mid 90's. In the pre-RV era, the volatility over the interval $[t-1, t]$ was proxied by the squared return over the entire interval. That is, daily volatility was proxied by the squared daily return, without considering any subintervals. It is clear that this is a very

bad proxy. To give an example, if an asset behaves very wildly during one day, but its opening price happens to equal its closing price, its volatility during the day is estimated zero, when considering just the squared daily return instead of returns over several subintervals.

In the ARCH(q) model, time t variance is a function of q past squared returns. Some years later, Bollerslev (1986) introduced the Generalized ARCH (GARCH) class of models. The GARCH(p,q) model specifies time t variance as a function of the previous p variances and previous q squared return shocks. In practice, GARCH(p,q) models with small values of p and q are preferred. In fact, out of the GARCH class models the GARCH(1,1) has become the most widely used. As mentioned in the previous section, it is the decomposition in (2), here repeated in (8), that underlies the GARCH type models.

(8) $\qquad r_t = \mu_t + \sigma_t z_t, \qquad z_t \ i.i.d, \qquad E[z_t] = 0, \qquad Var[z_t] = 1$

The GARCH(1,1) model for the conditional variance is then defined by the recursive relationship

(9) $$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

where $\epsilon_t := \sigma_t z_t$, $\omega, \alpha, \beta \geq 0$ and $\alpha + \beta < 1$. The constraints $\omega, \alpha, \beta \geq 0$ are required to ensure that the conditional variance will never be negative. The constraint $\alpha + \beta < 1$ is needed to guarantee stationarity of the process. The parameters can be estimated by Maximum likelihood procedure. As we see, in this model a very high or very low time t return will directly lead to high time $t+1$ volatility.

Since the introduction of the GARCH model, several extensions have been proposed that account for features that are typically found in daily volatility estimates. The result is a long list of GARCH variants. For example, the Threshold GARCH (TGARCH) of Glosten et al. (1993), the Asymmetric GARCH (AGARCH) of Engle and Ng (1993) and the Exponential GARCH (EGARCH) model of Nelson (1991) all capture the stylized fact that a negative return shock leads to a higher conditional variance in the subsequent period than an equally large positive shock would. The Fractionally Integrated GARCH (FIGARCH) (p,d,q) model of Baillie et al. (1996) takes into account the so called long memory behavior of volatility. The plain GARCH model implies that shocks to volatility decay at an exponential rate (see section 3.2). However, empirical research has shown that volatility shocks last much longer and decay hyperbolically. The FIGARCH model is designed to capture this hyperbolical decay.

For a comprehensive overview of all GARCH variants, we refer to Hansen and Lunde (2005).

Stochastic volatility

The defining property of a stochastic volatility (SV) model is that it allows the volatility of the underlying asset to be partly stochastic. For example, in a continuous time setting, a SV model might specify the logarithmic price process as

$$dp(t) = \mu(t)dt + \sigma(t)dW(t)$$

where the instantaneous volatility $\sigma(t)$ is again some stochastic diffusion process. For example , in the popular Heston model, the differential equation for the variance takes the form

$$\mathrm{d}\sigma^2(t) = \theta(\omega - \sigma^2(t))\mathrm{dt} + \xi\sigma(t)\mathrm{d}\widetilde{W}(t)$$

where ω is the mean long-term volatility, θ is the rate at which the volatility reverts toward it's long-term mean, ξ is the volatility of the volatility process and $\mathrm{d}\widetilde{W}(t)$ is again a Wiener process. The correlation between $\mathrm{d}W(t)$ and $\mathrm{d}\widetilde{W}(t)$ is constant and equal to ρ.

The presence of a second Wiener process renders both the estimation as the forecasting problem far more complex for the SV models than for the GARCH models. Instead of simple maximum likelihood procedures, simulation based procedures will now have to be used. It is this disadvantage of the SV model that makes many researchers prefer GARCH over SV as their forecasting model.

Realized volatility

In section 2.1 we explained that in theory, the realized variance converges to the integrated variance, which is squared volatility, when the length of the intraday intervals goes to zero, see equation (7). Therefore it can be used to assess the quality of forecasts generated by GARCH or stochastic volatility models. It is logical however, to also model and forecast RV directly. It is understandable that, whereas volatility has long been modeled and forecasted using mainly GARCH and SV models, there has been a strong tendency the last years towards directly modelling the RV. Compared to the GARCH and SV models, these models basicly just skip one step, which is the parametrization of σ. We will indicate the models that directly forecast the RV simply by "RV-based models".

It turns out that almost all RV series share a few fundamental statistical properties. One property is that the logarithmic RV series is much more homoskedastic than the series itself and is approximately normal. In addition, RV seems to be a long memory process, that is shocks decay hyperbolically. It is because of these features that RV is typically modeled using Autoregressive Fractionally Integrated (ARFI) models:

$$(10) \qquad\qquad \Phi(L)(1 - L)^d(y_t - \mu) = \epsilon_t$$

Here d denotes the long-memory parameter, y_t represents the log of the RV and μ is the unconditional mean of the y_t series. $\Phi(L)$ is a polynomial lag operator accounting for autogressive structure, that is $\Phi(L) = 1 - \phi L - \phi L^2 - \phi L^3 - ...$, where $L^p y_t = y_{t-p}$ and $L^p \epsilon_t = \epsilon_{t-p}$.

Other RV-based models have been proposed though, like the Heterogeneous Autoregressive model for the Realized Volatility (HAR-RV) by Corsi (2004). This model is a simple autoregressive model for the RV that takes into account volatilities realized over several horizons. At day t, the model's forecast of the RV over day $t + 1$ is given by

$$(11) \qquad RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + w_{t+1}$$

where $RV_t^{(d)}$ is the RV over day t and $RV_t^{(w)}$ and $RV_t^{(m)}$ are the average daily realized volatilities over the past week and month, respectively.

Corsi (2004) shows that his model outperforms the widely used ARFI class of models. Furthermore, the model seems to capture the main empirical features of financial time series (leptokurtosis, long memory), while at the same time it stays very simple and easy to work with.

2.3 Overview of empirical research

In the previous section we discussed the basics of the three main types of volatility forecasting models. Most important however, when selecting a model for our forecasting purposes, is of course its forecasting power. In this section we compare the forecasting performance of each of the three types of models, on the basis of past empirical research. Keeping in mind our final objective, the focus is on the long horizon. We also discuss some of the variables that have been proposed as explanatory for stock index volatility.

Forecasting performance of the different models

Volatility has long been modeled and forecasted using mainly SV and GARCH models, with a tendency towards the last one, because of the earlier discussed complexity of SV models. Poon and Granger (2002) offer a very elaborate overview of former research on volatility forecasting. Nonetheless they report only four studies that directly compare the performance of these two models, three of which report superior performance of SV. But there is obviously no general conclusion to draw with such a small sample size.

Whereas there is no clear winner between the GARCH and SV models as far as predictive power is concerned, models that rely on RV have in fact been shown to clearly outperform models that do not, at least on the short horizon. In other words, on the short horizon the use of RV adds power to a forecasting model. This is no great surprise, since as explained the RV based models directly model the sum of squared subinterval returns and the GARCH and SV models don't. There is a long list of studies that have come to this conclusion. In most of these studies a comparison of predictive performance is made between models that in some way incorporate the concept of RV and (plain) GARCH type models.

Examples are Andersen et al. (2003), who show that ARFI models, as defined in equation (10), produce superior short horizon forecasts compared to the daily GARCH(1,1) model and Martens (2001), who finds that adding intraday information leads to improved forecasts.

It has been exhaustively proved by now that for the short horizon RV based models provide superior volatility forecasts. However, for a horizon beyond a month they no longer are clearly superior. The question remains which models have been found to provide good long-horizon forecasts, if any. Some studies have found volatility forecasting at horizons beyond a few weeks to be difficult, see for example West and Cho (1995) and Christoffersen and Diebold (2000). However, some more recent studies show that reasonable longer horizon forecasts in fact can be made. For example, Calvet and Fisher (2004) propose a discrete time, regime switching stochastic volatility model which they show to substantially dominate some of the

basic GARCH type models at horizons of up to 50 days. Another example is the model by Brandt and Jones (2006), who combine various EGARCH models with data on the daily range (highest minus lowest return during one day) and find substantial forecastability of volatility as far as one year ahead.

Explanatory variables

Even though some proposals have been made, little attention has been paid so far to the problem of finding a long horizon volatility forecasting model. On the other hand, there is a long list of studies that propose single variables that could be explanatory for long horizon volatility. We discuss a few.

It has been suggested in several studies (e.g. Franks and Schwartz (1991)) to incorporate the slope of the yield curve as an explanatory variable to the volatility model. This variable is usually referred to as *term spread* and can be proxied by the ten years interest rate less the three months rate. When the economic outlook is bad, the short term rate is often lowered, while usually the long-term rate is not affected much. So the term spread is expected to increase in this case. Since bad economic conditions in general go together with high volatility of stock returns, the term spread is expected to be positively correlated with volatility.
Using the same arguments, some studies (e.g. Whitelaw (1994), Harvey (1991)) use just the *three months treasury bill yield* as explanatory variable.

The use of the *credit spread*, defined as the Baa - Aaa corporate bond yield spread, to forecast volatility is suggested by Schwertz (1989) and Whitelaw (1994), among others. Schwertz finds that even in the presence of other variables this spread is positively related to future stock market volatility. Whitelaw finds less strong evidence of explanatory ability but does find some significance.

The *commercial paper - short term treasury yield* is found to be very informative in several studies, see for example Bernanke (1990) and Whitelaw (1994). Bernanke argues that the forecasting power arises from the spread's ability to proxy for the stance of monetary policy. It should be positively correlated with volatility, since commercial paper yields will increase when economic conditions are bad and consequently volatility is high. In combination with a decreasing treasury yield, this should lead to an increase of the spread, such that high volatility goes together with a high spread. The yields on six months commercial paper and three months treasury bills can be taken to measure the spread.

The *dividend yield* is used as explanatory variable by Harvey (1991), among others. Contrary to the other explanatory variables, this variable should be negatively correlated with volatility: if the economic outlook is bad and so volatility is high, companies will in general pay out low dividends.

Finally, the *daily high-low range*, defined as the highest minus the lowest quoted price during the day, has been pointed out by many as very informative. We already mentioned Brandt and Jones (2006), but also Andersen and Bollerslev (1998) and Taylor (1987), among others, point out the power of the range when forecasting volatility. volatility. When making a forecast for several months ahead, we could

also consider using the *weekly* or even *monthly* high-low range as explanatory variable. By construction, the range is positively correlated with volatility.

2.4 Model choice

In the previous section we concluded that not very much models have been proposed in the existing literature for forecasting volatility at the long horizon, there is definitely no clearly dominating model. There is however for the short horizon, namely the (daily) RV based models. As our forecasting models we choose the GARCH(1,1) model and the HAR-RV model that was briefly discussed in section 2.2. We select an RV-based model because it is most logical to directly model and forecast the RV, without using additional, unneccessary parametrizations. Out of of all RV-based models we choose the HAR-RV model, because even though other RV based models, like the ARFI model in equation (10), are more widely used, Corsi's results seem promising. Furthermore, the simplicity of the HAR-RV is a big advantage compared to the ARFI model, which is not as trivial to estimate. Another positive point is that it has a clear economic interpretation, which we will come back to in the next chapter. We test the forecasting power of the model both with and without the additional explanatory variables which were discussed in the previous section.

In addition to the HAR-RV model, we also test the GARCH(1,1) model. Again we test the model both with and without the extra variables added in the variance equation. We emphasize that although the GARCH model has its shortcomings, it does serve as a natural benchmark for the forecast performance of the traditional, not RV-based volatility models. Furthermore, we just forget about the SV models for now. Their complexity makes them unsuitable for our practical purposes.

In the next chapter we elaborate the two proposed models in more detail.

Chapter 3

In the previous chapter we decided to continue with the GARCH(1,1) and the HAR-RV model. In this chapter some additional properties of both models will be discussed. In the case of the GARCH(1,1) model we first derive the model's h-step ahead forecast of conditional volatility. Next the problem of choosing an appropriate sampling frequency is discussed. Finally our testing procedures are outlined. As for the HAR-RV model, first some background is provided, since in contrast to most other models, this model has a nice economic interpretation. Next also for this model some practical issues related to its implementation are discussed.

3.1 The GARCH(1,1) model

In this section we derive the GARCH(1,1) model's h-step ahead forecast of conditional volatility. We first derive an expression for the unconditional variance, since this turns out to both simplify calculations and provide us with an interpretation of the forecasts.

