ON MODELING AND FORECASTING STOCK VOLATILITY

Heiko Ebens

A dissertation submitted to the Johns Hopkins University
in conformity with the requirements for the degree of
Doctor of Philosophy

Baltimore, Maryland

September, 2000

Abstract

Accurate measures and forecasts of volatility are indispensable for asset pricing, portfolio selection and risk management. Instrumental for these tasks have been statistical approaches such as Autoregressive Conditional Heteroskedastic (ARCH) and Stochastic Volatility models as well as economic formulations that imply future volatility estimates using the information contained in options. The resulting volatility estimates are however model driven and therefore sensitive to the particular specification chosen. Our first chapter introduces this dissertation and outlines in greater detail the limitations of the existing methods when measuring, modeling and/or forecasting volatility. In the second chapter we derive that under general conditions daily volatility can be measured model-free and to any degree of accuracy from intradaily return observations. Using the record of each transaction underlying the Dow Jones Industrials Average portfolio, we next document the empirical properties of such 'realized' volatility measure and capture its characteristics using a time-series model. On the basis of *ex ante* one-day-ahead prediction criteria we find that this specification yields unbiased and accurate volatility predictions and that these clearly improve upon the ones obtained by various ARCH models, including those that closely match the volatility regularities we document. The third chapter concerns the identification of models that provide good volatility forecast over short and long horizons *ex ante* and *ex post*. We also examine whether the very 'model-free' advantage of the employed realized volatility measure makes it inefficient as it ignores any structural dependence in the intradaily data. As this structure turns out to be quite complex, we extend traditional ARCH specifications via semi-parametric methods to model intradaily returns. We find that the realized volatility and semi-parametric specifications perform equally well and that, for various in-sample and out-of-sample horizons, both of them yield far better forecasts than the ones that are obtained using numerous daily ARCH models. The last chapter concludes the dissertation with recommendations for future research.

Advisors: Pedro de Lima and Tim Bollerslev

Readers : Burt Barnow, Lance James, Joel Spruck, and Geert Rider

Acknowledgements

I wish to tank my primary advisor Pedro de Lima for his help, efforts and thoughtful guidance. My secondary advisor, Tim Bollerslev, is deeply appreciated for his encouragement and support. Special thanks also to Joe Gruber, Gonçalo Fonseca, Daniel Hinze and Junko Nakaya – without their help my efforts would have been far less fruitful. Finally, I am grateful to the Johns Hopkins Library for acquiring the intradaily data that provide the basis for the empirical analyses.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Asset return volatility is indispensable for the theory and practice of finance. Pricing models dictate that investors require higher returns for volatile securities. Portfolio strategies rest on the notion that the return on a large number of assets fluctuates less than the average portfolio component return. An option contract is more valuable when the price of the underlying asset becomes erratic. Moreover, financial market volatility is an important economic risk factor. The Asian Currency Crises had adverse macroeconomic implications for several countries throughout the world and billions of dollars have been spent to bailout Long Term Capital Management with the purpose of maintaining financial and economic stability.

It becomes evident that asset return volatility is crucially important to investors and of substantial concern to policy makers. However, as volatility is not a directly observable variable, our understanding of it is limited, the utility of financial models is compromised, and regulatory policy is difficult to formulate.

We can identify three distinct ways in the literature that attempt to cope with the measurement problem. The first approach obtains mode-free volatility estimates directly from return observations, while the second and third approach consider volatility as an unobserved variable that can be recovered using either a statistical or an economic model.

The direct measurement method consists of calculating the sample return variance from daily data. This approach gives model-free volatility estimates, but requires an important tradeoff. If one chooses a large number of daily returns to calculate a variance observation, the resulting statistic does not contain any information about the behavior of volatility within the used timeframe. If, on the other hand, a small number of returns are used, the variance estimator is subject to great measurement error – at the extreme, only one return observation is used for a daily variance estimate.

The second measurement approach treats volatility as an unobservable variable that is to be obtained by a statistical model filtering the data. Models of the Autoregressive Conditionally Heteroskedastic (ARCH) and Stochastic Volatility class fall in this category. Both types of specifications have in common that the return process is assumed a product of two components, standard deviations and noise. The ARCH approach rests on the notion that the volatility component can be identified by making it a function of past returns and noise innovations. The closely related Stochastic Volatility models are more general in that the volatility process is additionally allowed to be a function of past innovations that come from a volatility specific noise term. In either case, the proper specification of the functional form that relates volatility to past information is subject to very active research and no single model has emerged that is viewed as best. Naturally, competing models provide different estimates.

The third approach to measuring volatility consists of backing out estimates from option pricing models. The celebrated Black-Scholes formula is most frequently used in this context. This pricing model holds under an array of technical assumptions, including no market imperfections, unlimited borrowing and lending at a risk-free rate and no arbitrage. From the observed call and stock price, the risk-free rate and dividend payments, a return variance estimate for the underlying asset over the duration of the option contract can be traced out. Unlike the approaches discussed above, the 'implied' volatility has therefore a forward-looking character and thus may be regarded as a volatility forecast implied by option market participants. This volatility measurement approach has however several disadvantages, among those are: estimates are only available for the limited number of asset on which option contracts can be written, the option implied measures do not contain any information about the volatility process within the time to expiration of contracts and, finally, the volatility estimates are only valid if the assumptions of the employed pricing formula hold.

Given the difficulties with the aforementioned methods, the approach taken in this thesis is to measure daily volatility from the sample variance of intraday return data that have recently become available. Specifically, we use the transaction record of the Dow Jones Industrials Average (DJIA) portfolio to obtain a time series of 'realized' daily variance estimates. These are free of the assumptions necessary when the statistical or economic approaches are employed. Moreover, as we have an (almost) continuous record of returns for each day, we can calculate interdaily variances with negligible error. Using our methods, volatility thus becomes an observed variable that can be analyzed using conventional time series techniques.

The second chapter of this dissertation, entitled *"Realized Stock Volatility"*, first gives a formal justification of the employed volatility measure and derives its theoretical properties. Using our data for the DJIA, we next document the empirical regularities of this volatility variable. These findings set the stage for the development of a time series model that captures the temporal dependency of realized volatility. Finally, we compare the predictive one-step-ahead ability of such specification to various ARCH formulations fitted to daily data in order to investigate whether this new approach is of practical relevance.

The third chapter, entitled *"Forecasting Stock Volatility"*, concerns the identification of models that provide good out-of-sample predictions of asset return volatility at alternative forecast horizons. Next to forecasting volatility over multiple days using daily ARCH and realized volatility specifications, we also consider a third approach, specifically, forecasting volatility over multi-day horizons using ARCH models fitted directly to intradaily returns. If such model imposes the correct structure on the return process, it should yield superior forecasts. However, moving to higher frequency data, we face the problem of intraday seasonality, *i.e.* that volatility changes systematically within the day. We explicitly address this problem by extending traditional ARCH models via nonparametric methods.

The fourth chapter concludes this dissertation with a summary of our main findings. We additionally formulate several recommendations for future research.

Chapter 2

Realized Stock Volatility

2.1 Introduction

Financial market volatility is indispensable for asset and derivative pricing, asset allocation, and risk management. As volatility is not a directly observable variable, large research areas have emerged that attempt to best address this problem. By far the most popular approach is to obtain volatility estimates using the statistical models that have been proposed in the ARCH and Stochastic Volatility literature. Another method of extracting information about volatility is to formulate and apply economic models that link the information contained in options to the volatility of the underlying asset. All these approaches have in common that the resulting volatility measures are only valid under the specific assumptions of the models used and it is generally uncertain which or whether any of these specifications provide a good description of actual volatility.

A model-free measure of volatility is the sample variance of returns. Using daily data, for instance, it may be freely estimated using returns spanning over any number of days and, as such, one can construct a time series of model-free variance estimates. When one chooses the observation frequency of this series, an important trade-off has to be made, however. When the variances are calculated using a large number of observations (*e.g.* the returns over an entire year), many interesting properties of volatility tend to disappear (the volatility clustering and leverage effect, for instance). On the other hand, if only very few observations are used, the measures are subject to great error. At the extreme, only one return observation is used for each daily variance estimate.

The approach taken in this dissertation is to calculate the daily volatility from the sample variance of intraday returns, the 'realized' volatility. Specifically, we use the transaction record of the Dow Jones Industrials Average (DJIA) portfolio over the period extending from January 1993 to May 1998, to obtain a time series of 1366 daily realized variances. These are free of the assumptions necessary when the statistical or economic approaches are employed and, as we have an (almost) continuous record of returns for each day, we can calculate the interdaily variances with little or perhaps negligible error.

In this chapter, we shall first give a through account of the theoretical properties that underlie the concept of realized volatility measurement. Using our data for the DJIA, we next document the empirical regularities of this volatility variable and then capture these using a parametric model. Finally, we compare the predictive ability of the realized volatility model to various ARCH formulations.

Almost all of the work on daily volatility is within the confines of ARCH and Stochastic Volatility models or derivative pricing formulas. There are exceptions, however. Schwert (1990) and Hsieh (1991) have computed sample standard deviations from intradaily returns on the S&P 500 index. However, the modeling and investigation of the properties of volatility have not been the major focus and consequently these two papers do not present a thorough analysis of the constructed series. More recently, Andersen and Bollerslev (1998) have calculated a time series of realized exchange rate variances to evaluate one-day-ahead GARCH model forecasts while Andersen, Bollerslev, Diebold and Labys (1999) use realized variance estimates to document the properties of daily exchange rate volatility. Our study is in spirit close to the latter paper, but distinct in two key aspects. Firstly, our analysis is on stock return volatility and as a result we characterize important empirical regularities not found for exchange rates. Secondly, we not only examine but also model realized volatility and determine whether this new approach is of practical relevance.

Following this introduction, Section 2.2 gives an account of the theoretical underpinnings of the realized volatility measure. Section 2.3 details the construction of the data that provide the basis for our subsequent empirical analysis. In Section 2.4 we investigate the properties of stock return volatility and, in Section 2.5, we fit parametric models to our volatility series. We compare the performance of these models to four ARCH formulations in Section 2.6. We finish in Section 2.7 with concluding remarks.

2.2 Realized Volatility Measurement

A common model-free indicator of volatility is the daily squared return. In this paper we measure interdaily volatility using intradaily high-frequency returns. We highlight in this section the relation between these two measures and discuss their individual properties.

To set forth the notation, let $p_{n,t}$ denote the time $n \geq 0$ logarithmic price at day t. The discretely observed time series of continuously compounded returns with N observations per day is then defined by:

$$r_{n,t} = p_{n,t} - p_{n-1,t}$$

where $n = 1, \ldots, N$ and $t = 1, \ldots, T$. If $N = 1$, for any series we ignore the first subscript n and thus r_t denotes the time series of daily return.

We shall assume that:

A.1: $E[\, r_{n,t} \,] = 0$

A.2: $E[\, r_{n,t} \, r_{m,s} \,] = 0 \quad \forall\, n, m, s, t$ but not $n = m$ and $s = t$

A.3: $E[\, r_{n,t}^2 \, r_{m,s}^2 \,] < \infty \quad \forall\, n, m, s, t$

Hence, returns are assumed to have mean zero and to be uncorrelated and it is assumed that the variance and covariances of squared returns exist and are finite.

The continuously compounded daily squared returns may be decomposed as:

$$r_t^2 = \left(\sum_{n=1}^{N} r_{n,t} \right)^2 = \sum_{n=1}^{N} r_{n,t}^2 + \sum_{\substack{n=1 \\ n \neq m}}^{N} \sum_{m=1}^{N} r_{n,t}\, r_{m,t} = \sum_{n=1}^{N} r_{n,t}^2 + 2 \sum_{n=1}^{N} \sum_{m=n+1}^{N} r_{n,t}\, r_{m-n,t} \quad (2.2.1)$$

Assuming that A.1 holds, the squared daily return is therefore the sum of two components: the sample variance (at the daily unit) and twice the sum of $N-1$ sample autocovariances (at the $1/N$th day interval unit). In this decomposition it is the sample variance that is of interest – the sample autocovariances are measurement error and induce noise in the daily squared return measure.

From (2.2.1) and A.1 and A.2 it therefore follows that an unbiased estimator of the daily return volatility is the sum of intraday squared returns, the realized volatility:

$$s_t^2 = \sum_{i=1}^{N} r_{n,t}^2$$

as:

$$E\left[s_t^2\right] = \sigma_t^2$$

where σ_t^2 is daily population variance.

Because the realized volatility s_t^2 is an estimator, it has itself a variance which can be interpreted as measurement error. From now on we shall assume that A.1 to A.3 hold, and then the variance of s_t^2 is given by:

$$\begin{aligned} V(s_t^2) &= E\left[\left(\sum_{n=1}^{N} r_{n,t}^2 - \sigma_t^2 \right)^2 \right] \\ &= E\left[\sum_{n=1}^{N} \sum_{m=1}^{N} \left(r_{n,t}^2 - \frac{\sigma_t^2}{N} \right) \left(r_{m,t}^2 - \frac{\sigma_t^2}{N} \right) \right] \end{aligned}$$

Thus the variance of s_t^2 depends on the sum of all covariances of the squared return process. Upon separating the double sum for all $n \neq m$, taking expectations and rearranging terms it follows:

$$= E\left[\sum_{n=1}^{N}\left(r_{n,t}^2 - \frac{\sigma_t^2}{N}\right)^2\right] + 2\,E\left[\sum_{n=1}^{N}\sum_{m=n+1}^{N}\left(r_{n,t}^2 - \frac{\sigma_t^2}{N}\right)\left(r_{m,t}^2 - \frac{\sigma_t^2}{N}\right)\right]$$

The first term is the variance of the intraday squared returns process (at the daily unit) and the second term is the sum of all squared return autocovariances (at the $1/N$th day interval unit). Upon dividing the term on the right by 1 over N times the expression on the left and taking expectations one obtains:

$$= E\left[\sum_{n=1}^{N}(r_{n,t}^2 - \frac{\sigma_t^2}{N})^2\right]\left(1 + 2\sum_{n=1}^{N}\frac{N-n}{N}\rho_{N,n,t}\right)$$

where $\rho_{N,n,t}$ the nth autocorrelation of $\{r_{n,t}^2\}_1^N$. Finally, after expanding the factor on the left and taking expectations it follows:

$$V(s_t^2) = \frac{\sigma_t^4}{N}\left(K_{N,t} - 1\right)\left(1 + 2\sum_{n=1}^{N-1}\frac{N-n}{N}\rho_{N,n,t}\right) \tag{2.2.2}$$

where $K_{N,t}$ denotes the kurtosis of $\{r_{n,t}\}_1^N$. Note that the kurtosis and autocorrelations have subscript N as these may change with the number of intraday returns. From (2.2.2) follows that for any particular value of N, measurement error increases with the daily population variance, with the kurtosis of intraday returns and with the autocorrelations of intraday squared returns.

Special cases of equation 2.2.2 reduce to familiar expressions. For instance, if $r_{n,t}$ is *i.i.d.* normal with $E[\,r_{n,t}^2\,] = \sigma_t^2/N$ (variances are constant within the day), equation (2.2.2) becomes: $V[\,s_t^2\,] = 2\,\sigma_t^4\,/\,N$. This result can be found in Kendall and Stuart (1963, p. 243), for instance. Note that under these assumptions the variance of the realized volatility decreases at rate N. However, for various assets it is well documented that returns have kurtosis in excess of three and that the squares of returns are correlated (the ARCH effect). Under these circumstances, this expression will therefore give the lower bound of measurement error.

To establish consistency of s_t^2, we require the two additional assumptions that:

$$\text{A.4:} \qquad K_{N,t} < \infty \quad \forall\, N$$

$$\text{A.5:} \qquad \exists\, \rho_{N,n,t} \;\text{ s.t }\; \rho_{N,n,t} < 1$$

Boundedness of $K_{N,t}$ rules out jump-diffusions (Drost, Nijman and Werker 1998) and implies continuity of the sample paths of $\sigma^2_{n,t}$ by the Kolmogorov criterion (Revuz and Yor 1991, Theorem I.1.8). Assumption A.5 states that the squared return process has at least one autocorrelation that is less than unity.

Suppose $\rho_{N,n,t} = 1$ for $n = 1, \dots, N$, then the last factor in (2.2.2) becomes: $\big(1 + 2\,(N-1) - 2\,N^{-1} \sum_{n=1}^{N-1} n\big) = N$, since $\sum_{n=1}^{N-1} n = 0.5\,(N-1)\,N$. Therefore, $V(s_t^2) = \sigma_t^4\,(K_{N,t} - 1)$. By A.5, however, $V(s_t^2)$ will decrease in N and by A.4 it follows therefore that:

$$\lim_{N\to\infty} V[\,s_t^2\,] = 0$$

Thus, the realized volatility measure converges in mean-square and is consistent.[1] The daily variance may therefore be estimated to any desired degree of accuracy by the realized volatility.

Recall that the results reported thus far are derived under the assumption that returns are uncorrelated. This assumption is questionable when N is large, as serial correlation in returns is a common symptom of market micro-structure effects such as price discreteness,

[1] Consistency may alternatively be established under the assumption that the price process $p_{n,t}$ follows $dp_{n,t} = \sigma_{n,t}\, dW_{n,t}$, where $W_{n,t}$ denotes a Wiener process. Under the assumption that $\sigma_{n,t}$ is continuous, it follows from the results in Karatzas and Shreve (1988, Chapter 1.5) or Barndorff-Nielsen and Shephard (1999) that $\mathrm{plim}_{N\to\infty} \sum_{i=1}^{N} r^2_{n,t} = \int_1^\infty \sigma^2_{n,t}\, dn = \sigma_t^2$. See also Andersen *et al.* (1999) for a thorough treatment along these lines in the context of special semi-martingales.

bid-ask bounces and non-synchronous trading (see for instance the textbook treatment by Campbell, Lo and MacKinlay 1997; Chapter 3). Any violation of this assumption can easily studied when considering the MA (q) (moving average) representation of $r_{n,t}$:

$$r_{n,t} = \epsilon_{n,t} + \sum_{i=1}^{q} \psi_{i,t}\, \epsilon_{n-i,t} \tag{2.2.3}$$

where the innovations $\epsilon_{n,t}$ are assumed to be uncorrelated across all leads and lags. Note that we allow the moving average representation to change across t. This simply reflects that our realized volatility measure does not require processes to remain constant over time. Upon squaring (2.2.3), taking expectations, and summing over $n = 1, \ldots, N$, it follows that:

$$E\left[\sum_{n=1}^{N} r_{n,t}^2\right] = E\left[\, s_t^2 \,\right] = (1 + \sum_{1}^{q} \psi_{i,t}^2)\, E\left[\sum_{n=1}^{N} \epsilon_{n,t}^2\right] \tag{2.2.4}$$

where $E[\sum_{n=1}^{N} \epsilon_{n,t}^2] = \sigma_t^2$. At day t therefore, the cumulative squared returns measure has a multiplicative bias that is given by the squared dynamic coefficients of the moving average representation. Under conditions of serial correlation, the realized variance will therefore unambiguously overestimate actual volatility. One may, of course, test for the statistical significance of the parameters that are used to capture any temporal dependence in returns and use (2.2.4) to determine whether any bias is economically important.

2.3 Data Source and Construction

Our empirical analysis is based on data from the NYSE Transaction and Quote (TAQ) database which records all trades and quotations for the securities listed on the NYSE, AMEX, NASDAQ, and the regional exchanges (Boston, Cincinnati, Midwest, Pacific, Philadelphia, Instinet, and CBOE). Our sample consists of the Dow Jones Industrials Average (DJIA) index constructed from the transaction prices of the 30 stocks that are contained in it.[2] The data span from January 4, 1993 to May 29, 1998 (1366 observations). Within

[2]The reconstruction of the index is straightforward. The DJIA is the sum of all prices adjusted by a devisor that changes once an index stock splits, pays a stock dividend of more than 10% or when one

each day, we consider the transaction record extending from 9:30 to 16:05, the time when the number of trades noticeably dropped. Next to transaction prices, volume and time (rounded to the nearest second) TAQ records various codes describing each trade. We used this information to filter trades that were recorded in error and out of time sequence.[3]

Taking all 30 stocks together, we observe a trade about every one second. Naturally, the trading frequency of the index components is lower. It also varies greatly across stocks: the median time between trades in a single stock ranges from a low of 7 seconds to a high of 54 seconds. This suggests that one should worry that non-synchronous trading induces serial correlation in the returns process which, in turn, would render the cumulative squared returns measure biased. Since we are focusing on an index, the market micro-structure effects that are due to price discreteness and bid-ask bounces are of less concern as these tend to wash out in the aggregate (see Gottlieb and Kalay 1985, Ball 1988 and Harris 1990 for instance).[4]

To mitigate the problem of bias, following Andersen and Bollerslev (1998) and Andersen *et al.* (1999), we shall rely on five-minute returns to obtain daily variance estimates. These are constructed from the logarithmic difference between the prices recorded at or immediately before the five-minute marks. When considering the transactions record extending from 9:30 to 16:05, this provides us with $N = 79$ returns for each of the $T = 1366$ days.

---

company in the group of thirty is replaced by another. Over our sample, the composition of the DJIA index changed March 17, 1997, when four stocks were replaced. Naturally, we accounted for this.

[3]Specifically, we omitted all trades that carry the 'correction indicators' 2 (symbol correction; out of time sequence), 7 (trade canceled due to error), 8 (trade cancelled) and 9 (trade canceled due to symbol correction). Moreover, we filtered all trades with the 'condition indicator' G (bunched sold; a bunched trade not reported within 90 seconds of execution time), L (sold last; a transaction that occur in sequence but is reported to the tape at a later time) and Z (sold sale; a transaction that is reported to the tape at a time later than it occurred and when other trades occurred between the time of the transaction and its report time). We refer to corresponding data manual for a more complete description of these and other codes.

[4]The theoretical literature on price discreteness suggests that the upward bias of the volatility estimate decreases with the price level. This suggests that one can detect the presence of bias by examining whether a time series of volatility estimates is negatively correlated with the corresponding prices. For our data we find a slight positive correlation – suggesting therefore that the discreteness of prices should not be of concern.