In section 2.2 we introduced the very basics of the GARCH(1,1) model. In particular, we gave the model's definition of time t conditional variance σ_t^2 in equation (9). Remember that σ_t^2 is defined conditional on the information set $F_{t-1} = \{p_1, p_2, ..., p_{t-1}\}$, that is: $\sigma_t^2 = E[(r_t - E[r_t])^2|F_{t-1}]$. The *unconditional* variance σ^2 is defined as the variance of the entire time series, i.e. $\sigma^2 = E[(r_t - E[r_t])^2]$. Taking into account the independence of z_t and σ_t and the fact that $E[z_t] = 0$ and $Var[z_t] = 1$, we find that σ^2 equals:

$$\sigma^2 = E[(r_t - E[r_t])^2] = E[(\sigma_t z_t)^2] = E[\sigma_t^2]E[z_t^2] = E[\sigma_t^2]$$

$$= E[\omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2] = E[\omega + \alpha\sigma_{t-1}^2 z_{t-1}^2 + \beta\sigma_{t-1}^2]$$

$$= \omega + (\alpha + \beta)E[\sigma_{t-1}^2]$$

$$= \omega + (\alpha + \beta)[\omega + (\alpha + \beta)E[\sigma_{t-2}^2]]$$

$$= \omega + (\alpha + \beta)\omega + (\alpha + \beta)^2\omega + ... = \frac{\omega}{1-\alpha-\beta}$$

provided that $\alpha + \beta < 1$. So the unconditional variance is finite and equal to $\frac{\omega}{1-\alpha-\beta}$ if and only if $\alpha + \beta < 1$.

Now we derive the model's h-step ahead forecast of conditional volatility, for any $h \geq 1$. At time t the one step ahead conditional variance σ_{t+1}^2 is deterministic. For $h \geq 2$ the h-step ahead conditional variance is not deterministic anymore, but its

expectation, i.e. a forecast is very easily derived:

$$E[\sigma^2_{t+h}|F_t] = E[\omega + \alpha \epsilon^2_{t+h-1} + \beta \sigma^2_{t+h-1}|F_t]$$

$$= \omega + \alpha E[\sigma^2_{t+h-1} z^2_{t+h-1}|F_t] + \beta E[\sigma^2_{t+h-1}|F_t]$$

$$= \omega + \alpha E[\sigma^2_{t+h-1}|F_t] E[z^2_{t+h-1}|F_t] + \beta E[\sigma^2_{t+h-1}|F_t]$$

$$= \omega + (\alpha + \beta) E[\sigma^2_{t+h-1}|F_t] = \sigma^2(1 - \alpha - \beta) + (\alpha + \beta) E[\sigma^2_{t+h-1}|F_t]$$

$$= \sigma^2 + (\alpha + \beta)[E[\sigma^2_{t+h-1}|F_t] - \sigma^2]$$

And iterating we find:

$$E[\sigma^2_{t+h}|F_t] = \sigma^2 + (\alpha+\beta)\left[\sigma^2 + (\alpha + \beta)[\sigma^2 + (\alpha + \beta)[E[\sigma^2_{t+h-2}|F_t] - \sigma^2] - \sigma^2]\right]$$

$$= \sigma^2 + (\alpha + \beta)^2[E[\sigma^2_{t+h-2}|F_t] - \sigma^2] = ...$$

(12) $$= \sigma^2 + (\alpha + \beta)^{h-1}[\sigma^2_{t+1} - \sigma^2]$$

So, as would be expected, expectations of future variance revert towards the unconditional variance as the forecast horizon increases. The term $(\alpha + \beta)$ is the rate of reversion.

Furthermore, since for any $i \geq 1$,

$$Cov[r_t, r_{t+i}] = Cov[\mu_t + \epsilon_t, \mu_{t+i} + \epsilon_{t+i}]$$

$$= Cov[\epsilon_t, \epsilon_{t+i}] = E[\sigma_t z_t \sigma_{t+i} z_{t+i}]$$

$$= E[\sigma_t z_t \sigma_{t+i}] E[z_{t+i}] = 0,$$

the return process is uncorrelated and so $Var[\sum_{i=t}^{h} r_i] = \sum_{i=t}^{h} Var[r_i]$.
Since the r_t are logreturns, the return over the multiple period interval $[t, t+h]$ is the sum of the single period sums. In other words, the above equation says that the variance of the total return over the interval $[t, t+h]$ simply is the sum of the one period variances. In particular, the time t forecast of the conditional variance over the interval $[t, t+h]$ equals the sum of the one-period variance forecasts derived in equation (12):

$$Var[r_{t+1} + ... + r_{t+h}|F_t] = \sum_{j=1}^{h} Var[r_{t+j}]$$

$$= \sum_{j=1}^{h} \left[\sigma^2 + (\alpha + \beta)^{j-1}(\sigma^2_{t+1} - \sigma^2)\right]$$

$$= h\sigma^2 + \sum_{j=1}^{h}(\alpha + \beta)^{j-1}(\sigma^2_{t+1} - \sigma^2)$$

And since

$$\sum_{j=1}^{h}(\alpha+\beta)^{j-1} = \sum_{j=0}^{h-1}(\alpha+\beta)^{j} = \sum_{j=0}^{\infty}(\alpha+\beta)^{j} - \sum_{j=h}^{\infty}(\alpha+\beta)^{j}$$

$$\frac{1}{1-\alpha-\beta} - (\alpha+\beta)^{h}\sum_{j=h}^{\infty}(\alpha+\beta)^{j-h} = \frac{1}{1-\alpha-\beta} - (\alpha+\beta)^{h}\sum_{j=0}^{\infty}(\alpha+\beta)^{j}$$

$$= \frac{1-(\alpha+\beta)^{h}}{1-\alpha-\beta}$$

we find

$$(13) \qquad Var[r_{t+1} + .. + r_{t+h}|F_t] = h\sigma^2 + \frac{1-(\alpha+\beta)^{h}}{1-\alpha-\beta}(\sigma_{t+1}^2 - \sigma^2)$$

as our time t forecast of the variance over $[t, t+h]$.

Testing procedure

In equation (13) we derived an expression for the GARCH forecast of the volatility over the period $[t, t+h]$. It is important to realize that this forecast relies on the sampling frequency, i.e. changing the sampling frequency will change the forcast. Assume for example that we want to make a forecast of the conditional variance over the next two weeks. We could work with daily returns, such that this forecast becomes a forecast of volatility over the next ten time periods. On the other hand we could use weekly returns, constructed from the daily returns, in which case the estimates for α and β will change. Furthermore, we are now making a forecast for only two periods ahead. It is clear from the expression in (13) that in general this will change the forecast.

So different frequencies give different forecasts, so it is important to choose well our sampling frequency. However, in the existing literature not much attention has been paid to this problem of finding the optimal sampling frequency given some forecasting horizon and some evaluation criterion. Most of the time the choice of frequency seems random, guided by intuition. In particular when it comes to inter-daily horizons, which is what we are interested in, almost no past research has been conducted on this subject. Andersen et al (1999) seem to be the only ones that do provide an answer. It is questionable though how valuable their results are, for our research. They compare the GARCH(1,1) forecasts of the one day, one week and one month ahead Deutschemark-US dollar volatility (proxied by the sum of squared returns, as in our case), for sampling frequencies ranging from five mintues to monthly. For all three horizons, they find that the best forecasts are made when using an hourly sampling frequency. However, they also find that with the increase of forecasting horizon the effect of changing from a daily to an hourly sampling frequency quickly declines. For the monthly horizon the difference is marginal. We consider even longer horizons than Andersen et al. That is why we will test the GARCH(1,1) model considering sampling frequencies of daily and lower only.

More precisely, we proceed as follows. We start with testing the "bare" GARCH(1,1) model, that is without any explanatory variables added to the variance equation. Over our sample period we make forecasts of 10, 20, 40, 60 and 120 days ahead S&P 500 stock index volatility, using a 1000 days rolling window. That is, at each step the GARCH parameters α, β and ω are estimated on the last 1000 days of data

and these values are plugged into equation (13) to generate the forecasts. Since we do not know which sampling frequency is optimal, forecasts are made using various sampling frequencies: daily, weekly, two-weekly and monthly. As said, our ex-post volatility measure over a given period is the sum of squared daily returns over that period. Various statistical criteria and an economic criterion are used to assess the quality of the forecasts. These criteria are discussed in chapters 5 and 6.

Next we repeat this testing procedure for the GARCH(1,1) model with the explanatory variables described in section 2.3 added to the GARCH variance equation. To start with, these variables are added one at a time. If it turns out that there are several variables that improve results of the bare GARCH model, we will also add combinations of these variables to see what their combined effect is.

3.2 The HAR-RV model

The HAR-RV model is a realized volatility model proposed by Corsi (2004). He shows that it is able to incorporate the main stylized properties of asset volatility series. At the same time the model stays very simple and therefore easy to implement. The model in particular is designed to reproduce the stylized fact of long memory in the volatility of (daily) return series. Even though in general there is no significant correlation between returns for different days, the correlations between the *magnitudes* of returns on nearby days, i.e. between squared or absolute returns, are in fact positive and significant. This positive correlation between magnitudes is indicative of positive correlation in the volatility process: remember the definition of volatility as a measure of the magnitude of the return shock.

The standard GARCH models imply that shocks to volatility decay at an exponential rate. To give an example, remember that at time t the GARCH(1,1) model expresses the h-step ahead forecast of volatility as in equation (12), that is:

$$\sigma^2_{t+h|t} = \sigma^2 + (\alpha + \beta)^{h-1}(\sigma^2_{t+1|t} - \sigma^2)$$

where

$$\sigma^2_{t+1|t} := \omega + \alpha\epsilon^2_t + \beta\sigma^2_t$$

Consequently, in this model the effect of the time t return shock ϵ^2_t on the h-step ahead volatility forecast is:

$$\frac{\partial\sigma^2_{t+h|t}}{\partial\epsilon^2_t} = \alpha(\alpha + \beta)^{h-1}$$

And since α, $\beta \geq 0$ and $\alpha + \beta \leq 1$, we can infact write

$$\frac{\partial\sigma^2_{t+h|t}}{\partial\epsilon^2_t} = cx^h$$

where $0 < x < 1$, showing that the GARCH(1,1) model indeed implies exponential decay of shocks to the volatility.

However, empirical observations provide evidence that autocorrelations of squared and absolute returns do not decline at an exponential rate, but at a much slower hyperbolic rate. In other words, the effect of a volatility shock lasts much longer then implied by for example the GARCH(1,1) model.

Several models have been proposed that do take into account this long memory be-
havior of volatility. Often just a fractional differencing operator of the form $(1-L)^d$
is added to some existing model. Examples are the ARFI model in equation (10)
and the Fractionally Integrated GARCH (FIGARCH) (1,d,1) model proposed by
Baillie et al. (1996).

Corsi (2004) agues that even though these fractionally integrated (FI) models do
incorporate the long memory property, they have their drawbacks. First of all
he argues that fractional integration lacks an economic interpretation. Second, FI
models are often difficult to estimate and heuristic methods may lead to largely
biased parameter estimates. That is why Corsi proposes an alternative model,
which he names the Heterogenous Autoregressive model for the Realized Volatility
(HAR-RV). The HAR-RV model is just a simple AR model for realized volatility,
that takes into account volatilities realized over several horizons. Corsi focuses on
modeling the daily RV. In this case the HAR-RV model is specified by the single
equation

$$(11) \qquad RV_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + w_{t+1}$$

where $RV_t^{(d)}$ represents the volatility realized over day t and $RV_t^{(w)}$ and $RV_t^{(m)}$ are
the average daily realized volatilities over the past week and month, respectively:

$$RV_t^{(w)} = \frac{1}{5}(RV_{t-1}^{(d)} + RV_{t-2}^{(d)} + RV_{t-3}^{(d)} + RV_{t-4}^{(d)} + RV_{t-5}^{(d)})$$

Analogously, $RV_t^{(m)}$ is defined as the average of the realized volatilities over the
past 20 days.

The basic idea behind the model is that in a market not all agents are identi-
cal. In particular, agents operate on different time horizons. It is because of this
heterogeinity that different reaction times to news can be observed, which causes
different volatility components. Roughly speaking, the $RV_t^{(d)}$ term in equation (11)
corresponds with the volatility caused by market participants that in their trading
activities only consider the very short, (intra)daily horizon, like market makers. On
the other hand, $RV_t^{(w)}$ and $RV_t^{(m)}$ reflect the decisions of agents that focus on the
weekly and monthly horizons as well.