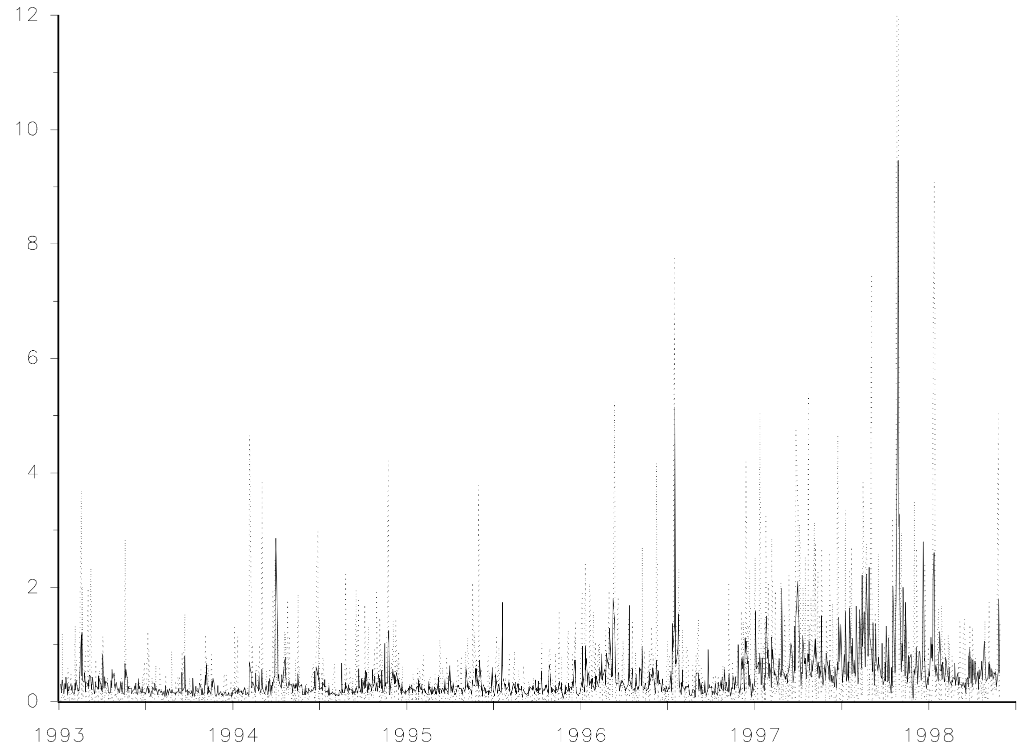
It remains an empirical question whether the five-minute cut-off is sufficient large enough so that the problem of bias due to market micro-structure effects is of no practical concern. For our data we find that the first two sample autocorrelations are 0.080 and -0.018 and these are significant judged by the $\pm 1.96\sqrt{(1/NT)}$ 5% confidence interval. Consistent with the spurious dependencies that would be induced in an index by non-synchronous trading, the first order autocorrelation is positive. The consequences of serial correlation are minimal, however. Upon estimating the MA (2) model defined by equation (2.2.3), we obtain $\hat{\psi}_1 = 0.0431$ and $\hat{\psi}_2 = -0.0317$. Form (2.2.4) it follows that the bias resulting from serial correlation scales the volatility estimates upward by a factor of only 1.0029. Considering that the mean realized variance equals 0.4166, we view this bias too small to be of any economic significance – no correction is therefore made.

The reduction in measurement error when using intradaily data to calculate volatility measures becomes readily apparent in Figure 2.1. The solid line plots realized variances, *i.e.* cumulative 5-minute squared returns, while the dotted line displays squared daily returns. For the latter series the two highest values are 39.4 (October 27, 1997) and 48.5 (October 28, 1997) and these two observations are outside the plot region. Although both series are correlated, the variability in the realized volatility series is small compared to the variability in the squared returns.

2.4 Properties of Realized Volatility

The main subject of this section is to investigate the properties of realized stock return volatility. Using intradaily returns on the DJIA portfolio over the period January 4, 1993 to May 29, 1998 (1366 observations), we focus on three volatility measures: variances (s_t^2) standard deviations (s_t) and logarithmic variances ($\ln(s_t)^2$). For each of these three measures, we investigate the distribution, persistency and relation to current and lagged returns. Our findings shall set the stage for the development of realized volatility models in our next section. In the literature on ARCH and Stochastic Volatility models considerable

Figure 2.1: Squared Daily Returns and Realized Variances



The graph displays realized variances (solid line) and daily squared returns (dashed line) for the Dow Jones Industrials Average over the period January 4, 1993 to May 29, 1998 (1366 observations). Variances are obtained using cumulative 5-minute squared returns. For the squared return series the two highest values are outside the plot region (39.4 and 48.5 recorded for October 27 and 28, 1997).

interest is in the distribution of daily returns divided by their daily standard deviations. Therefore, we characterize this distribution as well using our measures of volatility. Finally, as some of our analysis overlaps with the work by Andersen *et al.* (1999) on exchange rates, we will compare our results to theirs at the end of this section.

2.4.1 Distribution of Volatility

We graph the distributions of the variance, standard deviation and logarithmic variance series in Figure 2.2.[5] The skewness ($\hat{S}$) and kurtosis ($\hat{K}$) coefficients are displayed in the top-right corner of each plot. The distributions of variances and standard deviations are

[5] Density estimates throughout this paper are based on the Gaussian kernel. The bandwidths are calculated according to equation 3.31 of Silverman (1986).

clearly non-normal – both are skewed right and leptokurtic. The square root transformation of variances to standard deviations, however, reduces the skewness estimate from 8.19 to 2.57 and the kurtosis estimate from 121.59 to 16.78. The distribution of logarithmic variances appears to be approximately normal (the normal density is displayed by dashed lines). Nonetheless, standard tests reject normality. For instance, under the null hypothesis of *i.i.d.* normality, $\hat{S}$ and $\hat{K}$ are distributed normal as well with standard errors equal to $\sqrt{6/T} = 0.066$ and $\sqrt{24/T} = 0.133$; the skewness and kurtosis estimates for logarithmic variances are several standard errors away from their hypothesized values.[6]

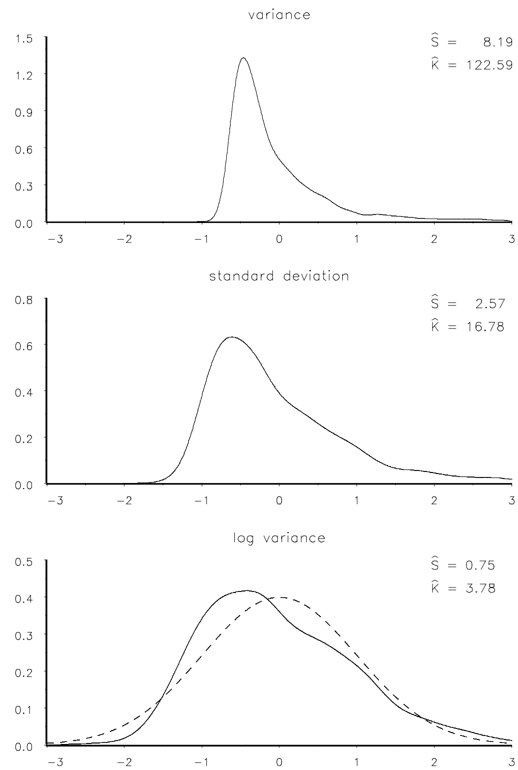
2.4.2 Persistency of Volatility

Dating back to Mandelbrot (1963) and Fama (1965), volatility clustering has a long history as a salient empirical regularity characterizing speculative returns. The literature on volatility modeling has almost universally documented that any such temporal dependency is highly persistent. The time series plot of the realized volatility in Figure 2.1 – displaying identifiable periods of high and low variances – seems to already intimate that view. The temporal dependency of volatility is reinforced by Figure 2.3, where we plot the sample autocorrelation function for each series (boxes). For all three volatility measures, the autocorrelations begin around 0.65 and decay very slowly. The 100 day correlation is 0.15, 0.28 and 0.32 for the variance, standard deviation and logarithmic variance series, respectively. At the 200 day offset, the functions take a value of 0.08, 0.15 and 0.18.

The low first-order autocorrelations already suggests that the three volatility measures do not contain a unit root. The Augmented Dickey-Fuller test, allowing for a constant and 22 lags, yields test statistics of -4.692, -3.666 and -3.096 – rejecting therefore the unit root hypothesis, I(1), at the 5% level. The slow hyperbolic decay, however, indicates the presence

[6] It is often noted that tests for normality can be grossly incorrect in finite samples and/or when the observations are dependent (see for instance Beran 1994, Chapter 10 and the references therein). Upon simulating the fractionally integrated process $(1-L)^{0.4}\, y_t = -0.04 + \varepsilon_t$, $\varepsilon_t \sim i.i.d.\, N(0, 0.2)$ and $t = 1, \ldots, 1366$ (see Table 2.1 later in this chapter), we find for y_t, using 100,000 trials, that the 95% confidence interval for $\hat{S}$ and $\hat{K}$ is given by $[-0.195, 0.194]$ and $[2.707, 3.305]$, respectively. For ε_t, we obtain $[-0.130, 0.130]$ and $[2.760, 3.278]$. The asymptotic intervals under *i.i.d.* normality are $[-0.130, 0.130]$ and $[2.740, 3.260]$. Under these conditions, the degree of bias is therefore not severe.

Figure 2.2: Distribution of Realized Volatility



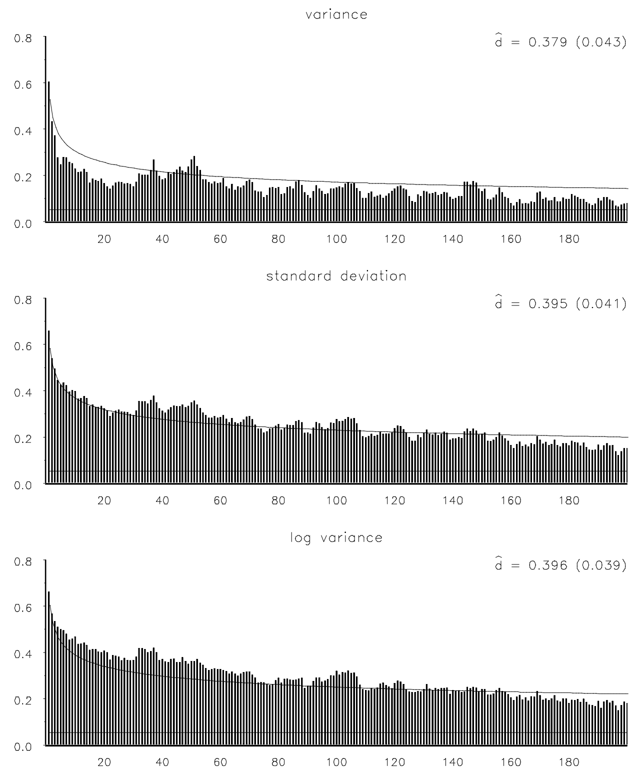
The graphs display the density estimates of variances (top panel), standard deviations (middle panel) and logarithmic variances (bottom panel). All series are standardized to mean zero and variance one. The bottom panel graphs along with the density estimates the standard normal probability distribution function (dashed line). Skewness ($\hat{S}$) and kurtosis ($\hat{K}$) coefficients are displayed in the top-right corner of each plot.

of long-memory. Using squared returns (or some transform thereof), this phenomenon has been documented by Ding, Granger and Engle (1993) and Crato and de Lima (1994), among others.

A covariance stationary fractionally integrated process, I(d), has the property of long-memory in the sense that the autocorrelations decay at a slow hyperbolic rate when $0 < d < 0.5$. In the top-right corner of each plot in Figure 2.2 we display the Geweke and Porter-Hudak (1993) log-periodogram estimate for the fractional integration parameter d; standard errors are given in parentheses.[7] The theoretical autocorrelation functions implied

[7]The estimates are obtained using the first to $m = T^{4/5} = 322$ spectral ordinates and this choice is optimal according to Hurvich, Deo and Brodsky (1998). Standard errors are obtained using the usual OLS regression formula and are slightly higher than the asymptotic standard error of the estimator, $\pi/\sqrt{24\,m} = 0.036$.