This idea, in combination with the findings of LeBaron (2001) that the sum of
three AR(1) processes, each considering a different time horizon, can give rise to
hyperbolically decaying memory, leads to Corsi's proposal of the HAR-RV model.

Testing procedure

Like for the GARCH(1,1) model, we can use different sampling frequencies
when testing the HAR-RV model. Corsi (2004) only considers the daily frequency,
i.e. he makes one week ahead forecasts by five times recursion of equation (11). But
we could also make a one week ahead forecast by means of the following regression:

$$(12) \qquad RV_{t+1}^{(w)} = c + \beta^{(w)} RV_t^{(w)} + \beta^{(2)} RV_t^{(2)} + \beta^{(3)} RV_t^{(3)} + w_{t+1}$$

where $RV_t^{(w)}$ again denotes the average daily RV over the past week and $RV_t^{(2)}$ and
$RV_t^{(3)}$ are the average daily RV's over *some* time periods. We could for example
set $RV_t^{(2)}$ equal to the average daily RV over the previous month and $RV_t^{(3)}$ to

the average daily RV over the previous three months. It is clear though, that for the HAR-RV model finding the optimal sampling frequency for a given forecasting horizon is less straightforward than for the GARCH(1,1) model, since changing the length of the last two components changes the forecast. In other words, compared with the GARCH(1,1) model, we are now dealing with two additional parameters. So a first decision to make is how to choose these components. For each frequency we have a very large number of possibilities. To start with, we just choose one "intuitively logical" realisation for each sampling frequency and in the next chapters we first test how the model performs for this choice. For the daily frequency we simply follow Corsi and set the components equal to the previous one day, one week and one month average RV's. For the weekly frequency we set them equal to the previous one week, one month and three months average RV's, for the twoweekly frequency to the previous two weeks, two months and six months average RV's and for the monthly frequency to the previous one, three and six months average RV's. In chapter 7 we try to improve the forecasting results of the HAR model by choosing regressors more smartly by means of two modern regression techniques called boosting and forward stepwise regression.

The rest of the testing procedure is the same as for the GARCH model: 10, 20, 40, 60 and 120 days forecasts are made using a 1000 days rolling window, which in this case means that the β's and the constant c are estimated on the last 1000 days of data. Again, tests will be performed on the bare HAR model as well as the HAR model with the explanatory variables from section 2.3 added. Like for the GARCH model, these are added one at a time to start with.

CHAPTER 4

Data

Our testing data consists of S&P 500 stock index returns during the period
from 6 December 1988 to 18 September 2007. The S&P 500 is a very notable stock
market index, containing the stocks of 500 (mostly American) corporations. Our
sample period covers both relatively calm as well as turbulent periods. To explain
the seemingly random starting date, we note that the full set of exogenous variables
has been available at a daily basis no earlier then 6 December 1988. In figure 1 the
annualized realized volatilities over 10, 60 and 120 days are plotted for our entire
sample period. Table 1 shows their summary statistics. As would be expected,
we find that the longer the horizon, the more the volatility spikes are smoothed
out: the standard deviation decreases, as does the maximum of the series, while
the minimum increases.

From the graphs in figure 1 we see that in the period 1996-2003 realized volatility
has been consistently high. The first five years of this period correspond with the
IT bubble, a speculative bubble during which stock markets saw their value increase
rapidly from growth in the new Internet sector and related fields. The five years of
increasing stock prices were followed by a two year bear market in which the S&P
500 lost approximately 50% of its value. Remember that the RV is calculated as the
square root of the sum of squared returns, meaning that both consistently highly
positive as consistently highly negative returns lead to high RV. This explains why
during the whole period 1996-2003 RV has been high. During the two year bear
market the most dramatic declines happened in the period July-October 2002. The
most pronounced spikes in the 60 and 120 days RV series correspond with this
period. Also in the 10 days RV series a cluster of spikes is observed for this period,
but the highest peak occurred in August 1998. On August 17 Russia defaulted on
its government debt, triggering major drops in worldwide stock markets. In one
month the S&P 500 lost 14.5 % of its value. The third high spike in the 10 days RV
series corresponds with the mini-crash of 27 October 1997. On this day worldwide
stock markets crashed, caused by an economic crisis in Asia. In one day the S&P
500 lost over 7%.
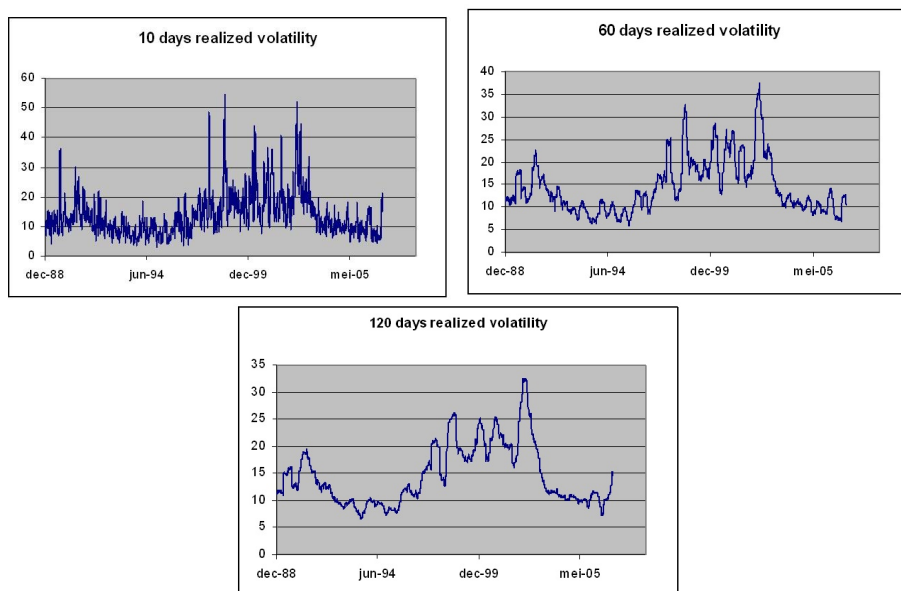
FIGURE 1. **Annualized realized volatility over 10, 60 and 120 days for the period 6 December 1988 to 18 September 2007. The annualized realized volatility over a n-day period is calculated as the square root of the sum of squared daily returns during that period, times a factor $\frac{252}{n}$.**

	10 days RV	60 days RV	120 days RV
mean	13.89	14.43	14.60
st.dev	7.15	5.98	5.59
min	2.95	5.91	6.64
max	54.41	37.45	32.55

Table 1. Summary statistics for the series plotted in figure 1.

CHAPTER 5

Empirical results: statistical analysis

To asses the quality of the forecasts, various criteria can be used. In this chapter the focus will be on *statistical* criteria. Using several performance measures, we discuss the statistical accuracy of our forecasts. Under these measures an upward error and an downward error of the same magnitude are judged equally bad, without taking into account any economic interpretation. Performances of the HAR and GARCH forecasting models are compared to the performance of a very simple benchmark model.

5.1 Performance measures

We use different performance measures to evaluate the quality of our forecasts. First of all, we calculate the mean squared error (MSE) with the actual (ex-post) RV series for each series of forecasts:

$$(12) \qquad MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i - \sigma_i)^2$$

where N denotes the total number of forecasts, $\hat{\sigma}_i$ the period i volatility forecast and σ_i the ex-post RV over period i.

In addition we report, for each forecast series $\hat{\sigma}$, the so called coefficient of determination, R^2, obtained from the regression of realized volatility on forecasted volatility:

$$\sigma = a + b\hat{\sigma} + \epsilon$$

where $\epsilon \sim N(0,1)$. R^2 is the fraction of the sample variance $\frac{1}{N} \sum_{i=1}^{N} (\sigma_i - \sigma_{mean})^2$ that is explained by the forecasting model. It equals

$$R^2 = \frac{b^2 \sum_{i=1}^{N} (\hat{\sigma}_i - \hat{\sigma}_{mean})^2}{\sum_{i=1}^{N} (\sigma_i - \sigma_{mean})^2}$$

where $\hat{\sigma}_{mean}$ and σ_{mean} are the means of the forecasting series and realized volatility series respectively. To understand this equation, note that $\sigma_{mean} = a + b\hat{\sigma}_{mean}$ and so $\sigma_i - \sigma_{mean} = b(\hat{\sigma}_i - \hat{\sigma}_{mean}) + \epsilon_i$. So the difference from the mean $(\sigma_i - \sigma_{mean})$ can be decomposed as the sum of a component corresponding to the mean of the forecast series and an unexplained component, described by the residual ϵ_i, implying that the fraction of $\frac{1}{N} \sum_{i=1}^{N} (\sigma_i - \sigma_{mean})^2$ that is explained by the forecasting model is indeed the above expression for R^2.

In addition to the R^2's, we also report the coefficients a and b of the regression, since these can provide us extra information about the quality of a series $\hat{\sigma}$. In particular, we are interested whether b significantly differs from 0, since this would

mean that there is a certain degree of predictability in the volatility. For each regression we test the null hypotheses $H_0^1 : b = 0$ and $H_0^2 : a = 0$, by means of t-tests.

5.2 Results

In tables 3a-b MSE's are reported for all HAR and GARCH forecasting models (i.e all horizons, all sampling frequencies, all extra variables). In the Appendix the R^2's and the regression coefficients a and b with corresponding p-values can be found in tables A1, A2 and A3. These results are evaluated against the results from a benchmark model which forecasts tomorrow's volatility as a linear combination of today's volatility and a constant:

$$\hat{\sigma}_{t+1} = \hat{c} + \hat{d}\sigma_t$$

where $\hat{c}$ and $\hat{d}$ are least squares estimates, obtained from the regression

$$\sigma = c + d\sigma^{(-1)} + \epsilon$$

where the notation $\sigma^{(-1)}$ is used to indicate the one period lag with respect to σ.

So for example, at day t a forecast of the RV over the next 120 days is made as:

$$\overset{\overbrace{t+120}}{\underset{t+1}{\sum}} r_i^2 = \hat{c} + \hat{d} \sum_{i=t-119}^{t} r_i^2$$

So this model could be described as a sophisticated version of "the best prediction of tomorrow's volatility is today's volatility". Like the HAR and GARCH models, this benchmark model uses a 1000 days rolling window, meaning that at each time step c and d are estimated based on the last 1000 days of data. Table 2 below gives the results for the benchmark model.

benchmark			horizon		
	10	20	40	60	120
MSE's	1.458	2.373	3.619	5.441	9.887
R^2's	0.393	0.436	0.504	0.470	0.357
a	-0.145	-0.082	0.033	0.478	2.645
	(0.479)	(0.826)	(0.942)	(0.397)	(0.001)
b	0.986	0.963	0.959	0.906	0.727
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 2: MSE's with the ex-post RV series for the forecast series generated by the benchmark model and regression R^2's and regression coefficients of the regression $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the ex-post RV series and $\hat{\sigma}$ the forecast series generated by the benchmark model. In brackets the p-values from the t-tests for a=0 and b=0.

GARCH

Based on tables 3a and A1a our most important conclusion is that for the GARCH model the daily frequency gives best results for all horizons. From table

3a we clearly see that MSE's increase as the sampling frequency gets lower. The R^2's in table A1a show similar results: for all horizons the highest values are found when using a daily sampling frequency.

GARCH MSE's			horizon		
Bare model	10	20	40	60	120
daily	1.329	2.047	3.659	5.432	11.418
weekly	1.539	2.548	4.551	6.448	11.385
two-weekly	1.908	3.221	5.943	8.822	17.551
monthly		5.682	9.959	14.634	25.284
Added: range					
daily	1.385	2.172	3.814	5.603	11.705
weekly	1.582	2.629	4.834	6.867	12.020
two-weekly	1.923	3.507	6.370	9.204	17.097
monthly		5.130	8.206	11.789	21.663
Added: credit spread					
daily	1.340	2.081	3.697	5.472	11.463
weekly	1.596	2.627	4.638	6.577	11.694
two-weekly	1.830	3.149	5.901	8.736	17.588
monthly		5.013	8.945	12.279	21.270
Added: dividend yield					
daily	1.326	2.036	3.652	5.412	11.409
weekly	1.574	2.615	4.637	6.570	11.656
two-weekly	1.894	3.278	6.130	9.000	17.800
monthly		5.214	9.141	13.132	22.917
Added: 3 months treasury bill yield					
daily	1.333	2.061	3.681	5.466	11.494
weekly	1.584	2.626	4.652	6.610	11.802
two-weekly	1.780	3.032	5.801	8.635	17.574
monthly		5.020	9.001	12.817	22.189
Added: term spread					
daily	1.324	2.040	3.652	5.424	11.410
weekly	1.619	2.651	4.524	6.442	11.531
two-weekly	1.782	2.992	5.809	8.645	17.520
monthly		5.610	9.584	9.316	23.343
Added: 6m com.paper - 3m treasury bill yield					
daily	1.329	2.053	3.661	5.430	11.496
weekly	1.561	2.569	4.662	6.635	11.762
two-weekly	1.844	3.053	5.781	8.573	17.379
monthly		5.253	8.807	12.595	21.431

Table 3a: MSE's with the ex-post RV series for all forecast series generated by the GARCH model.

Comparing table 3a with table 2, we find that for this daily sampling frequency, the GARCH model at the 10 and 20 days horizon beats the benchmark model, that is MSE's are lower. This result holds no matter which explanatory variable is added. At the 40 and 120 days horizon the benchmark model is never beaten, while for 60 days ahead this depends on which variable is added: the bare model and models with dividend yield, term spread or 6 months commercial paper minus 3 months treasury bill yield all lead to decreases in MSE with respect to the benchmark, the others don't. Improvements do not seem very large though. Therefore, in the next section we test whether the decreases in MSE wrt the benchmark model are

significant, by means of the Wilcoxon signed rank.

Comparing tables 2 and A1a (concentrating again on the daily sampling frequency), we find that also R^2's are higher for the GARCH model than for the benchmark model, for all horizons.

Looking at the tables A2a-g, the most striking result is that for all horizons, all sampling frequencies and all explanatory variables added, the null hypothesis $H_0^1 : b = 0$ is very clearly rejected, that is at all reasonable confidence levels. This is a nice result, since it means that all these series are better forecasts then when taking just a constant, indicating that there is predictability in the volatility. This does not tell us anything about the quality of the forecasts though, that is with respect to the benchmark model. In fact, we see from table 2 that also for the benchmark model the null hypothesis $H_0^1 : b = 0$ is very clearly rejected for all horizons. Summarizing, we find evidence of predictability but in the next section we test whether a very simple benchmark model can be beaten.

Finally, in the figures 2a-b the 20 and 120 days ahead forecast series generated by the daily sampled bare GARCH model are plotted. As we see, no strange spikes occur in the forecast series, i.e. all spikes are a result of spikes in the RV series. So in this sense the model actually behaves "nicely". However, for the 120 days horizon, forecasts tend to be too low in the first part of the sample and in the last part they are consistently too high. This effect is less strong when extra variables are added, see Appendix figures A1a-d for some examples.
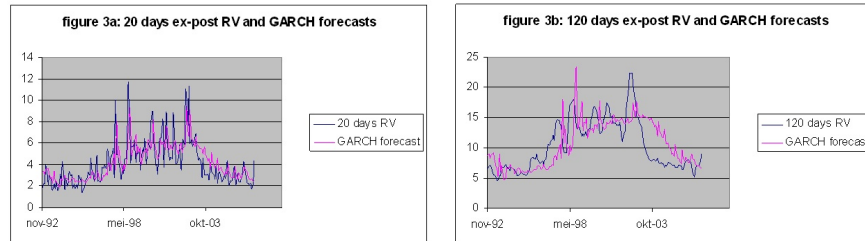
FIGURES 2A-B:**20 and 120 days RV and corresponding forecasts generated by the daily sampled bare GARCH model.**

HAR

A first conclusion we can make from the results in tables 3b and A1b is that for the HAR model the weekly sampling frequency gives best results: MSE's are lowest and R^2's highest for this frequency.

HAR MSE's			horizon		
Bare model	10	20	40	60	120
daily	1.552	2.448	4.601	6.966	14.502
weekly	1.405	2.236	3.952	5.758	11.718
two-weekly	1.421	2.451	4.167	6.037	12.296
monthly		2.897	4.963	7.553	15.388
Added: range					
daily	1.449	2.149	3.934	5.648	10.925
weekly	1.404	2.230	3.747	5.214	9.941
two-weekly	1.487	2.460	4.307	6.217	12.129
monthly		3.060	4.561	6.770	11.724
Added: credit spread					
daily	1.549	2.539	4.559	6.662	12.706
weekly	1.450	2.284	4.009	5.814	11.401
two-weekly	1.462	2.537	4.326	6.216	12.029
monthly		2.793	4.539	6.558	12.153
Added: dividend yield					
daily	1.605	2.703	5.351	8.316	18.237
weekly	1.508	2.442	4.333	6.343	13.087
two-weekly	1.473	2.517	4.217	6.106	12.340
monthly		3.151	5.546	8.293	15.690
Added: 3 months treasury bill yield					
daily	1.554	2.454	4.676	7.112	14.917
weekly	1.443	2.343	4.302	6.435	13.602
two-weekly	1.459	2.556	4.486	6.690	14.197
monthly		2.993	5.234	8.003	16.486
Added: term spread					
daily	1.557	2.490	4.797	7.400	16.085
weekly	1.403	2.256	4.118	6.166	13.263
two-weekly	1.437	2.525	4.479	6.688	14.240
monthly		3.073	5.428	8.333	17.020
Added: 6m com.paper - 3m treasury bill yield					
daily	1.552	2.441	4.658	6.967	14.703
weekly	1.434	2.333	4.314	6.458	13.750
two-weekly	1.456	2.556	4.504	6.701	14.345
monthly		3.003	5.290	8.039	16.968

Table 3b: MSE's with the ex-post RV series for all forecast series generated by the HAR model.

Comparing tables 3b and 2, we find that for this weekly sampling frequency the HAR model at the 10 and 20 days horizon beats the benchmark, except when the dividend yield is added to the model. For all other variables added and for the bare HAR model, MSE's are lower, but the dividend yield for some reason makes MSE's increase with respect to the benchmark. Like for the GARCH model, the benchmark model is never beaten at the 40 and 120 days horizon. At the 60 days horizon, only the HAR model with the range added beats the benchmark. Again, significance of the decreases in MSE's is tested in the next section.

From tables A4a-g we see that, like for the GARCH model, the null hypothesis $H_0^1 : b = 0$ is, for all horizons, sampling frequencies and explanatory variables added, rejected at all reasonable confidence levels. So again, predictability of volatility is found, but based on these tables we cannot draw any conclusions yet about the

quality of the forecasts.

In the figures 3a-b the 20 and 120 days ahead forecast series generated by the weekly sampled bare HAR model are plotted. Like for the GARCH forecast series in figures 2a-b, all spikes can be explained by spikes in the RV series. The same holds when extra variables are added to the bare model, see Appendix figures A2a-d for a few examples.
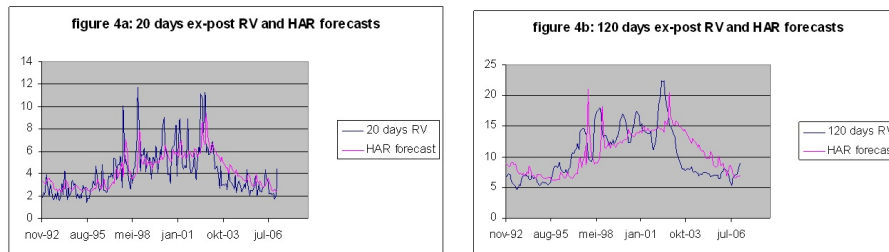
FIGURES 3A-B: **20 and 120 days RV over the period November 1992-February 2007 and corresponding forecasts generated by the weekly sampled bare HAR model.**

5.3 The Wilcoxon signed rank test

To test whether one model significantly outperforms the other, we use the Wilcoxon signed rank test.

Let $\{\sigma\}$ denote the true, ex-post volatility series and let $\{\widehat{\sigma}^{(1)}\}$ and $\{\widehat{\sigma}^{(2)}\}$ denote two competing forecast series of σ. The time t forecast errors from the two models are:

$$\epsilon_t^{(1)} = \sigma_t - \widehat{\sigma}_t^{(1)}$$

and

$$\epsilon_t^{(2)} = \sigma_t - \widehat{\sigma}_t^{(2)}$$

The Wilcoxon signed rank test statistic is based on the differences, D_t, between these forecasting errors:

$$D_t := \epsilon_t^{(1)} - \epsilon_t^{(2)}$$

The null hypothesis is that the two forecasting error series have the same distribution and so that no model significantly outperforms the other. That is, under the null hypothesis the distribution of the D_i is symmetric around 0. The test statistic is calculated as follows. To start with, the *absolute values* of the D_t's are ranked. That is, letting N denote the total number of D_t's, each D_t obtains a unique rank $R_t \in \{1, .., N\}$, such that if $R_i < R_j$, then it must hold that $|D_i| < |D_j|$. Next, *signed* ranks are obtained by restoring the sign of the D_t's to the R_t's. Finally, the test statistic W is defined as the sum of the ranks with *positive* sign. If the two forecasting errors series follow the same distribution, we expect approximately half of the D_i to be positive and half to be negative, in which case W will not be very large or very small. Therefore, the null hypothesis is rejected when W does in fact take on a too extreme value.

It can be easily shown (see Rice (1995)) that, under the null hypothesis, $E(W) = \frac{1}{4}N(N+1)$ and $Var(W) = \frac{1}{24}N(N+1)(2N+1)$ and that the normalized test statistic

$$Z = \frac{W - E(W)}{\sqrt{Var(W)}}$$

is asymptotically normally distributed. Since our sample period covers a period of almost 20 years, we can in fact consider normalized test statistics. That is, we test the null hypothesis $H_0 : Z \sim N(0,1)$, where

$$(13) \qquad Z = \frac{W - \frac{1}{4}N(N+1)}{\sqrt{N(N+1)(2N+1)/24}}$$

We want to test whether the benchmark forecast series is beaten by any of the HAR or GARCH models, that is whether the forecasting errors resulting from these competing forecasting models are significantly lower. Therefore, we set $\{\hat{\sigma}^{(1)}\}$ equal to the forecasting error series implied by the benchmark series and $\{\hat{\sigma}^{(2)}\}$ to the error series implied by the various HAR and GARCH forecast series. In the tables A4a and A4b, in the Appendix, the normalized test statistic Z is reported for each choice of $\{\hat{\sigma}^{(2)}\}$. We are interested in the case $Z > 1.96$. In this case the forecasting errors of the benchmark model are significantly higher, at the 2.5% confidence level, than the errors of the competing HAR and GARCH forecast model.

In table A4a the results for the GARCH forecast series are reported. Values greater than 1.96 are in bold. As we see, these are only found at the 20 days horizon and when using a daily sampling frequency. This result holds, no matter which explanatory variable is added. Results for the HAR forecast series, which are reported in table A4b, are even worse: There are no values greater than 1.96 at all, meaning that the benchmark model is beaten by none of the HAR models.

Finally, it must be stressed that when testing at the 2.5% confidence level, we accept the possibility to falsely reject the null hypothesis 2.5 out of 100 times. In our case we accept the possibility of falsely rejecting three out of the test statistics in tables A4a and A4b, since in each table a total of 133 test statistics is reported. To correct for this falsely rejecting the null hypothesis, which is expected to happen more often the larger the number of times the hypothesis is tested, several proposals have been made. The most rigorous example is the Bonferroni correction, which states that when testing a hypothesis n times at the $\alpha\%$ confidence level, a p-value threshold of $\frac{\alpha}{n}$ should be used. A less restrictive criterion is the false discovery rate, which says that first the p-values of the n tests should be ordered in increasing order. Denoting these n orderes p-values by $p_{(1)}, p_{(2)}, ...p_{(n)}$, the largest k should be found such that $p_{(k)} \leq \frac{k}{n}\alpha$. Then the null hypothesis should be rejected for the first k tests. Under these criteria also for the GARCH model all test statistics, in table A4a, drop to insignificance.

Summarizing, we find no convincing evidence that any of the HAR or GARCH forecasting error series is significantly lower than the benchmark error series. This is of course a disappointing result. Therefore, in chapter 7 we will try to improve results by applying two modern regression techniques. In the next chapter we first

test the HAR and GARCH models again, but under a very different (but equally important) criterion.

CHAPTER 6

Empirical results: economic interpretation

In the previous chapter we discussed the quality of our forecast series using statistical criteria, considering only the *magnitude* of the forecasting errors. In practice however, the *sign* of the error is often also important as it can mean the difference between a loss or a profit. In this chapter we first discuss the variance swap, as an example of a contract of which the payoff depends not only on the magnitude but also on the sign of the forecasting error. Next we calculate and discuss the so called hitratios for all forecasting models and all horizons, i.e. how often does a model predict well the direction of change in volatility. Again the HAR and GARCH models' performances are compared with results of our simple benchmark model, described in the previous chapter.