Figure 2.3: Realized Volatility Sample Autocorrelation Functions



The graphs display the first 200 sample autocorrelations of variances (top panel), standard deviations (middle panel) and logarithmic variances (bottom panel). The horizontal lines are the upper limit of the 95% confidence interval, $1.96/\sqrt{T}$. Geweke and Porter-Hudak estimates for the fractional integration parameter are given in the top-right corner of each plot; standard errors are reported in parentheses. The lines are the theoretical autocorrelations implied by these estimates.

by these estimates for d match rather well the sample autocorrelations and the parameter estimates for d are more than two standard errors away from 0.5 and several standard errors away from zero.[8] This suggests that realized volatilities are covariance stationary and fractionally integrated.

[8]The unweighted minimum distance estimator proposed by Tieslau, Schmidt and Baillie (1996) minimizes the sum of squares between the theoretical autocorrelation function of an I(d) process and the sample autocorrelation function. Upon applying this estimator to the first 200 autocorrelations, we obtain for the fractional integration parameter estimates of 0.346, 0.389 and 0.401 for variances, logarithmic variances and standard deviations, respectively.

2.4.3 Volatility and Returns

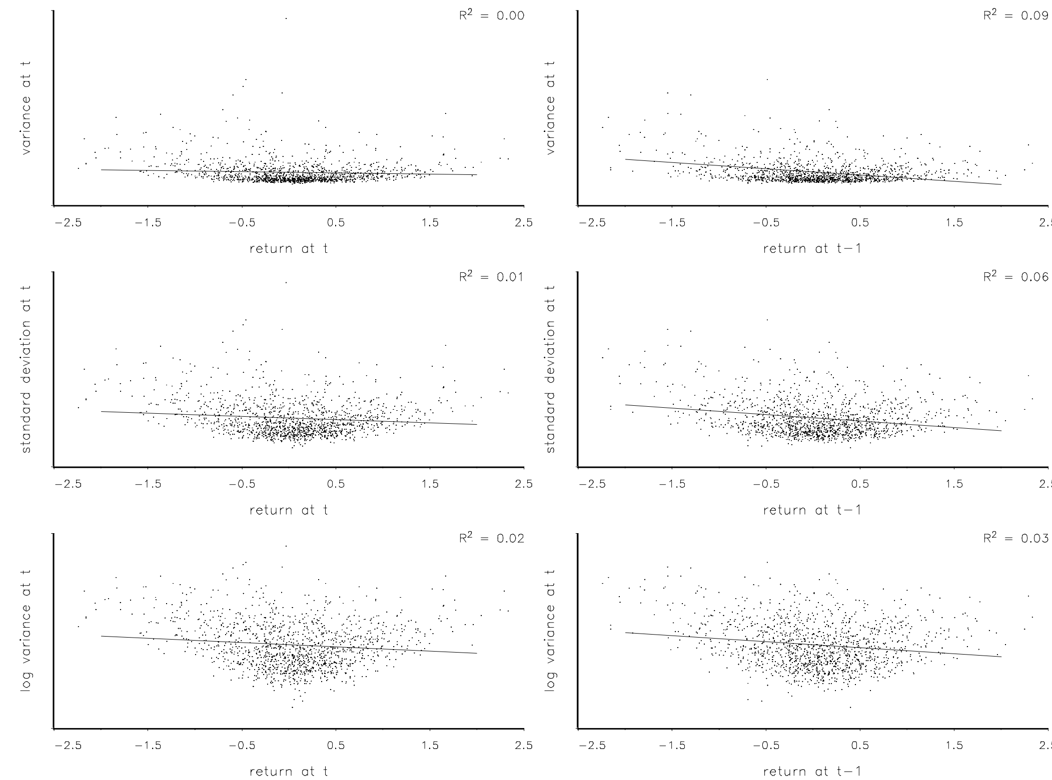
The relation between volatility and returns is of interest for two reasons. First, theoretical models such as Merton's (1973) intertemporal CAPM model relate current expected excess returns to volatility. This has motivated the ARCH in mean, or ARCH-M, model introduced by Engle, Lilien and Robins (1987). In the ARCH-M specification the conditional mean of returns (or excess returns) is a linear function of the conditional variance. Second, Black (1976), Pagan and Schwert (1990) and Engle and Ng (1993), among others, have documented asymmetries in the relation between news (as measured by lagged returns) and volatility – suggesting that good and bad news have different predictive power for future volatility. Generally it is found that a negative return tends to increase subsequent volatility by more than would a positive return of the same magnitude. This phenomenon is known as the 'leverage' or 'news' effect.

In Figure 2.4 the relation between our three volatility measures and current returns is displayed on the left whereas the relation with lagged returns is given on the right. Through the scatters, the graphs display an ordinary least squares regression line which is based on the displayed variables and a constant term. The R^2 statistic from each regression is given in the top-right corner of the plots.

Focusing on the three graphs on the left, we can see that there is no 'important' linear relation between the three volatility measures and current returns – suggesting that the ARCH-M effect is negligible for our data. It becomes however obvious that volatilities are non-linear in returns; all three volatility measures increase with positive and negative returns. Note also how in each of the three plots a convex frontier seems to shape out. This implies that a particular daily price change generates some minimum level of volatility.

The plots on the right of Figure 2.4 suggest the presence of the leverage effect: lagged negative returns yield high volatility more frequently than lagged positive returns. This phenomenon is most pronounced for variances and least obvious for logarithmic variances.

Figure 2.4: Realized Volatilities and Current and Lagged Returns



The graphs display returns (left panels) and lagged returns (right panels) against variances (top panel), standard deviations (middle panel) and logarithmic variances (bottom panel). The lines are OLS regression lines which are based on the displayed variable and a constant term. The regression R^2 measures are given in top-right corner of the plots. In all graphs we omit four observations that are to the left and three observations that are to the right of the plot region.

It is quite surprising that such asymmetry is less evident when looking at the graphs using current returns. If the news-effect is indeed the source of asymmetry, one would expect that current news, rather than past news, yield the suggested effect. Possibly it takes time for some market participants to react.

To investigate further the asymmetric response of volatility to past returns, we fit via ordinary least squares the following regression models to our data:

$$\begin{aligned} s_t^2 &= \omega_1 + \omega_2\,\mathrm{I}^- + \omega_3\, r_{t-1}^2 + \omega_4\, r_{t-1}^2\,\mathrm{I}^- + \varepsilon_t \\ s_t &= \omega_1 + \omega_2\,\mathrm{I}^- + \omega_3\, r_{t-1} + \omega_4\, r_{t-1}\,\mathrm{I}^- + \varepsilon_t \\ \ln(s_t^2) &= \omega_1 + \omega_2\,\mathrm{I}^- + \omega_3\, r_{t-1} + \omega_4\, r_{t-1}\,\mathrm{I}^- + \varepsilon_t \end{aligned} \tag{2.4.1}$$

where s_t^2 denotes realized variances; the indicator I^- takes value one when $r_{t-1} < 0$ and is zero otherwise. Note that we allow for asymmetry in intercepts as well as slopes and that we consider for variances a quadratic relation between lagged returns and volatility.[9]

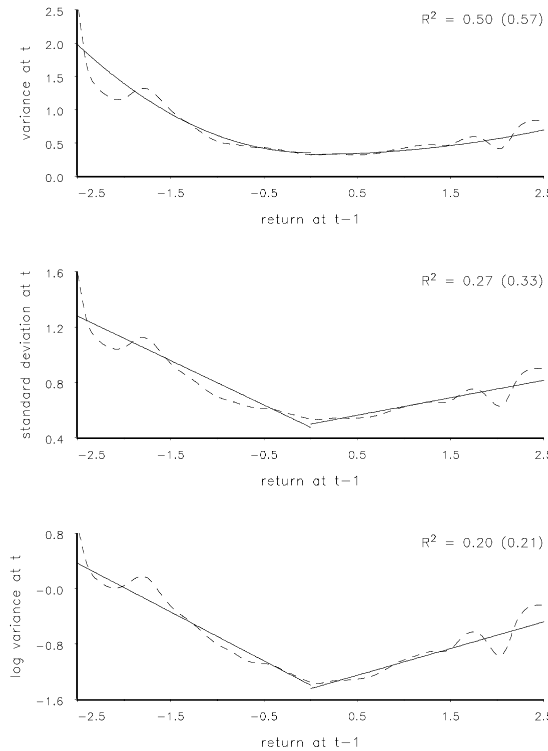
In Figure 2.5 we plot the regression lines implied by estimates of equation 2.4.1 (solid line) along with the regression lines implied by the nonparametric models of lagged returns on each of the three volatility measures (dashed line).[10] The R^2 statistics from the parametric and nonparametric regressions (in parentheses) are displayed in the top-right corner of each plot.

Both the parametric and nonparametric regressions confirm the asymmetric news-effect – volatility increases more steeply with negative than with positive returns. The news-impact functions are centered around $r_{t-1} = 0$; this suggests that asymmetry is only in slopes and not in intercepts. The close correspondence between the parametric and nonparametric regression lines indicates that the models given by equation 2.4.1 characterize well the news-impact functions for the DJIA portfolio. There are no obvious discrepancies that would suggest any other parametric specification to capture the lagged return volatility

[9] In our estimations we find, as one would expect from the results in Section 2.4.2, that the residual innovations $\hat{\varepsilon}_t$ are serially correlated and non-normal (see also Figure 2.6 which we shall discuss shortly). Nonetheless, the least squares estimator yields under these circumstances still unbiased, albeit not efficient, coefficient estimates.

[10] The nonparametric regression estimates are obtained using the Nadaraya-Watson estimator with a Gaussian kernel. The bandwidth parameters are determined using cross-validation scores. Estimation was done over the entire sample, yet the plot regions are restricted to returns in the -2.5 to 2.5 interval. Four observations are smaller than -2.5 and three observations are greater than 2.5. Note that the kernel estimator is consistent despite non-normal and correlated residuals. However bandwidth selection by cross-validation gives under-smoothed estimates (see Härdle and Linton 1994).

Figure 2.5: Parametric and Nonparametric News Impact Functions

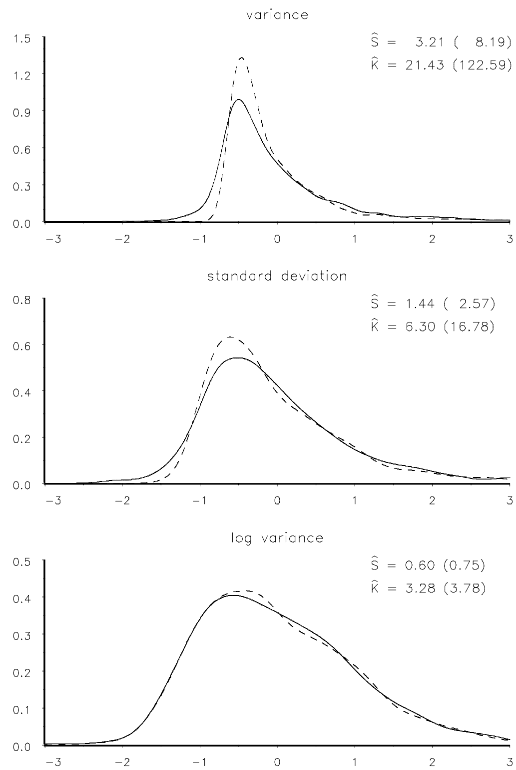


The graphs display the regression lines implied by estimates of equation 2.4.1 and nonparametric regression estimates of lagged returns on variances (top panel), standard deviations (middle panel) and logarithmic variances (bottom panel). The R^2 of both the parametric and nonparametric regressions (in parentheses) are given in top-right corner of each plot.

relation.