6.1 Variance swap

Variance swap contracts are swap contracts that offer investors direct exposure to the volatility of the underlying asset, usually a stock index. A pre-agreed variance level will be exchanged for the actual variance realized over the contract period. The net payoff P will be the difference between these two:

$$P = X(\sigma^2 - K^2)$$

where K is the variance swap strike price, σ^2 is the realized variance of the underlying asset and X is the variance swap notional, the amount of money invested in the swap.

The realized variance σ^2 is calculated as the annualized sum of squared daily log returns:

$$\sigma^2 = \frac{252}{T} \sum_{i=1}^{T} [\ln(\frac{P_i}{P_{i-1}})]^2$$

where P_i is the (closing) stock price on day i and T is the number of days of the contract period.

So apart from the term $\frac{252}{T}$ which annualizes the expression, this is the same definition we used when making our forecasts. So if our model predicts that the variance level over the next T days ($\hat{\sigma}_T$) will be higher then the variance realized over the last T days (σ_{-T}), a strategy could be to take a long position in a variance swap with strike level K equal to σ_{-T}. It is clear that the most important thing is that the variance realized over the next T days will in fact be higher then σ_{-T}, since otherwise a loss will be made. So in this example it is important to predict well the direction of change in volatility.

6.2 hitratios

Tables 5a and 5b report the hitratios for the various HAR and GARCH forecast series. For a forecast series, the hitratio is calculated as the number of times that the direction of change in realized volatility is forecasted correctly, divided by the total number of forecasts:

$$\frac{1}{N}\sum_{t=1}^{N}(I_{\hat{\sigma}_t>\sigma_{t-1}\&\sigma_t>\sigma_{t-1}} + I_{\hat{\sigma}_t\leq\sigma_{t-1}\&\sigma_t\leq\sigma_{t-1}})$$

where N is the total number of forecasts, $\{\sigma\}$ denotes the true, ex-post volatility series and $\{\hat{\sigma}\}$ is the forecast series of $\{\sigma\}$. In other words: given today's volatility, how often is forecasted correctly wether tomorrow's volatility will be higher or lower. To start with, the hitratios for the benchmark model, against which the HAR and GARCH hitratios will be evaluated, are reported in table 4 below.

benchmark hitratios	horizon				
	10	20	40	60	120
	0.624	0.626	0.618	0.650	0.651

Table 4: hitratios for the forecast series generated by the benchmark model

GARCH

In table 5a hitratios are reported for the GARCH model. Based on this table, a first conclusion is that for all horizons hitratios are highest when using a daily sampling frequency and decline as the frequency declines. This observation is in agreement with the statistical results from the previous chapter, where for the GARCH series we found full sample MSE's to be lowest for the daily frequency as well.

A second observation is that the GARCH model leads to highest hitratios at the shortest horizons. From table 5a we find the hitratios to be highest for the 10 days horizon and lowest for the 120 days horizon. This result is quite in contrast with the results of the benchmark model, for which the highest hitratio is found for the 120 days horizon. Consequently, comparing tables 4 and 5a (concentrating on the daily sampling frequency, which we found to give best results), we see that at the 10 and 20 days horizon the GARCH model always beats the benchmark, that is no matter what explanatory variable added. On the other hand, at the 40, 60 and 120 days horizon, the benchmark model is never beaten.

HAR

For the HAR model, a first observation is that none of the sampling frequencies leads to overall highest hitratios. Actually, it seems to be quite random for which frequency hitratios are highest, for a given horizon. Remember that under statistical criteria the weekly sampling frequency did outperform the other frequencies for all horizons.

Secondly, like for the GARCH model, the highest hitratios are reported for the shortest horizons. Again, the model does best at the 10 days horizon and worst at the 120 days horizon. However, in contrast to the GARCH results, the HAR

model outperforms the benchmark model not only at the shortest howizons, but at all except the 120 days horizon. It does though, for every horizon, really depends on which sampling frequency is used and which explanatory variable is added to generate a forecast series, wether or not that series beats the benchmark series. There is only one combination of sampling frequency and explanatory variable for which the benchmark is beaten for all four horizons, which is the combination of credit spread and weekly sampling frequency.

GARCH	horizon				
Bare model	10	20	40	60	120
daily	0.669	0.665	0.612	0.644	0.580
weekly	0.649	0.620	0.596	0.571	0.523
two-weekly	0.657	0.626	0.590	0.559	0.511
monthly		0.597	0.551	0.503	0.494
Added: range					
daily	0.671	0.659	0.601	0.644	0.580
weekly	0.652	0.609	0.596	0.588	0.534
two-weekly	0.663	0.603	0.601	0.537	0.517
monthly		0.581	0.545	0.492	0.5
Added: credit spread					
daily	0.663	0.643	0.618	0.644	0.586
weekly	0.632	0.609	0.601	0.582	0.511
two-weekly	0.638	0.592	0.584	0.548	0.529
monthly		0.581	0.584	0.554	0.5
Added: dividend yield					
daily	0.677	0.659	0.612	0.638	0.569
weekly	0.646	0.620	0.618	0.582	0.540
two-weekly	0.652	0.587	0.562	0.514	0.5
monthly		0.553	0.534	0.486	0.489
Added: 3 months treasury bill yield					
daily	0.660	0.654	0.612	0.633	0.586
weekly	0.644	0.620	0.590	0.576	0.529
two-weekly	0.644	0.637	0.584	0.537	0.506
monthly		0.575	0.562	0.497	0.494
Added: term spread					
daily	0.663	0.665	0.618	0.644	0.586
weekly	0.635	0.615	0.590	0.565	0.523
two-weekly	0.646	0.620	0.573	0.508	0.494
monthly		0.575	0.534	0.475	0.529
Added: 6m com.paper - 3m treasury bill yield					
daily	0.669	0.665	0.612	0.644	0.586
weekly	0.644	0.615	0.601	0.593	0.540
two-weekly	0.638	0.615	0.584	0.559	0.506
monthly		0.592	0.545	0.508	0.517

Table 5a: hitratios for all forecast series generated by the GARCH model

Finally, when comparing the HAR and GARCH results for the 10 and 20 days horizon (for which both models outperform the benchmark model), we see that for these horizons the overall highest values are reported for the GARCH model. So based on the information in tables 4, 5a and 5b, our conclusion is that at the 10 and 20 days horizon best results are found for the GARCH model, at the 40 and 60 days horizon for the HAR model and at the 120 days horizon for the benchmark model.

HAR	horizon				
Bare model	10	20	40	60	120
daily	0.639	0.626	0.596	0.610	0.489
weekly	0.644	0.637	0.573	0.621	0.506
two-weekly	0.649	0.620	0.573	0.616	0.511
monthly		0.620	0.596	0.588	0.552
Added: range					
daily	0.65	0.654	0.685	0.633	0.569
weekly	0.627	0.615	0.579	0.633	0.632
two-weekly	0.657	0.609	0.573	0.638	0.557
monthly		0.615	0.590	0.610	0.661
Added: credit spread					
daily	0.667	0.637	0.601	0.655	0.563
weekly	0.660	0.659	0.622	0.672	0.580
two-weekly	0.683	0.631	0.607	0.661	0.575
monthly		0.637	0.579	0.599	0.580
Added: dividend yield					
daily	0.633	0.615	0.579	0.605	0.546
weekly	0.641	0.615	0.601	0.621	0.540
two-weekly	0.641	0.609	0.590	0.616	0.563
monthly		0.654	0.629	0.621	0.592
Added: 3 months treasury bill yield					
daily	0.642	0.643	0.607	0.599	0.483
weekly	0.630	0.631	0.601	0.593	0.477
two-weekly	0.652	0.626	0.618	0.610	0.494
monthly		0.620	0.567	0.588	0.477
Added: term spread					
daily	0.639	0.598	0.601	0.593	0.471
weekly	0.657	0.603	0.596	0.605	0.477
two-weekly	0.666	0.620	0.624	0.576	0.460
monthly		0.615	0.545	0.565	0.511
Added: 6m com.paper - 3m treasury bill yield					
daily	0.642	0.615	0.596	0.582	0.580
weekly	0.644	0.637	0.601	0.582	0.563
two-weekly	0.657	0.620	0.579	0.576	0.563
monthly		0.615	0.596	0.588	0.546

Table 5b: hitratios for all forecast series generated by the HAR model

CHAPTER 7

Modern regression techniques

So far we tested the HAR model choosing the regressors more or less randomly, in a statistical sense: for each sampling frequency we tested the model for just one combination of three RV's over different past periods, which seemed "intuitively logical" from an economical point of view. We just followed Corsi (2004) at this point, who's choice of regressors is also just economically motivated. This procedure is of course not very acceptable from a statistical point of view. Also, we tested whether the macroecomic variables add predictive power by simply adding them to the model one at a time and then running the model with one variable added over the whole sample period. Again, it is clear that this is not the best way of testing. In this chapter we test the HAR model again, but this time select the regressors by means of two more advanced regression techniques: forward stepwise regression and boosting. Both techniques select a set of regressors at each time step, so this set will possible change over time. Remember that in chapter 5, using statistical criteria, we did not find evidence that the HAR model led to significant improvement of the benchmark model. It is interesting to find out whether these two modern regression techniques can improve the accuracy of the forecasts such that the benchmark is in fact beaten, which is our ultimate goal. Since in chapter 5 we found the weekly sampling frequency to be optimal under statistical criteria, this is the only frequency we consider in this chapter. That is, we try to improve the accuracy of the weekly sampled HAR model only.

7.1 Forward stepwise regression

Say we want to predict a certain value Y by means of a linear regression model and say that we have a set of k possible regressors: $\{X_j\}_{j=1}^k$. Given the histories of Y and the X_j, we could just decide to use *all* k regressors and estimate the linear regression model

$$Y = \beta_0 + \sum_{j=1}^k X_j \beta_j$$

by means of the least squares method. One thing we know is that the sum of squared errors , which for a given set of m regressors is defined as

$$SSE(\beta) = \sum_{j=1}^m (y_i - \beta_0 - \sum_{j=1}^m X_{ij}\beta_j)^2$$

will get lower when more regressors are added. Consequently it will be lowest when *all* regressors are added. However, in general the variance of the estimates will get larger when adding more regressors. When a lot of regressors are added, prediction accuracy may be very low. Furthermore, with a large number of regressors it is probable that a smaller subset exhibits the strongest effects.

These two considerations are the basic ideas behind forward stepwise regression. Instead of using all regressors we use only a few. We start with an intercept and then step by step out of the remaining regressors, the regressor which most improves the fit is added, until none of the remaining regressors significantly improves the model anymore. This improvement in fit is often based on the F-statistic:

$$F = \frac{SSE(\hat{\beta}) - SSE(\widetilde{\beta})}{SSE(\widetilde{\beta})/(N - m - 2)}$$

where the current model of m regressors corresponds with parameter estimates $\hat{\beta}$ and the model with one of the remaining predictors added to this current model with parameter estimates $\widetilde{\beta}$. N is the sample size. It must be stressed that the total set of parameters is estimated again each time an extra regressor is added. In other words, the values of $\widetilde{\beta}$ are not just equal to $\hat{\beta}$ with the exception of the one extra element, but it is a whole different set of values.

In our case, our set of regressors consists of the six macroeconomic variables defined earlier and a set of average RV's over various past periods. Remember that in the previous chapters we used the average RV's over the past week, month and three months as regressors for the weekly sampled HAR model (see section 3.2). However, in principle RV's over *all* past periods could be chosen as regressors, that is there is an infinite number of potential regressors. To avoid ending up with too large sets of possible regressors and to avoid any problems with data availability, we choose only 26 regressors out of this infinite set. These 26 regressors are the average RV's over the past week, past two weeks, past three weeks etc, with the last regressor being the average RV over the last half year. These 26 regressors bring the total number of possible regressors to 32. To make the comparison with the simple HAR model we tested in the previous chapters as fair as possible, we again use a 1000 days rolling window. That is, at each time step we select a subset out of these 32 regressors to make a forecast and determine parameters, based on the last 1000 days of data. Our strategy is to sequentially add the regressor which produces the largest value of the F-statistic and stop when no predictor leads to a value for F greater than the 95th percentile of the $F_{1,N-m-2}$ distribution.

Table 6 below presents the results for the HAR model when regressors are selected by means of forward stepwise regression. Reported are again the MSE's as defined in equation (12), the R^2 from the regression of realized volatility σ on our forecasts $\hat{\sigma}$:

$$\sigma = a + b\hat{\sigma} + \epsilon$$

and the a and b from this regression together with the p-values from the t-tests of $H_0^1 : b = 0$ and $H_0^2 : a = 0$. Also the Wilcoxon signed rank test statistics with corresponding p-values are reported, to test for significant improvement of the benchmark model. Using the notation of section 5.3, we test MSE's of the benchmark forecast series, $\{\hat{\sigma}^{(1)}\}$, against the MSE's of the forecast series generated by the foreward stepwise regression, $\{\hat{\sigma}^{(2)}\}$.

| Forward stepwise | horizon | | | | |
regression	10	20	40	60	120
MSE's	1.187	2.150	2.816	4.816	9.376
W. signed rank test-statistics	**2.211**	**2.019**	1.789	1.003	0.537
R^2's	0.560	0.578	0.676	0.528	0.508
a	0.473 (0.000)	0.943 (0.000)	1.194 (0.000)	2.517 (0.000)	4.354 (0.000)
b	0.775 (0.000)	0.722 (0.000)	0.768 (0.000)	0.625 (0.000)	0.562 (0.000)

Table 6: results for the forward stepwise regression: MSE's with the ex-post RV series, Wilcoxon signed rank test statistics with $\{\hat{\sigma}^{(1)}\}$ equal to the benchmark forecast series and $\{\hat{\sigma}^{(2)}\}$ to the forecast series resulting from the forward stepwise regression, regression R^2's and regression coefficients of the regression $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the ex-post RV series and $\hat{\sigma}$ the forecast series generated by the forward stepwise regression. In brackets the p-values from the t-tests for a=0 and b=0.

Comparing tables 1, 3b and 6, a first observation is that for the stepwise regression R^2's are very high for all horizons, higher than both the benchmark R^2's and the simple HAR R^2's. Remember from section 5.1 that this means that the forward stepwise regression model explains a larger part of the variance of the actual volatility than the benchmark model and the simple HAR model. Like for the benchmark model, also for the stepwise regression model the highest R^2 is reported for the 40 days horizon and the lowest R^2 for the 120 days horizon. However this minimum is still as high as 0.508.

In addition, we see that for all horizons the MSE's implied by the forward stepwise regression are (much) lower than those implied by both the simple HAR model from the previous chapters and the benchmark model. However, at the 2.5% confidence level, the Wilcoxon signed rank test statistic is greater then 1.96 only at the 10 and 20 days horizons, meaning that only at these horizons the forecasting errors implied by the forward stepwise regression model are significantly lower than the errors implied by the benchmark model. However, under the false discovery rate, described in section 5.3, also at the 20 days horizons the test statistic becomes insignificant. At the 10 days horizon the test statistic remains significant. So all together, results indicate that the forward stepwise regression model does better than the simple HAR model but the model does not outperform the benchmark model convincingly beyond the 10 days horizon.

7.2 Boosting

Boosting is a general method of producing an accurate prediction by combining a large set of rough, inaccurate predictions. It was originally designed for classification problems, but it can be extended to regression problems as well.

Let us again assume we want to predict a variable Y and have a vector X of k

predictor variables. In general terms, a boosting algorithm produces a basis functions expansion of the form

$$f(x) = \sum_{m=1}^{M} \beta_m b(x, \gamma_m)$$

where the β_m are the expansion coefficients and the $b(x, \gamma_m)$ the basis functions: simple functions of the predictor variables and a set of parameters γ. The expansion resulting from the boosting algorithm, $f_{opt}(x)$, typically minimizes some loss function over the training data:

$$f_{opt}(x) = \min_{\{\beta_m, \gamma_m\}_1^M} \sum_{i=1}^{N} L(y_i, \sum_{m=1}^{M} \beta_m b(x_i, \gamma_m))$$

where N is the size of the training sample. For many loss functions, finding f_{opt} is computationally very intensive. Therefore often f_{opt} is just approximated.

We now explain how to apply the concept of boosting to our forecasting problem. In our case, we are dealing with squared error loss, so the loss function is:

$$L(y, f(x)) = (y - f(x))^2$$

Our set of predictor variables is the same as in the previous chapter, that is the six macroeconomic variables and the 26 RV's over past periods. We use the 32 (32×1)-vectors with 31 elements equal to 0 and one element equal to one as the basis functions, γ is empty. So actually the basis functions are simply the regressors. To approximate the solution which minimizes the loss function, we will use an approximation algorithm called forward stagewise (FS) additive modelling. The basic idea is similar to the forward stepwise regression discussed in the previous section: also FS additive modelling sequentially adds new regressors (basis functions) to the expansion, but *without* adjusting the parameters and coefficients of those that are already added. Remember that the regression model described in the previous section estimates the coefficients again at each time step. Also, while the model from the previous chapter does not add any regressor more than once to the expansion, the FS additive modelling can add the same regressor multiple times. In fact, the number of iterations M may be (and often will be) larger than the number of predictor variables, implying that at least one of the regressors *has* to be added more than once. Note that adding the same regressors more than once only makes sense due to the fact that coefficients of all previous added regressors do not change. At each iteration m, the FS additive modelling solves for the optimal basis function and corresponding coefficients β_m to add to the current expansion $f_{m-1}(x)$. Adding this optimal term to the current expansion produces $f_m(x)$ and the procedure is repeated.

Boosting	horizon				
	10	20	40	60	120
MSE's	0.741	1.723	2.341	4.067	7.935
W. signed rank test statistics	**4.231**	**2.339**	**2.410**	1.829	1.526
R^2's	0.721	0.673	0.578	0.535	0.510
a	0.500	1.193	1.610	1.749	4.532
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
b	0.908	0.792	0.741	0.717	0.662
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 7: results for the boosting model: MSE's with the ex-post RV series, Wilcoxon signed rank test statistics with $\{\hat{\sigma}^{(1)}\}$ equal to the benchmark forecast series and $\{\hat{\sigma}^{(2)}\}$ to the forecast series resulting from the boosting model, regression R^2's and regression coefficients of the regression $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the ex-post RV series and $\hat{\sigma}$ the forecast series generated by the boosting model. In brackets the p-values from the t-tests for a=0 and b=0.

Table 7 presents results for the forecast series generated by the boosting model when using FS additive modelling to approximate the optimal solution. We set the number of iterations M equal to 50. We find that for all horizons the MSE's implied by the boosting model are much lower than the benchmark and HAR MSE's in tables 1 and 3b respectively. They are also lower than the MSE's implied by the forward stepwise regression model in table 6. The Wilcoxon signed rank test statistic is greater than 1.96 for the 10,20 and 40 days horizon and is significant at these horizons even under the false discovery rate criterion!

In addition, the boosting R^2's are for all horizons higher than the R^2's for the other three models. The boosting R^2's differ from the benchmark and forward stepwise regression R^2's in the sense that they decrease monotonically with the increase of horizon, while for the other models the maximum is reported at the 40 days horizon. For the 10 days horizon the boosting R^2 equals 0.721, which is an extremely high value in the context of volatility forecasting. In the existing literature values higher than 0.55 are rarely reported.

Summarizing, there is very strong evidence that the boosting model outperforms the benchmark model at the 10, 20 and 40 days horizon. Taking also in consideration the very R^2's for the boosting model we may conclude the boosting model is preferred over the benchmark model.

CHAPTER 8

Conclusion

We studied the predictability of S&P 500 stock index volatility for forecasting horizons from 10 to 120 days. Forecasts were made with the GARCH and HAR model and using daily, weekly, two-weekly and monthly sampling frequencies. The added value of several explanatory variables was tested as well. Results of the two forecasting models were compared to the outcomes of a very simple benchmark model.

Under both statistical criteria (MSE's, R^2's) and the economic criterion (hitratios), we find the GARCH model to work best when using the daily sampling frequency. Best results are reported for the shortest horizons (10 and 20 days). Under the Bonferroni and false discovery rate criteria, we found the GARCH MSE's to be never significantly lower than the benchmark MSE's, for none of the forecasting horizons and none of the extra variables added. Under the economic criterion, the GARCH model beats the benchmark at the 10 and 20 days horizon.

Under statistical criteria, the HAR model gives best results when using the weekly sampling frequency. The model has highest predictive power for the 60 days horizon. However, we also found the HAR MSE's to be never significantly lower than the benchmark MSE's. Based on the hitratios there is no superior sampling frequency for the HAR model. Best results are found for the shortest horizons. For all horizons except the 120 days horizon, the HAR model leads to higher hitratios than the benchmark model, when certain macroeconomic variables are added to the model. The overall highest hitratios are found when the credit spread is added.

We tried to improve the HAR model's results by smartly choosing the regressors by means of forward stepwise regression and boosting. Both of these regression techniques lead to a considerable increase in R^2's and decrease in MSE with respect to both the simple HAR model and the benchmark model, for all forecasting horizons. Wilcoxon signed rank test statistics are significant at the lowest horizons. At the 10 days horizon, the forward stepwise regression model's MSE is significantly lower than the corresponding benchmark MSE even under the false discovery rate criterion. For the boosting model this is the case for the 10, 20 and 40 days horizons, providing very strong evidence that the boosting model outperforms the benchmark at the lowest horizons. However, also at the longest horizons the boosting model does better than the benchmark model. Our final recommandation is to use the boosting model for further volatility forecasting purposes.

Appendix

Table A1a					
GARCH R^2's			horizon		
Bare model	10	20	40	60	120
daily	0.458	0.512	0.507	0.486	0.420
weekly	0.344	0.381	0.374	0.371	0.376
two-weekly	0.262	0.299	0.288	0.266	0.227
monthly		0.063	0.091	0.079	0.105
Added: range					
daily	0.437	0.485	0.489	0.472	0.407
weekly	0.326	0.362	0.338	0.333	0.342
two-weekly	0.261	0.260	0.266	0.260	0.253
monthly		0.109	0.167	0.161	0.153
Added: credit spread					
daily	0.454	0.505	0.503	0.483	0.419
weekly	0.320	0.363	0.364	0.362	0.367
two-weekly	0.293	0.324	0.304	0.285	0.242
monthly		0.086	0.106	0.113	0.129
Added: dividend yield					
daily	0.459	0.514	0.508	0.487	0.419
weekly	0.329	0.365	0.361	0.357	0.358
two-weekly	0.269	0.293	0.275	0.259	0.221
monthly		0.115	0.145	0.138	0.149
Added: 3 months treasury bill yield					
daily	0.456	0.509	0.505	0.483	0.417
weekly	0.325	0.362	0.358	0.353	0.351
two-weekly	0.310	0.345	0.313	0.290	0.239
monthly		0.130	0.138	0.138	0.165
Added: term spread					
daily	0.460	0.514	0.509	0.487	0.421
weekly	0.311	0.358	0.375	0.368	0.365
two-weekly	0.316	0.359	0.320	0.297	0.249
monthly		0.081	0.118	0.131	0.148
Added: 6m com.paper - 3m treasury bill yield					
daily	0.458	0.510	0.507	0.486	0.420
weekly	0.335	0.376	0.361	0.356	0.360
two-weekly	0.289	0.337	0.310	0.289	0.239
monthly		0.102	0.155	0.154	0.190

Table A1b					
HAR R^2's			horizon		
Bare model	10	20	40	60	120
daily	0.365	0.413	0.371	0.328	0.239
weekly	0.419	0.462	0.459	0.441	0.372
two-weekly	0.412	0.418	0.437	0.422	0.357
monthly		0.332	0.349	0.311	0.254
Added: range					
daily	0.412	0.486	0.468	0.459	0.423
weekly	0.427	0.473	0.503	0.516	0.507
two-weekly	0.393	0.429	0.440	0.433	0.409
monthly		0.310	0.400	0.379	0.404
Added: credit spread					
daily	0.396	0.435	0.451	0.451	0.455
weekly	0.418	0.472	0.490	0.488	0.472
two-weekly	0.416	0.434	0.468	0.469	0.459
monthly		0.373	0.419	0.413	0.408
Added: dividend yield					
daily	0.414	0.443	0.422	0.397	0.335
weekly	0.428	0.473	0.491	0.487	0.448
two-weekly	0.436	0.456	0.498	0.496	0.463
monthly		0.349	0.371	0.355	0.347
Added: 3 months treasury bill yield					
daily	0.371	0.413	0.373	0.335	0.260
weekly	0.408	0.442	0.423	0.395	0.319
two-weekly	0.401	0.401	0.407	0.381	0.307
monthly		0.323	0.332	0.296	0.243
Added: term spread					
daily	0.362	0.399	0.347	0.297	0.196
weekly	0.414	0.453	0.435	0.405	0.313
two-weekly	0.402	0.400	0.400	0.372	0.291
monthly		0.306	0.313	0.276	0.224
Added: 6m com.paper - 3m treasury bill yield					
daily	0.381	0.421	0.388	0.351	0.278
weekly	0.405	0.439	0.418	0.391	0.313
two-weekly	0.399	0.398	0.404	0.381	0.306
monthly		0.319	0.324	0.290	0.228

Tables A1a-b: Regression R^2's obtained from the regressions of the ex-post RV series on the forecast series generated by the GARCH (A1a) and HAR (A1b) models: $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the RV series and $\hat{\sigma}$ the forecast series.