For the modeling of volatility it becomes of interest whether the news-effect can account for the asymmetry and excess kurtosis we observe in the distribution of our volatility series (Figure 2.2). In Figure 2.6, we graph the distribution of the variance, standard deviation and logarithmic variance series after (using solid lines) and before (dashed lines) accounting for the news effect. For the 'after news-effect' distributions we use the residuals from the models defined by equation 2.4.1. The skewness and kurtosis coefficients are displayed in the top-right corner of each plot. The estimates before accounting for news-effects are reported in parentheses.

Figure 2.6: Distribution of Realized Volatility and News Impact



The graphs display the density estimates of variances, standard deviations and logarithmic variances before and after accounting for news-effects. The density estimates after accounting for news-effects (solid line) are obtained from the OLS residuals of the models defined by equation 2.4.1. The density estimates before accounting for news-effects (dashed lines) are identical to the ones displayed in Figure 2.2. Skewness ($\hat{S}$) and kurtosis ($\hat{K}$) coefficients are displayed in the top-right corner of each plot. The estimates before accounting for news-effects are reported in parentheses.

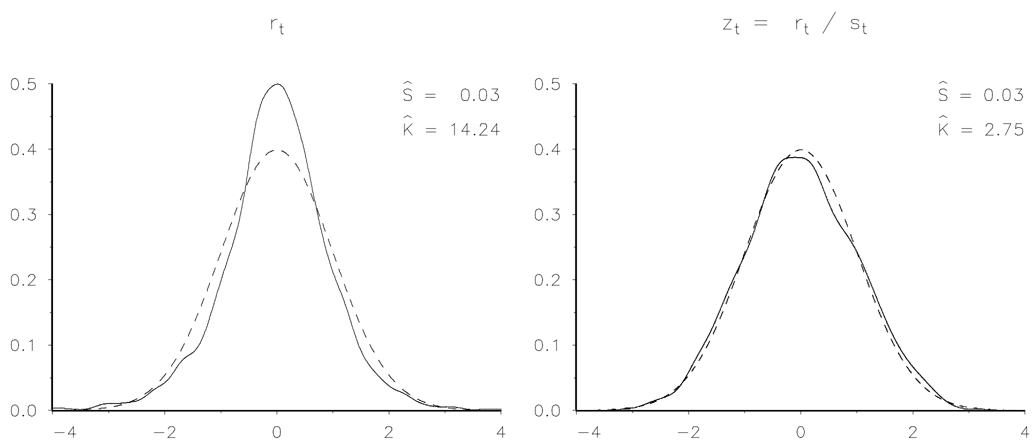
Even after accounting for the asymmetric response of volatility to lagged returns, the distribution of variances and standard deviations remains clearly non-normal. News-effects can however remove some of the asymmetry and flatness in the distribution of the volatility measures. For variances, standard deviations and logarithmic variances, the skewness coefficient is reduced from 8.19 to 3.21, 2.57 to 1.44 and 0.75 and 0.60, respectively. The kurtosis coefficient decreases from 122.59 to 21.43, 16.78 to 6.30 and 3.78 to 3.28 for the respective volatility measures. As there is little asymmetry in the distribution of logarithmic variances, the corresponding reduction in the skewness and kurtosis estimates is only modest, however. Judged by the standard errors of these estimates (see Section 2.4.1), normality of logarithmic variances is again rejected.

2.4.4 Distribution of Returns and Standardized Returns

An empirical regularity found almost universally across all assets is that high frequency returns are leptokurtic. Early evidence for this dates back to Mandelbrot (1963) and Fama (1965). Clark (1973) established that a stochastic process is thick tailed if it is conditionally normal with changing conditional variance. ARCH and Stochastic Volatility models have this property, but it is often found that these models do not adequately account for leptokurtosis. Specifically, returns divided by the estimated standard deviations ($z_t = r_t/\hat{\sigma}_t$) display frequently excess kurtosis. As a result, several other conditional distributions have been employed to fully capture the degree of tail fatness (see for instance Hsieh 1989 and Nelson 1991).

Realized standard deviations allow us to characterize the distribution of standardized returns without modeling changing variances. In Figure 2.7, we plot the density of the daily return series (r_t) on the left whereas we depict the density of this series scaled by daily standard deviations ($z_t = r_t/s_t$) on the right. In each graph we also plot the normal density. Skewness ($\hat{S}$) and kurtosis ($\hat{K}$) estimates are given in the top-right corner of the plots.

Figure 2.7: Distribution of Returns and Standardized Returns



The graphs display the density estimates of daily returns r_t (left panel) and scaled returns $z_t = r_t/s_t$ (right panel), where s_t are daily realized standard deviations. The displayed series are standardized to mean zero and variance one (the mean and standard deviation of z_t equal 0.132 and 1.041, respectively). The standard normal density is plotted with dashed lines.

Returns are hardly skewed, but leptokurtic as expected. From the kurtosis estimate of scaled returns it becomes evident that changing variances can fully account for fat tails in returns – the estimate even suggests that this distribution is platykurtic. The density of z_t is very close to the one implied by the normal distribution. Based on the standard errors of the skewness and kurtosis estimates (see Section 2.4.1), normality cannot not be rejected at the 5% level – $\hat{S}$ and $\hat{K}$ are within two standard errors of their hypothesized values.

Recall from Section 2.4.1 that we found that logarithmic variances are distributed nearly normal – implying that standard deviations and variances are distributed approximately lognormal. Combined with the normality of z_t, this suggests that returns are approximately a normal-lognormal mixture which has been proposed by Clark (1973). In Clark's model, however, the volatility process is assumed *i.i.d.* whereas we find that it is serially correlated (see Section 2.4.2).[11]

2.4.5 Comparison to Exchange Rates

Our results regarding the distribution and persistency of realized stock volatility are remarkably similar to the ones obtained by Andersen *et al.* (1999) in the setting of exchange rates. They also found that the distribution of variances and standard deviations is skewed right and leptokurtic, but that logarithmic variances are distributed approximately normal.[12] Exchange rate return variances, standard deviations and logarithmic variances display a high degree of persistency as well. Depending on the volatility measure and exchange rate series used, Andersen *et al.* report Geweke and Porter-Hudak (1993) log-periodogram estimates ranging between 0.346 and 0.421. Results on news impact however differ. Contrary to our results, they did not find much evidence for the asymmetric volatility effect. This is

[11] In Clark it is also assumed that the series z_t is independent. The BDS test (Brock, Dechert, LeBaron and Scheinkman 1987) yields test statistics of $W_2 = -2.721$, $W_3 = -2.839$ and $W_4 = -2.089$. As these are distributed standard normal, we therefore have to reject independence of z_t at the 5% level. However, Ljung-Box portmanteau statistics for up to $\{10, 20, 100\}$ th-order serial correlation in z_t and z_t^2 are insignificant at the 5% level.

[12] Without accounting for the leverage effect, our skewness and kurtosis estimates are higher than the ones reported for exchange rates by Andersen *et al.* After adjusting for the effect (see Section 2.4.3), our estimates become quite close to theirs, however.

to be expected, however, as this phenomenon is generally observed for equities only.

2.5 Realized Volatility Modeling and Predictions

In this section we first build models aimed to capture the temporal dependency of realized volatility. Treating volatility as observed instead of latent allows us to utilize the time series techniques employed when modeling the conditional mean. We thus can sidestep the relatively more complicated ARCH and Stochastic Volatility formulations that model and measure volatility simultaneously. Later in this section we investigate how well the developed models predict volatility *ex ante* one-step-ahead. We shall compare these predictions to the ones obtained by ARCH models in our next section.

2.5.1 Realized Volatility Modeling

As far as the modeling of our three volatility measures is concerned, the main findings of our previous section are: (1) the distributions of variances and standard deviations are asymmetric and leptokurtic, but logarithmic variances are distributed approximately normal; (2) realized volatilities appear covariance stationary and fractionally integrated; and (3) volatility is correlated with lagged negative and positive returns. Before detailing the specific models we shall employ to account for (2) and (3), we discuss first the implications of the distributional characteristics of our volatility measures for the modeling of these series.

Assuming that the only deterministic component of a covariance stationary process y_t is a (possibly non-zero) mean ω, then it is well known that the Wold representation of y_t is a (possibly infinite) moving average, *i.e.* $y_t = \omega + \varepsilon_t + \sum_{i=1}^{q} \alpha_i \, \varepsilon_{t-i}, \ \varepsilon_t \sim WN(0, \sigma^2)$ where WN denotes serially uncorrelated white noise. Estimation and inference generally require the stronger assumption that $\varepsilon_t \sim i.i.d. \, WN(0, \sigma^2)$ and using this premise it is straightforward to show that:

$$K_y - 3 \;=\; \frac{1 + \sum_i^q \alpha_i^4}{(\,1 + \sum_i^q \alpha_i^2\,)^2}\,(\,K_\varepsilon - 3\,)$$

$$S_y \;=\; \frac{1 + \sum_i^q \alpha_i^3}{(\,1 + \sum_i^q \alpha_i^2\,)^{3/2}}\,S_\varepsilon$$

where $K_y,\ S_y\ (K_\varepsilon,\ S_\varepsilon)$ denotes the kurtosis and skewness of y_t (ε_t). Note, as $S_y > 0$, $K_y > 3$, $\exists\, i$ s.t. $\alpha_i \neq 0$ and $0 \leq \alpha_i \leq 1\ \forall\, i$ then $S_\varepsilon > S_y$ and $K_\varepsilon > K_y$. Because we found in our previous section that variances as well as standard deviations – even after accounting for the news effect – are highly skewed and leptokurtic and that the sample autocorrelation functions of these volatility measures are positive and slowly decaying (suggesting $0 \leq \alpha_i$), a model with a moving average representation would leave these distributional characteristics unexplained and even amplified in the residuals. When estimation is done by maximum likelihood, as it is commonly the case, this in turn would require one to either rely on quasi-maximum likelihood estimates or to condition the residuals on a density that allows for skewness and excess kurtosis. However, the former approach may not yield consistent estimates of the parameters and variance-covariance matrix whereas the latter would complicate analysis as it requires additional coefficients.[13] We found, however, that the distribution of logarithmic variances is almost symmetric and subject to little excess kurtosis. For these reasons we restrict our attention to modeling this series only. Of course, logarithmic variances are rarely of interest. We address this issue by investigating in our next subsection whether logarithmic predictions transformed into variances and standard deviations provide useful descriptions of these two volatility measures.

[13]One density that allows for skewness and kurtosis is the exponential generalized beta (McDonald and Xu 1995). We used this density to estimate the ARFIMAX specification discussed below to model variances and standard deviations directly. Any improvements, as measured by *ex ante* one-day-ahead prediction criteria, were minor only. Alternatively, one may consider estimation in the frequency domain, which can allow one to relax the normality assumption. However, such models do not easily allow for the type of exogenous variables we consider.