Table A2a						
GARCH				horizon		
Bare model		10	20	40	60	120
daily	a	0.283	0.259	0.647	1.071	2.602
		(0.074)	(0.387)	(0.116)	(0.035)	(0.000)
	b	0.852	0.904	0.864	0.828	0.725
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.010	0.103	0.562	0.933	1.903
		(0.963)	(0.795)	(0.297)	(0.148)	(0.027)
	b	1.007	0.998	0.935	0.903	0.850
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.903	1.220	2.142	3.003	5.248
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	b	0.780	0.804	0.724	0.662	0.553
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.987	4.068	5.213	7.096
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.350	0.400	0.364	0.405
			(0.001)	(0.000)	(0.000)	(0.000)

Table A2b						
GARCH				horizon		
Range		10	20	40	60	120
daily	a	0.362	0.407	0.801	1.221	2.766
		(0.024)	(0.181)	(0.054)	(0.017)	(0.000)
	b	0.829	0.873	0.843	0.812	0.713
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.039	0.079	0.675	1.053	1.899
		(0.861)	(0.850)	(0.236)	(0.125)	(0.039)
	b	1.004	1.010	0.922	0.892	0.856
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.854	1.458	2.320	3.034	4.762
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	b	0.812	0.753	0.710	0.682	0.635
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.570	3.214	4.070	6.088
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.473	0.573	0.551	0.514
			(0.000)	(0.000)	(0.000)	(0.000)

Table A2c						
GARCH				horizon		
Credit spread		10	20	40	60	120
daily	a	0.303	0.306	0.699	1.125	2.653
		(0.057)	(0.308)	(0.089)	(0.026)	(0.000)
	b	0.846	0.894	0.857	0.822	0.721
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.104	0.234	0.683	1.088	2.190
		(0.637)	(0.560)	(0.204)	(0.090)	(0.010)
	b	0.997	0.970	0.918	0.886	0.827
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.816	1.184	2.139	2.978	5.249
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	b	0.819	0.823	0.732	0.674	0.560
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.715	3.722	4.425	6.104
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.421	0.458	0.482	0.502
			(0.000)	(0.000)	(0.000)	(0.000)

Table A2d						
GARCH				horizon		
Dividend yield		10	20	40	60	120
daily	a	0.284	0.243	0.631	1.045	2.575
		(0.073)	(0.416)	(0.126)	(0.040)	(0.001)
	b	0.852	0.908	0.867	0.831	0.727
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.042	0.166	0.527	0.893	1.867
		(0.850)	(0.683)	(0.344)	(0.180)	(0.037)
	b	0.997	0.985	0.939	0.906	0.849
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.841	1.257	2.249	3.074	5.302
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	b	0.813	0.800	0.708	0.657	0.552
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.673	3.680	4.605	6.419
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.444	0.481	0.467	0.481
			(0.000)	(0.000)	(0.000)	(0.000)

Table A2e						
GARCH				horizon		
3 months						
treasury bill yield		10	20	40	60	120
daily	a	0.291	0.273	0.662	1.091	2.632
		(0.067)	(0.363)	(0.108)	(0.032)	(0.000)
	b	0.850	0.901	0.862	0.825	0.722
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.066	0.168	0.540	0.927	1.970
		(0.766)	(0.681)	(0.333)	(0.166)	(0.028)
	b	0.988	0.982	0.934	0.899	0.837
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.805	1.118	2.107	2.948	5.245
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	b	0.819	0.836	0.735	0.676	0.560
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.529	3.598	4.459	6.114
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.494	0.500	0.494	0.521
			(0.000)	(0.000)	(0.000)	(0.000)

Table A2f						
GARCH				horizon		
Term spread		10	20	40	60	120
daily	a	0.289	0.272	0.667	1.094	2.625
		(0.067)	(0.360)	(0.103)	(0.030)	(0.000)
	b	0.851	0.902	0.862	0.825	0.723
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.169	0.261	0.476	0.856	1.895
		(0.443)	(0.517)	(0.383)	(0.191)	(0.031)
	b	0.958	0.967	0.943	0.906	0.841
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.779	1.115	2.127	2.959	5.220
		(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
	b	0.840	0.842	0.736	0.679	0.566
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.897	3.861	3.768	6.501
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.380	0.446	0.464	0.475
			(0.000)	(0.000)	(0.000)	(0.000)

Table A2g						
GARCH				horizon		
6 months commercial paper						
-3m treasury bill yield		10	20	40	60	120
daily	a	0.291	0.270	0.652	1.074	2.596
		(0.066)	(0.368)	(0.113)	(0.034)	(0.000)
	b	0.851	0.903	0.865	0.829	0.726
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.003	0.066	0.675	1.091	2.125
		(0.990)	(0.871)	(0.213)	(0.093)	(0.014)
	b	1.010	1.010	0.920	0.885	0.831
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.827	1.089	2.038	2.866	5.161
		(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
	b	0.817	0.844	0.748	0.689	0.568
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		2.665	3.432	4.290	5.754
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.456	0.543	0.529	0.569
			(0.000)	(0.000)	(0.000)	(0.000)

Tables A2a-g: Regression coefficients of the regression $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the ex-post RV series and $\hat{\sigma}$ the forecast series generated by the GARCH model. In brackets the p-values from the t-test for a=0 and t-test for b=0. In the upper left corner is reported which explanatory variable is added.

Table A3a						
HAR				horizon		
Bare model		10	20	40	60	120
daily	a	0.229	-0.181	0.200	0.798	2.987
		(0.235)	(0.650)	(0.728)	(0.267)	(0.004)
	b	0.858	0.996	0.926	0.853	0.684
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.161	0.084	0.420	0.776	2.227
		(0.364)	(0.806)	(0.369)	(0.177)	(0.008)
	b	0.891	0.944	0.904	0.872	0.769
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.193	0.404	0.701	1.086	2.645
		(0.279)	(0.243)	(0.132)	(0.057)	(0.001)
	b	0.882	0.868	0.855	0.829	0.727
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		0.875	1.298	2.074	4.057
			(0.017)	(0.009)	(0.001)	(0.000)
	b		0.759	0.758	0.697	0.591
			(0.000)	(0.000)	(0.000)	(0.000)

Table A3b						
HAR				horizon		
Range		10	20	40	60	120
daily	a	0.297	0.096	0.529	0.890	2.193
		(0.084)	(0.768)	(0.242)	(0.104)	(0.004)
	b	0.840	0.935	0.878	0.849	0.769
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.390	0.563	1.055	1.511	2.921
		(0.016)	(0.061)	(0.007)	(0.001)	(0.000)
	b	0.822	0.814	0.778	0.750	0.674
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.388	0.622	1.175	1.692	3.289
		(0.025)	(0.053)	(0.006)	(0.001)	(0.000)
	b	0.822	0.814	0.778	0.750	0.674
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		1.206	1.358	1.992	3.125
			(0.001)	(0.002)	(0.000)	(0.000)
	b		0.701	0.772	0.735	0.713
			(0.000)	(0.000)	(0.000)	(0.000)

Table A3c						
HAR				horizon		
Credit spread		10	20	40	60	120
daily	a	0.799	1.140	1.884	2.540	4.161
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	b	0.727	0.757	0.728	0.703	0.659
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.621	0.844	1.480	2.050	3.644
		(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
	b	0.772	0.804	0.769	0.744	0.684
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.640	1.087	1.714	2.286	3.888
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	b	0.766	0.743	0.727	0.708	0.657
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		1.118	1.536	2.108	3.502
			(0.000)	(0.000)	(0.000)	(0.000)
	b		0.724	0.742	0.719	0.676
			(0.000)	(0.000)	(0.000)	(0.000)

Table A3d						
HAR				horizon		
Dividend yield		10	20	40	60	120
daily	a	0.442	0.652	1.375	2.053	4.049
		(0.007)	(0.036)	(0.001)	(0.000)	(0.000)
	b	0.751	0.753	0.681	0.633	0.535
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.411	0.592	1.111	1.582	3.061
		(0.011)	(0.048)	(0.005)	(0.001)	(0.000)
	b	0.776	0.784	0.745	0.717	0.644
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.341	0.636	0.999	1.435	2.811
		(0.035)	(0.037)	(0.012)	(0.003)	(0.000)
	b	0.797	0.775	0.763	0.737	0.668
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		1.091	1.700	2.392	3.732
			(0.001)	(0.000)	(0.000)	(0.000)
	b		0.671	0.652	0.615	0.582
			(0.000)	(0.000)	(0.000)	(0.000)

Table A3e						
HAR				horizon		
3 months						
treasury bill yield		10	20	40	60	120
daily	a	0.380	0.220	0.891	1.624	3.839
		(0.036)	(0.545)	(0.081)	(0.010)	(0.000)
	b	0.813	0.907	0.820	0.752	0.615
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.233	0.245	0.787	1.353	3.292
		(0.188)	(0.475)	(0.095)	(0.020)	(0.000)
	b	0.863	0.900	0.837	0.788	0.661
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.285	0.571	1.044	1.631	3.648
		(0.106)	(0.097)	(0.025)	(0.004)	(0.000)
	b	0.851	0.827	0.797	0.753	0.629
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		0.998	1.546	2.376	4.484
			(0.006)	(0.002)	(0.000)	(0.000)
	b		0.725	0.711	0.651	0.546
			(0.000)	(0.000)	(0.000)	(0.000)

Table A3f						
HAR				horizon		
Term spread		10	20	40	60	120
daily	a	0.317	0.069	0.719	1.539	4.226
		(0.092)	(0.858)	(0.194)	(0.026)	(0.000)
	b	0.838	0.949	0.853	0.766	0.578
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.221	0.192	0.677	1.218	3.249
		(0.209)	(0.572)	(0.149)	(0.036)	(0.000)
	b	0.886	0.936	0.880	0.830	0.687
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.283	0.573	1.083	1.684	3.782
		(0.108)	(0.096)	(0.021)	(0.004)	(0.000)
	b	0.866	0.842	0.808	0.764	0.633
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		1.190	1.865	2.770	4.958
			(0.001)	(0.000)	(0.000)	(0.000)
	b		0.702	0.682	0.621	0.522
			(0.000)	(0.000)	(0.000)	(0.000)

| Table A3g						
HAR				horizon		
6 months commercial paper						
-3m treasury bill yield		10	20	40	60	120
daily	a	0.427	0.338	1.014	1.755	3.944
		(0.015)	(0.332)	(0.037)	(0.003)	(0.000)
	b	0.808	0.892	0.812	0.746	0.613
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
weekly	a	0.318	0.397	1.005	1.612	3.658
		(0.066)	(0.233)	(0.029)	(0.004)	(0.000)
	b	0.856	0.889	0.826	0.777	0.648
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
twoweekly	a	0.352	0.678	1.214	1.828	3.938
		(0.042)	(0.044)	(0.008)	(0.001)	(0.000)
	b	0.842	0.817	0.785	0.742	0.616
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
monthly	a		1.059	1.643	2.458	4.749
			(0.003)	(0.001)	(0.000)	(0.000)
	b		0.720	0.704	0.647	0.526
			(0.000)	(0.000)	(0.000)	(0.000)

Tables A3a-g: Regression coefficients of the regression $\sigma = a + b\hat{\sigma} + \epsilon$ where σ denotes the ex-post RV series and $\hat{\sigma}$ the forecast series generated by the GARCH model. In brackets the p-values from the t-test for a=0 and t-test for b=0. In the upper left corner is reported which explanatory variable is added.

benchmark vs GARCH test statistics			horizon		
Bare model	10	20	40	60	120
daily	1.792	**2.587**	-0.209	0.019	-1.834
weekly	-1.027	-1.139	-3.017	-2.411	-1.597
two-weekly	-2.752	-2.251	-4.491	-4.736	-6.713
monthly		-3.629	-4.953	-4.917	-5.204
Added: range					
daily	1.459	**2.256**	-0.841	-0.453	-2.246
weekly	-1.621	-1.306	-3.173	-3.328	-2.421
two-weekly	-3.329	-3.241	-4.465	-4.641	-5.009
monthly		-3.863	-5.393	-6.102	-6.311
Added: credit spread					
daily	1.701	**2.450**	-0.199	-0.191	-1.853
weekly	-1.640	-1.450	-2.986	-2.645	-1.913
two-weekly	-3.226	-3.112	-4.954	-5.586	-6.569
monthly		-3.511	-4.123	-5.107	-6.200
Added: dividend yield					
daily	1.926	**2.503**	-0.113	0.068	-1.789
weekly	-1.293	-1.494	-3.088	-2.653	-1.845
two-weekly	-3.197	-3.246	-5.105	-6.234	-6.338
monthly		-3.851	-4.523	-4.513	-6.230
Added: 3 months treasury bill yield					
daily	1.615	**2.598**	-0.214	-0.009	-1.879
weekly	-1.780	-1.657	-3.001	-2.969	-2.002
two-weekly	-3.003	-2.966	-5.180	-5.799	-6.755
monthly		-4.523	-5.382	-5.734	-6.498
Added: term spread					
daily	1.745	**2.704**	-0.231	0.008	-1.837
weekly	-2.091	-1.474	-3.210	-2.742	-1.695
two-weekly	-2.996	-2.891	-5.009	-5.569	-6.752
monthly		-3.698	-5.192	-2.641	-6.015
Added: 6m com.paper - 3m treasury bill yield					
daily	1.874	**2.390**	-0.344	0.126	-1.713
weekly	-1.358	-1.262	-3.125	-2.729	-1.853
two-weekly	-3.416	-3.000	-5.029	-5.697	-6.435
monthly		-4.537	-5.173	-5.915	-6.532

Table A4a: Wilcoxon signed rank test statistics, as calculated in eq.(13), to test the null hypothesis of equal distributions of the benchmark model's forecasting errors and the GARCH models' forecasting errors. Values greater than 1.96 are in bold. In this case GARCH forecasting errors are significantly lower than the benchmark errors, at the 2.5% confidence level.
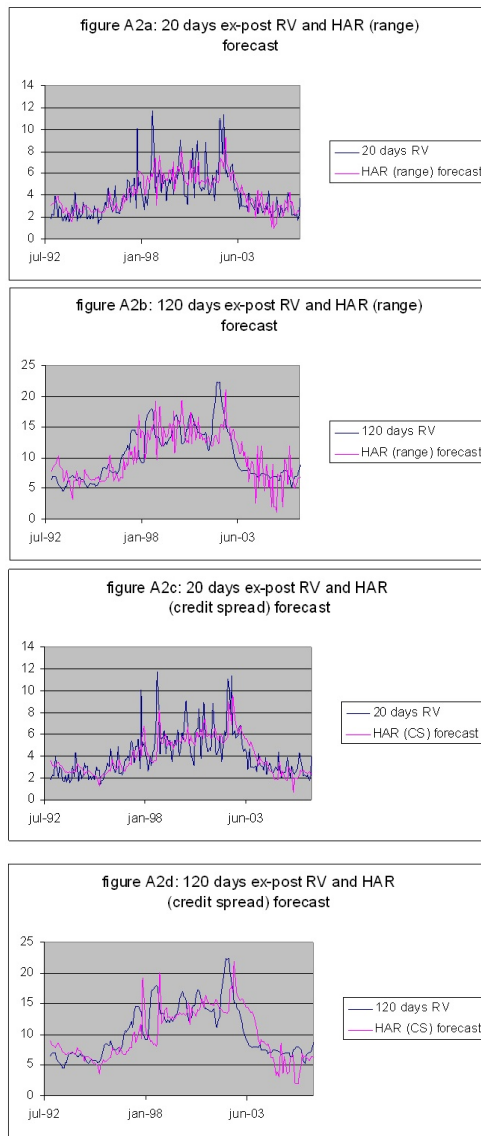
benchmark vs HAR test statistics Bare model	10	20	horizon 40	60	120
daily	-0.829	-1.007	-9.302	-3.901	-5.045
weekly	1.257	1.884	-2.049	-1.006	-2.547
two-weekly	0.679	-0.526	-2.621	-1.802	-2.190
monthly		-1.999	-3.683	-3.930	-4.911
Added: range					
daily	0.922	1.892	-0.830	-0.315	-0.932
weekly	0.847	1.500	-0.592	0.521	0.492
two-weekly	-0.348	-0.963	-2.443	-1.503	-1.527
monthly		-2.410	-2.831	-2.490	-1.257
Added: credit spread					
daily	-0.927	-0.638	-2.101	-2.141	-1.983
weekly	0.492	1.557	-1.920	-0.630	-1.283
two-weekly	-0.129	-0.462	-1.897	-1.702	-1.572
monthly		-1.627	-2.436	-2.092	-1.622
Added: dividend yield					
daily	-1.682	-2.071	-4.445	-4.345	-5.782
weekly	-0.839	-0.445	-2.323	-1.467	-3.172
two-weekly	-0.308	-1.001	-2.238	-1.131	-2.136
monthly		-2.913	-3.087	-2.901	-4.437
Added: 3 months treasury bill yield					
daily	-0.829	-0.854	-4.137	-3.746	-4.833
weekly	0.453	0.293	-2.635	-2.822	-3.625
two-weekly	-0.003	-1.245	-3.036	-2.753	-3.716
monthly		-2.635	-4.018	-4.826	-5.164
Added: term spread					
daily	-0.860	-0.837	-3.726	-3.262	-4.081
weekly	1.113	1.403	-2.173	-1.943	-2.487
two-weekly	0.735	-0.836	-2.028	-2.732	-3.254
monthly		-2.376	-4.726	-4.761	-4.992
Added: 6m com.paper - 3m treasury bill yield					
daily	-0.891	-0.816	-3.514	-3.243	-4.816
weekly	0.614	0.562	-2.382	-2.018	-3.128
two-weekly	0.104	-1.023	-2.941	-2.339	-3.182
monthly		-2.235	-4.032	-3.728	-5.112

Table A4b: Wilcoxon signed rank test statistics, as calculated in eq.(13), to test the null hypothesis of equal distributions of the benchmark model's forecasting errors and the HAR models' forecasting errors. Values greater than 1.96 are in bold. In this case HAR forecasting errors are significantly lower than the benchmark errors, at the 2.5% confidence level.
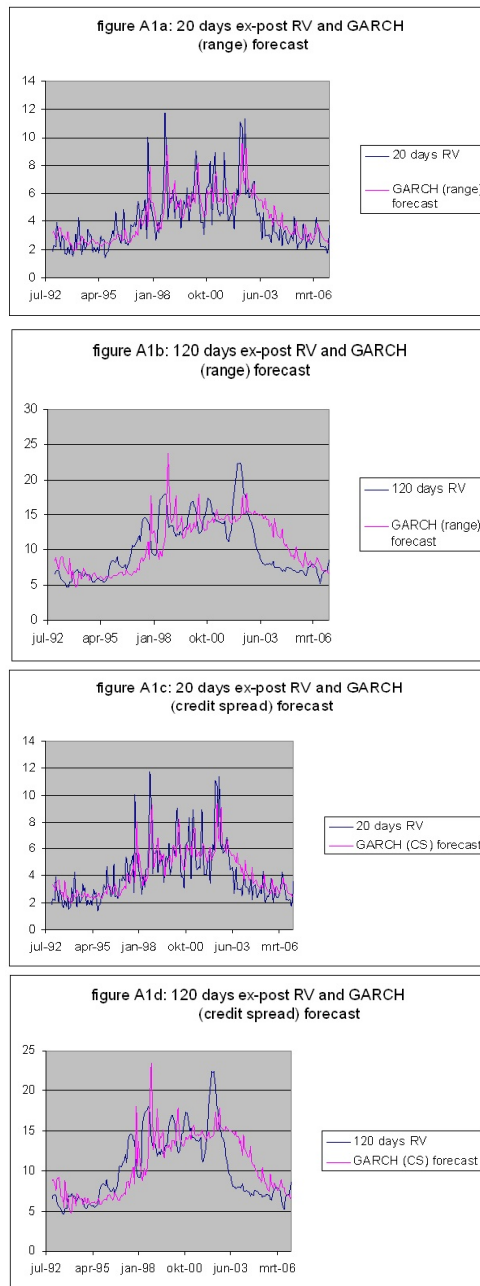
FIGURES A1A-D: **20 and 120 days RV over the period November 1992 - February 2007 and corresponding forecasts, generated by the daily sampled GARCH model with range (figures A1a-b) and credit spread (figures A1c-d) added.**

figure A2a: 20 days ex-post RV and HAR (range) forecast

figure A2b: 120 days ex-post RV and HAR (range) forecast

figure A2c: 20 days ex-post RV and HAR (credit spread) forecast

figure A2d: 120 days ex-post RV and HAR (credit spread) forecast

FIGURES A2A-D: **20 and 120 days RV over the period November 1992 - February 2007 and corresponding forecasts, generated by the weekly sampled HAR model with range (figures A2a-b) and credit spread (figures A2c-d) added.**

Bibliography

Andersen, T.G. and T. Bollerslev (1998), *Answering the sceptics:
Yes, standard volatility models do provide accurate forecasts*, International economic review,
39, 885-905

Andersen, T.G., T. Bollerslev, F.X. Diebold and P.Labys (2003),
Modeling and forecasting realized volatility, Econometrica, 71, 579-625

Andersen, T.G., T. Bollerslev and S. Lange (1999), *Forecasting financial market volatility:
sample frequency vis-a-vis forecast horizon*, Journal of empirical finance 6, 457-477

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996), *Fractionally integrated generalized
autoregressive conditional heteroskedasticity*, Journal of econometrics 74, 3-30

Bernanke, B.S. (1990), *On the predictive power of interest rates and interest rate spreads*,
New England Economic Review (Federal Reserve Bank of Boston) November-December
1990, pp. 51-68

Bollerslev, T. (1986), *Generalized autoregressive conditional heteroskedasticity*,
Journal of Econometrics 31, 307-327

Brandt, M.W. and C.S. Jones (2006),*Volatility forecasting with range-based EGARCH models*,
Journal of business and economic statistics, 24(4), 470-486

Calvet, L.E. and A.J. Fisher (2004), *How to forecast long-run volatility: regime switching and
the estimation of multifractal processes*, Journal of financial econometrics, 2(1), 49-83

Christoffersen, P. and F.X. Diebold (2000), *How relevant is volatility forecasting for financial
risk management?*, Review of Economics and Statistics, 82, 12-22

Corsi, F. (2004), *A simple long memory model of realized volatility*, working paper

Engle, R.F. (1982), *Autoregressive conditional heteroskedasticity with estimates
of the variance of UK inflation*, Econometrica 50, 987-1008

Engle, R.F., V.K. Ng (1993), *Measuring and testing the impact of news on volatility*,
Journal of finance 48, 1749-1778

Franks, J.R. and E.S. Schwartz (1991), *The stochastic behaviour of market variance implied
in the prices of index options*, The economic journal, 101

Glosten, L.R., R. Jagannathan and D. Runkle (1993), *On the relation between the expected value
and the volatility of the nominal excess returns on stocks*, Journal of finance 48, 1779-1801

Granger, C. and S.H. Poon (2002), *Forecasting volatility in financial markets: a review*,
Journal of economic literature, vol 41(2), 478-539

Hansen, P.R. and A. Lunde (2005), *A forecast comparison of volatility models: does anything beat a GARCH(1,1) model?*, Journal of applied econometrics 20(7), 873-889

Harvey, C.R. and R. Huang (1991), *Volatility in the Foreign Currency Futures Market*, Review of Financial Studies 4, 543-569

Hastie, T., R. Tibshirani and J. Friedman (2001), *The elements of statistical learning: Data mining, inference and prediction*, Springer, New York.

Kartazas, I. and S.E. Shreve (1988), *Brownian motion and stochastic calculus*, New York, Springer Verlag

LeBaron, B. (2001), *Stochastic volatility as a simple generator of financial power-laws and long memory*, Quantitative finance, 1(6), 621-631

Martens, M.P.E. (2001), *Forecasting daily exchange rate volatility using intraday returns* Journal of International Money and Finance, 20, 1-23.

Nelson, D.B. (1991), *Conditional heteroskedasticity in asset returns: a new approach*, Econometrica 59, 347-370

Rice, J.A. (1995), *Mathematical statistics and data analysis*, Duxberry Press

Schwertz, W.G. (1989), *Why does stock market volatility change over time*, Journal of finance, 44, 1115-1153

Taylor, S.J. (1987), *Forecasting the volatility of currency exchange rates*, International Journal of Forecasting 3, 159-170

West, K.D and D. Cho (1995), *The predictive ability of several models of exchange rate volatility*, Journal of econometrics 69, 367-391

Whitelaw, R.F (1994), *Time variations and covariations in the expectation and volatility of stock market returns*, Journal of finance, 44(2), 515-541