To account for long-memory and the correlation of volatility with lagged negative and positive returns we model logarithmic variances using the following ARFIMAX (p,d,q) model:

$$\big(1-L\big)^d\,\big(1-\beta(L_p)\big)\ln(s_t^2) = w_0 + w_1\,r_{t-1}\,\mathrm{I}^- + w_2\,r_{t-1}\,\mathrm{I}^+ + \big(1+\alpha(L_q)\big)\,\varepsilon_t \qquad (2.5.1)$$

where $\varepsilon_t \sim i.i.d.\ N(0,\sigma^2)$, $\alpha(L_q) = \sum_{i=1}^{q} \alpha_i\,L^i$ and $\beta(L_p) = \sum_{i=1}^{p} \beta_i\,L^i$. Realized variances are denoted by s_t^2, the indicator I^- (I^+) takes value one when $r_{t-1} < 0$ ($r_{t-1} \geq 0$) and is zero otherwise. Next to the standard ARMA (p,q) coefficients (w_0, $\beta(L_p)$, $\alpha(L_q)$) the above specification contains the following three coefficients: a fractional integration parameter (d) to capture the slow hyperbolic decay in the sample autocorrelation function; lagged negative (ω_1) and positive (ω_2) returns to allow for the leverage effect as well as to account for the slight asymmetry and tail fatness in the distribution of $\ln(s_t^2)$. We estimate the above model using the conditional sum-of-squares maximum likelihood estimator suggested by Hosking (1984). The finite sample properties of this estimator have been investigated by Chung and Baillie (1993).

Parameter estimates of three specifications nested within the above model – an ARFIMA (0,d,0) labeled FI, an ARFIMAX (0,d,0) with label FIX and an ARFIMAX (0,d,1) labeled FIMAX – are given in Table 2.1. Standard errors are reported in parentheses under the coefficient estimates. All of the parameters are statistically significant at the 5% level on the basis of either Wald or likelihood ratio tests. The table reports in addition the Schwartz Bayesian Information Criterion (SBC) and Ljung-Box portmanteau statistics for up to Kth-order serial correlation in the residuals (Q_K). The numbers in parentheses below these statistics report the probability that the K autocorrelations are not significant.

Paying attention to the estimates of the fractional integration parameter d first, we can see that our estimation results confirm our earlier suspicion that the logarithmic variance process is stationary and fractionally integrated. Estimates for d range between 0.324 and 0.392 and are several standard errors away from both zero and 0.5. The FI model estimate

Table 2.1: Realized Volatility Model Estimates

	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{d}$	$\hat{\alpha}_1$	$\hat{\sigma}^2$	*SBC*	Q_{10}	Q_{20}	Q_{100}
FI	-0.043			0.392		0.221	-918.8	8.540	12.969	97.410
	(0.156)			(0.020)		(0.009)		(0.287)	(0.738)	(0.469)
FIX	-0.153	-0.316		0.324		0.205	-870.6	16.064	21.866	102.520
	(0.020)	(0.030)		(0.017)		(0.008)		(0.013)	(0.148)	(0.306)
FIMAX	-0.170	-0.336	0.067	0.344	-0.100	0.203	-871.9	12.880	18.043	94.291
	(0.025)	(0.034)	(0.031)	(0.023)	(0.037)	(0.008)		(0.012)	(0.205)	(0.472)

Coefficients of the ARFIMAX model defined by equation (2.5.1) are obtained by conditional sum-of-squares maximum likelihood estimation using analytical gradients. The $(1-L)^d$ polynomial is truncated at lag 1000. Standard errors, based on the second derivatives of the log-likelihood function, are reported in parentheses under the coefficient estimates. *SBC* reports the Schwarz Bayesian Information Criterion ($SBC = \mathfrak{L}^* - 0.5\,k\,\ln(1366)$, where $\mathfrak{L}^*$ denotes the maximized log likelihood and k the number of estimated coefficients). Q_K refers to the Ljung-Box portmanteau tests for up to Kth-order serial correlation in the residuals. The numbers in parentheses below these statistics report the probability that the K autocorrelations are not significant.

of $d = 0.392$ corresponds closely to the Geweke and Porter-Hudak (1993) log-periodogram estimate of $d = 0.396$ obtained in our previous section. The FI model estimate is also in accordance with Breidt, Crato and de Lima (1998) who on estimating a ARFIMA(1,d,0) Stochastic Volatility process (without allowing for the asymmetric volatility effect) report $d = 0.444$ for the CRSP index. Upon fitting a FIEGARCH model to daily returns on the S&P 500 composite stock index, Bollerslev and Mikkelsen (1996) however found $d = 0.633$. This estimate is much higher than the ones we report and suggests, contrary to our results, that the logarithmic variance process is not covariance-stationary.

Looking next at the estimates for ω_1 and ω_2, we find support for the asymmetric news-effect. It becomes evident, however, that it is mostly negative and not positive returns that are important for the modeling of logarithmic variances; the estimate of ω_2 in the FIMAX specification is small and only marginally significant at the 5% level.

The addition of lagged negative and/or positive returns to the FI model induces some low-order serial correlation in the residuals. While for the FI model all reported Ljung-Box Q statistics are insignificant at the conventional levels, for the FIX and FIMAX model we cannot – at the 5% significance level – reject the null of no 10th-order serial correlation in the residuals. For the FIMAX model we mitigate this problem by allowing next to the two news-parameters a first-order moving average component; the coefficient on α_1 is however only small and accompanied by a relatively large standard error.

Judged by the Schwarz Bayesian Criterion the asymmetric return-volatility effect is important for the modeling of logarithmic variances. Among the two models that allow for lagged returns, the criterion favors the parsimonious FIX specification which does not include negative returns and a first-order moving average component.

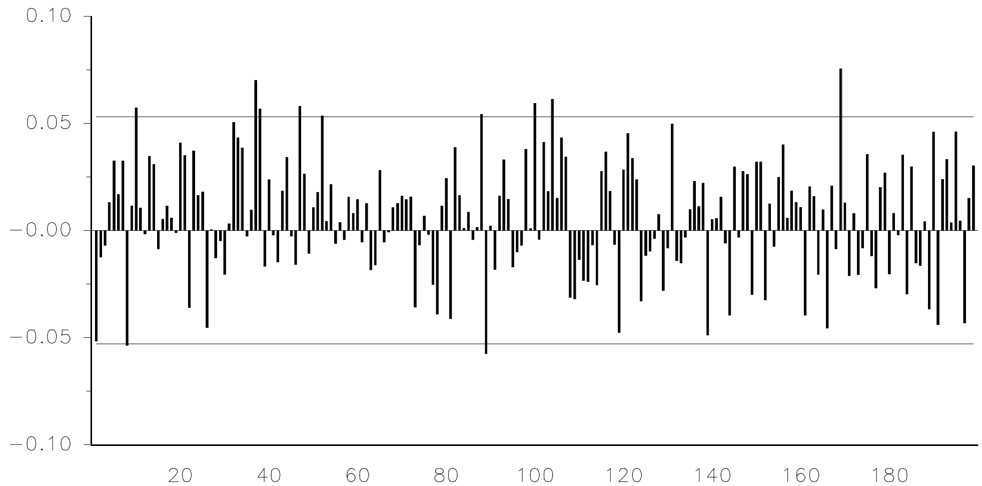
To investigate further the possibility that the FIX model leaves some time-dependency of volatility unexplained, we plot in Figure 2.8 its residual autocorrelation function. The significant Q_{10} statistic for this model is likely driven by the size of the first, eighth and tenth-order residual autocorrelation. Judged by the 95 percent confidence interval, $\pm 1.96/\sqrt{T}$, only the eight and tenth order autocorrelations are significant – however only marginally so. When considering all 200 autocorrelations it becomes evident that the FIX model captures logarithmic variance dynamics rather well. The eleven significant autocorrelations may be attributed to type II error of the test. Above all, the FIX model accounts fully for the slow hyperbolic decay found in the logarithmic variance autocorrelation function (see Figure 2.3). In Figure 2.8, no pattern of decay remains.

2.5.2 Realized Volatility Model Predictions

In this subsection we investigate how well the realized logarithmic volatility models set out above predict our three volatility series *ex ante* one-step-ahead. To determine the next-period predictions, it is convenient to rewrite the ARFIMAX model given by equation 2.5.1 more compactly as:

$$\ln(s_t^2) = f(\mathfrak{F}_{t-1}) + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.\ N(0, \sigma^2) \tag{2.5.2}$$

Figure 2.8: Realized Volatility Model Residual Autocorrelations



The graph displays the first 200 residual autocorrelations for the FIX model reported in Table 2.1. The parallel lines are the 95% confidence interval, $\pm 1.96/\sqrt{T}$.

where $\mathfrak{F}_{t-1}$ denotes the information set available at time $t-1$. The one-step-ahead variance, standard deviation and logarithmic variance predictions of (2.5.2) evaluated at the estimates given in Table 2.1 are given by:

$$\begin{aligned}
\hat{s}_t^2 &= E\big[\, s_t^2 \,|\, \mathfrak{F}_{t-1}\big] &&= e^{\hat{f}(\cdot)} \quad E\big[\, e^{\hat{\varepsilon}_t} \,|\, \mathfrak{F}_{t-1}\big] &&= e^{\hat{f}(\cdot)+\frac{1}{2}\,\hat{\sigma}^2} \\
\hat{s}_t &= E\big[\, s_t \,|\, \mathfrak{F}_{t-1}\big] &&= e^{\frac{1}{2}\hat{f}(\cdot)} \quad E\big[\, e^{\frac{1}{2}\hat{\varepsilon}_t} \,|\, \mathfrak{F}_{t-1}\big] &&= e^{\frac{1}{2}\hat{f}(\cdot)+\frac{1}{8}\,\hat{\sigma}^2} \\
\ln(\hat{s}_t^2) &= E\big[\ln(\hat{s}_t^2) \,|\, \mathfrak{F}_{t-1}\big] &&= \hat{f}(\cdot) + E\big[\, \hat{\varepsilon}_t \,|\, \mathfrak{F}_{t-1}\big] &&= \hat{f}(\cdot)
\end{aligned} \qquad (2.5.3)$$

Since in (2.5.2) it is assumed that $\varepsilon_t \sim N(0,\sigma^2)$ it follows that $\exp(\varepsilon_t) \sim LN(0,\sigma^2)$ and $\exp(\frac{1}{2}\,\varepsilon_t) \sim LN(0, \frac{1}{4}\,\sigma^2)$, where LN denotes the lognormal density.[14] Let y_t denote one of our three volatility series, *i.e.* s_t^2, s_t and $\ln(s_t^2)$, then we evaluate its predictions $\hat{y}_t$, *i.e.* $\hat{s}_t^2$,

[14] As one would expect form the discussion at the beginning of this section, the residual innovations coming from our logarithmic variance models display slight asymmetry and excess kurtosis. In particular, we obtain skewness and kurtosis estimates of $\{0.560, 4.304\}$, $\{0.380, 3.984\}$ and $\{0.362, 4.134\}$ for the FI, FIX and FIMAX model respectively. However, when we instead compute the expectations of $\exp(\hat{\varepsilon}_t)$ and $\exp(\frac{1}{2}\,\hat{\varepsilon}_t)$ using the mean of these two measures in order to obtain variance and standard deviation predictions, our subsequent results change only little. Furthermore, when we condition the residual innovations on non-normal densities, results hardly change.

$\hat{s}_t$ and $\ln(\hat{s}_t^2)$, using the OLS regression:

$$y_t = \alpha + \beta\, \hat{y}_t + \varepsilon_t \tag{2.5.4}$$

If a prediction is unbiased, $\alpha = 0$ and $\beta = 1$. Table 2.2 reports the ordinary least squares estimates of (2.5.4) and the associated R^2 statistic when applied to variances, standard deviations and logarithmic variances. Standard errors using White's (1980) heteroskedasticity correction are in parentheses.

Table 2.2: Realized Volatility Model *Ex Ante* Predictions

	variances			*standard deviations*			*log variances*		
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2
FI	-0.079	1.238	0.379	-0.048	1.086	0.486	0.028	1.024	0.515
	(0.058)	(0.162)		(0.031)	(0.057)		(0.039)	(0.030)	
FIX	0.000	1.026	0.627	-0.030	1.055	0.576	0.026	1.022	0.551
	(0.031)	(0.085)		(0.023)	(0.041)		(0.035)	(0.027)	
FIMAX	0.071	0.843	0.607	0.000	1.003	0.576	0.007	1.006	0.554
	(0.040)	(0.103)		(0.028)	(0.049)		(0.035)	(0.026)	

The table reports ordinary least squares coefficient estimates for the model defined by equation 2.5.4 using the variance, standard deviation and logarithmic variance predictions given by (2.5.3), *i.e.* the *ex ante* one-step-ahead volatility predictions coming from the FI, FIX and FIMAX models reported in Table 2.1. Standard errors using White's (1980) heteroskedasticity correction are in parentheses.

For all three models, the estimates of α and β are within two standard errors of their hypothesized values. From the R^2 statistics it becomes evident that the realized volatility specifications can explain much that is observed in volatility over our sampling period. The R^2 statistics for logarithmic variances range between 51.5% and 55.4%, for standard deviations between 48.6% and 57.6% and for variances between 37.9% and 62.7%. The addition of lagged negative returns to the FI model (therefore yielding the FIX model) improves only slightly the predictions for logarithmic variances, but has important consequences for the predictions of standard deviations and most notably for variances; the R^2 measure for this latter volatility measure increases by 24.8 percentage points to 62.7%. Little or nothing is however gained by adding positive lagged returns and a moving average component to the FIX model (therefore yielding the FIMAX model). The R^2 for logarithmic variances increases by only 0.3 percentage points. For standard deviations the R^2 measures are identical and for variances the parsimonious FIX specification yields an even higher R^2 measure than the FIMAX specification that requires two additional parameters.

We plot in Figure 2.9 the one-step-ahead *ex ante* variance predictions implied by the FIX model (solid line) along with the realized variance series (dotted line). Clearly, the one-day ahead predictions do a remarkable job of tracking realized variances over our sample period. Major discrepancies between the two depicted series are generally only noticeable when the realized volatility is unusually high (for instance March 31, 1994 and July 16, 1997). However, for the highest realized variance observation in the sample (October 28, 1997) the FIMAX model predicts a variance of 10.43 while the corresponding realized volatility measure takes value of 9.45 for that day.

The question remains whether employment of realized volatility measures to model volatility leads to improvements or whether perhaps one of the standard techniques yields similar or even better result. We tackle this issue in our next section.

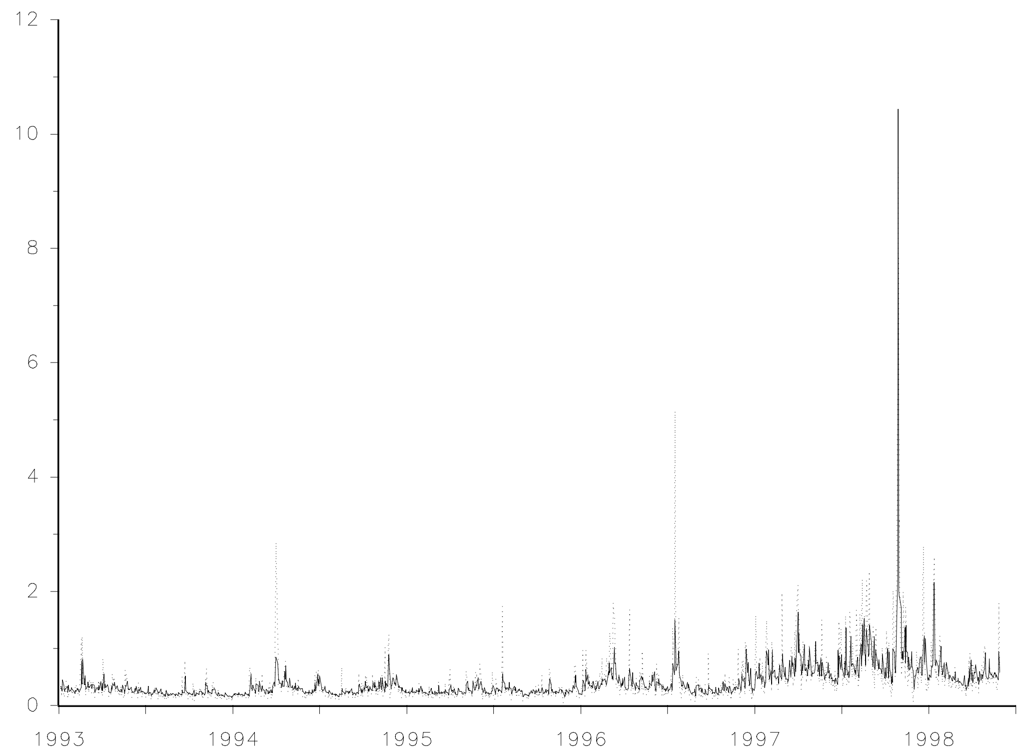
2.6 ARCH Volatility Modeling and Predictions

The most common tool for characterizing changing variances is to fit ARCH-type models to daily returns. The performance of some of these models relative to the ones just developed is the subject of this section. We detail next the exact formulations we shall be using. Later in this section we evaluate the volatility predictions implied by these models and this will allow us to directly compare the ARCH models to the realized volatility formulations employed before.

2.6.1 ARCH Volatility Modeling

For the parameterization of ARCH models the main findings of our previous sections are: (1) volatilities are covariance stationary and fractionally integrated, (2) volatilities are non-

Figure 2.9: Realized Volatility Model *Ex Ante* Variance Predictions



The graph displays the *ex ante* variance predictions implied by the FIX model (solid line) along with the realized variances (dashed line). The FIX model is defined by equation 2.5.1 and its estimates are reported in Table 2.1.

symmetric in lagged returns, (3) returns are not (at best only weakly) correlated with volatilities and (4) the distribution of returns divided by standard deviations is normal. In many applications with daily data it is however found that the distribution of standardized returns is leptokurtic. We shall therefore investigate whether our finding of normality is particular to the data underlying our study or whether perhaps non-normality is confined to the ARCH approach to modeling volatility.

Since the introduction of the ARCH model by Engle (1982) numerous extensions have been proposed.[15] However, only the FIGARCH model developed by Baillie, Bollerslev and Mikkelsen (1996) and the FIEGARCH model formulated Bollerslev and Mikkelsen (1996)

[15]Recent studies surveying the various ARCH models include Pagan (1996), Palm (1996), Bollerslev, Engle and Nelson (1994), Bera and Higgins (1993) and Bollerslev, Chou and Kroner (1992).

explicitly allow for the long-memory property of volatility. We shall focus on these two specifications although only the FIEGARCH model allows for the news-effect and can be covariance stationary while allowing for long-memory.

When modeling the conditional variance processes discussed below, we did not find any evidence for temporal dependencies in the conditional mean of returns (r_t) other than a constant term (μ). Since we in Section 2.4.3 found hardly any evidence for the ARCH-M effect, we consider return representations of the form:[16]

$$
\begin{aligned}
r_t &= \mu + \varepsilon_t \\
\varepsilon_t &= \sigma_t \, z_t
\end{aligned}
\tag{2.6.1}
$$

where $E[z_t] = 0$ and $E[z_t^2] = 1$.

The conditional variance process in the FIGARCH (q,d,p) model is defined as:

$$
\sigma_t^2 = \frac{\omega + \Big[\big(1 - \beta(L_p)\big) - \big(1 - \alpha(L_q) - \beta(L_p)\big) \, \big(1 - L\big)^d \Big] \, \varepsilon_t^2}{\big(1 - \beta(L_p)\big)}
\tag{2.6.2}
$$

where $\alpha(L_q) = \sum_{i=1}^{q} \alpha_i \, L^i$, $\beta(L_p) = \sum_{i=1}^{p} \beta_i \, L^i$. The FIGARCH model is covariance stationary only in the special case where $d = 0$ and then it reduces to Bollerslev's (1986) GARCH specification. The FIGARCH model displays however the important property of having a bounded cumulative impulse-response function for any $d < 1$. As in Bollerslev (1987), we condition the innovations z_t in (2.6.1) on the Student t distribution, *i.e.* $z_t \sim T(0, 1, \eta_1)$. This density has thicker tails than the normal when $\eta_1 < \infty$.

Although the FIGARCH model is consistent with our finding of long-memory, for $d > 0$ the FIGARCH process is, contrary to our findings, not covariance stationary. Furthermore, variances are symmetric in lagged returns and therefore the FIGARCH model does not

[16] Nonetheless, for the GARCH (1,1), EGARCH (1,2) and FIEGARCH (0,d,1) models discussed below we still tested whether the ARCH-M specification is appropriate, *i.e.* $r_t = \mu_1 + \mu_2 \sigma_t^2 + \varepsilon_t$. As expected, the estimates for μ_2 were positive, yet insignificant.

permit the leverage effect. These two deficiencies are addressed by the FIEGARCH (p,d,q) model which is defined as:

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(L_q)\,\big(\,\gamma\, z_t + |z_t| - \mathrm{E}\big[\,|z_t|\,\big]\,\big)}{\big(\,1-L\,\big)^d\,\big(\,1-\beta(L_p)\,\big)} \tag{2.6.3}$$

with all polynomials defined as before. If the leverage effect holds, we expect to find $\gamma < 0$. This formulation nests Nelson's (1991) EGARCH model when $d = 0$. We condition – as in the original formulation of the EGARCH model – the innovations z_t on the generalized error distribution, *i.e.* $z_t \sim GED(0, 1, \eta_2)$. The density is normal when $\eta_2 = 2$ while it displays heavy tails for $\eta_2 < 2$. The fractional integration parameter in (2.6.3) has the same interpretation as in the models of our previous section, *i.e.*, the logarithmic variance process is covariance stationary if $d < 0.5$. For $d < 1$ the process is mean-reverting and shocks to volatility decay.

The FIEGARCH model is similar to our realized volatility model in that it seeks long-memory in the logarithmic variance process and allows for the asymmetric news-effect. Whereas our analysis of news impact in Section 2.4 and 2.5 suggests that logarithmic variances are linear in lagged positive and negative returns, the FIEGARCH model conjectures that logarithmic variances increase linearly with negative and positive standardized returns $(r_{t-1} - \mu)/\sigma_{t-1}$.[17] The main difference however is that our earlier specifications are in the spirit of Stochastic Volatility models and not of ARCH models.

Maximum likelihood estimates of some formulations nested within (2.6.2) and (2.6.3) – a GARCH (1,1), FIGARCH (1,d,1), EGARCH (1,2) and FIEGARCH (0,1,1) model – are reported in Table 2.3. Coefficient estimates for η carry suffix 1 when the ARCH innovations z_t are conditioned on the Student t density and suffix 2 when the generalized error density is used instead. Standard errors, based on the matrix of second derivatives of the log-likelihood function, are in parentheses. With the exception of the fractional integration parameter d

[17] Upon holding constant the information dated t-2 and earlier (as in the definition by Engle and Ng 1993), logarithmic variances are however linear in positive and negative r_{t-1}.

in the FIGARCH model, all reported estimates are significant at the 5% level on the basis of either Wald or log-likelihood ratio tests. $\mathfrak{L}^*$ reports the maximized log-likelihood.

Table 2.3: ARCH Model Estimates

	$\hat{\mu}$	$\hat{\omega}$	$\hat{\beta}_1$	$\hat{d}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}$	$\hat{\eta}_{1,2}$	$\mathfrak{L}^*$
GARCH	0.063	0.008	0.930		0.054			6.250_1	-1340.0
	(0.016)	(0.005)	(0.024)		(0.018)			(1.043)	
FIGARCH	0.063	0.021	0.652	0.375	-0.285			6.536_1	-1338.6
	(0.016)	(0.012)	(0.105)	(0.108)	(0.108)			(1.150)	
EGARCH	0.050	-0.884	0.972		0.231	-0.117	-0.596	1.425_2	-1329.1
	(0.015)	(0.125)	(0.014)		(0.043)	(0.047)	(0.158)	(0.075)	
FIEGARCH	0.065	-1.245		0.585	0.227		-0.668	1.418_2	-1326.7
	(0.014)	(0.274)		(0.056)	(0.041)		(0.039)	(0.072)	

Coefficients of the models defined by equations 2.6.1 and either 2.6.2 or 2.6.3 are obtained by conditional sum-of-squares maximum likelihood estimation using analytical gradients. Coefficient estimates for η carry suffix 1 when the ARCH innovations z_t are conditioned on the Student t density and suffix 2 when the generalized error density is used instead. Standard errors, based on the second derivatives of the log-likelihood function, are reported in parentheses under the coefficient estimates. $\mathfrak{L}^*$ reports the maximized log-likelihood. The $(1-L)^d$ polynomial in the FIGARCH and FIEGARCH model is truncated at lag 1000. The data are daily percentage returns for the Dow Jones Industrial Average from January 1993 to May 1998.

Consistent with prior literature on ARCH models, the innovations z_t are heavy tailed, the implied volatility processes are highly persistent and, when we allow for asymmetry in returns, coefficients on the news parameters suggest the leverage effect. The FIGARCH and FIEGARCH models indicate that shocks to volatility decay (eventually) at a slow hyperbolic rate. Our FIEGARCH estimate of $d = 0.585$ is in line with the one reported by Bollerslev *et al.* (1996), who found $d = 0.633$ for the S&P 500 composite stock index. Both estimates are however much higher than the ones we obtained in the context of our realized volatility models and imply, contrary to the findings of our previous sections, that the logarithmic variance process is not covariance stationary.

Judged by the maximized log-likelihood ($\mathfrak{L}^*$), the FIEGARCH model is the most promising ARCH specification for characterizing changing variances. We shall next investigate whether this or any other of the above formulations provide useful volatility predictions.

2.6.2 ARCH Volatility Model Predictions

ARCH model predictions are generally evaluated by means of criteria that match squared returns with the volatility predictions implied by a particular model (or some transform of these two series). As we made clear in Section 2.2, the daily squared return is a very noisy indicator of volatility. Following Andersen and Bollerslev (1998), we therefore use realized volatilities to evaluate the ARCH model predictions. Specifically, let y_t denote one of our three volatility series, *i.e.* s_t^2, s_t and $\ln(s_t^2)$, then we evaluate the corresponding ARCH predictions $\hat{y}_t$, *i.e.* $\hat{\sigma}_t^2$, $\hat{\sigma}_t$ and $\ln(\hat{\sigma}_t^2)$, using the regression:

$$y_t = \alpha + \beta\,\hat{y}_t + \varepsilon_t \tag{2.6.4}$$

If the predictions are unbiased, $\alpha = 0$ and $\beta = 1$. Table 2.4 reports the ordinary least squares estimates of (2.6.4) and the associated R^2 statistics when applied to variances, standard deviations and logarithmic variances. Standard errors using White's (1980) heteroskedasticity correction are in parentheses.

Table 2.4: ARCH Model *Ex Ante* Predictions

	variances			*standard deviations*			*log variances*		
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2
GARCH	0.146	0.526	0.228	0.130	0.682	0.334	-0.443	0.866	0.388
	(0.044)	(0.098)		(0.036)	(0.056)		(0.037)	(0.036)	
FIGARCH	0.128	0.562	0.283	0.110	0.713	0.379	-0.421	0.888	0.424
	(0.057)	(0.123)		(0.040)	(0.062)		(0.035)	(0.035)	
EGARCH	-0.165	1.225	0.518	-0.037	0.953	0.457	-0.366	0.904	0.410
	(0.079)	(0.175)		(0.037)	(0.059)		(0.034)	(0.033)	
FIEGARCH	-0.121	1.163	0.572	-0.008	0.926	0.495	-0.347	0.877	0.444
	(0.059)	(0.136)		(0.032)	(0.051)		(0.032)	(0.030)	

The table reports ordinary least squares coefficient estimates for the model defined by equation 2.6.4 using the ARCH model variance, standard deviation and logarithmic variance predictions. Standard errors using White's (1980) heteroskedasticity correction are in parentheses.

Turning to the results, we can see that the ARCH model volatility predictions are not always unbiased, but all models can capture much of the variation we observe for our three volatility measures. The R^2 statistics range between 22.8% and 57.2%. The FIEGARCH model clearly performs best. For variances and standard deviations the estimates for α and β are roughly within two standard errors of their hypothesized values and, compared to all the other ARCH specifications, the R^2 statistics are highest for all three volatility measures.

Recall that the FIX model employed in our previous section gave unbiased volatility predications and that we obtained for this specification R^2 statistics of 62.7%, 57.6% and 55.1% for variances, standard deviations and logarithmic variances, respectively. Judged by these measures, this realized volatility model clearly improves upon the four ARCH specifications. Yet, the extent of enhancement depends greatly on which formulation is employed. Compared to the standard GARCH model, the FIX model R^2 measures are higher by 39.9, 24.2 and 16.3 percentage points for the respective volatility measure. Compared to the FIEGARCH model, the gains are more modest. The R^2 measures are higher by only 5.5, 8.1 and 10.7 percentage points.

Our result that the realized volatility model performs better is of course only suggestive. There may exist other ARCH models that outperform the models used in this section. Nonetheless, only the FIEGARCH formulation is – in principle – consistent with all the empirical regularities we document. It is therefore doubtful that any other univariate model of the ARCH class could disinter anything more from returns that would be relevant for the prediction of stock return volatility. Moreover, the FIEGARCH model estimates suggest that scaled returns are non-normal and that the volatility process is not covariance stationary – two implications we did not observe using realized volatilities. This perhaps suggests some mis-specification that is confined to the ARCH approach to modeling volatility. An open question remains whether Stochastic Volatility models would perform better. The results throughout this paper suggest that any such formulation would need to account for the long-memory property of volatility. Although it is possible to obtain parameter estimates of fractionally integrated Stochastic Volatility models (*e.g.* Breidt *et al.* 1998), for these type of models one cannot extract volatility predictions from the data.[18] Any comparison along the lines we have pursued is therefore not possible.

2.7 Conclusions

Using 5-minute squared returns on the Dow Jones Industrials Average portfolio over the January 1993 to May 1998 period, we documented the properties of daily stock return volatility. We found that the distributions of variances and standard deviations are skewed-right and leptokurtic, but that logarithmic variances are distributed approximately normal. All three volatility measures are (a) covariance stationary, (b) highly persistent, (c) very little correlated with current returns (no ARCH-M effect) and (d) correlated more strongly with lagged negative than lagged positive returns (news-effect). The news effect can explain some of the asymmetry and flatness of tails in the distribution of the three volatility series – most notably for variances and standard deviations.

We fitted a fractionally integrated model that accounts for the news-effect directly to logarithmic variances. Using *ex ante* one-day-ahead prediction criteria we found that this model yields unbiased and accurate variance, standard deviation and logarithmic variance predictions and that these predictions are better than the ones obtained by the GARCH, FIGARCH, EGARCH and FIEGARCH models. Among these four ARCH specifications, the FIEGARCH formulation performed best. However, the estimate of the fractional integration parameter given by this specification implies that the logarithmic variance process is not covariance stationary. For all ARCH models we found that the distribution of returns divided by the implied standard deviations is leptokurtic. When using realized standard deviations instead, normality of this distribution cannot be rejected.

[18]A survey of Stochastic Volatility models can be found in Ghysels, Harvey and Renault (1996).

Chapter 3

Forecasting Stock Volatility

3.1 Introduction

Shortly after the introduction of the Autoregressive Conditional Heteroskedastic (ARCH) model by Engle (1982), a new and large research area has developed. The class of specifications that emerged has been applied to model the return volatility of virtually every financial asset available. The survey of the literature by Bollerslev, Chou and Kroner (1992) lists already several hundred papers related to this approach. Subsequent surveys include Bera and Higgins (1993), Bollerslev, Engle and Nelson (1994), Palm (1996) and Pagan (1996), among others. The closely related Stochastic Volatility (SV) class of models has engendered its own voluminous literature, see for instance the survey by Ghysels, Harvey and Renault (1996).

A factor fuelling this research area is, without doubt, that many interesting propositions in finance, such as asset and derivative pricing, risk hedging, portfolio selection and risk-management, explicitly depend upon the volatility of future returns. Thus forecasting volatility is of critical importance and ARCH and SV models are, in principle, capable to contribute to this task.

Of considerable interest is whether and to what extend these models proove useful. The investigation of this issue is, however, substantially complicated by the fact that volatility is not a directly observable variable. The standard resolution has been to compare the variance forecasts to squared daily returns. However, even at the one-day-ahead horizon, the correlation between these two variables tends to be disappointingly low (the R^2 is typically 5%). It is therefore not surprising that hardly any attention has been given to the forecasting performance of volatility models when the horizon extends. One exception is the study by West and Cho (1995) who found, as expected, that the performance of conventional models cannot be distinguished from a naive homoskedastic model. This suggests that the usefulness of ARCH and SV models is perhaps limited to one-step-ahead forecasts. Although such short horizon may be sufficient for trading risk management, in many applications, such as option-pricing and other forms of risk management, forecasts over longer horizons are certainly warranted.

Andersen and Bollerslev (1998) give evidence that the apparent poor forecasting performance of the standard models may simply result from the fact that the daily squared return has large measurement error. As an alternative, they suggest a better forecast evaluation measure: the sum of intraday squared returns or the 'realized' volatility. This measure is free of any error, provided that sufficiently many intraday returns are used.[1] In the setting of exchange rates, they found that the correlation between one-day-ahead GARCH variance forecasts and realized variances is substantially higher than when the daily square return is used (the R^2 measures are about 50%).

Our previous chapter suggested that the realized volatility should be viewed not only as a forecast evaluation device, but also as a measure of intrinsic interest. Specifically, we fitted a time series model directly to realized volatility measures and found that the resulting *ex ante* one-day-ahead forecasts are superior to the ones obtained from models of the ARCH

[1]As a convention, and perhaps abusive of language, we shall refer to 'intraday' returns as the sequence of returns within one day; by 'intradaily' we mean this intraday sequence spanning over several days. By 'daily' returns we mean the sequence of returns measured for an entire single day spanning over several days.

class, including those that give proper consideration to the long-run properties in the data (the realized volatility formulation gives an R^2 statistic of more than 60%).

Central to ARCH and SV models is the idea that the squared return process can be decomposed as a product of conditional variances and noise. If such decomposition can indeed be achieved in practice, then this suggests that the realized volatility measure is inefficient. In this case, it might be better to measure volatility using the sum of squares of intraday conditional variances. Furthermore, a model that accomplishes such decomposition should yield superior volatility forecasts.

The central question we seek to answer is which volatility models yield useful forecasts over long-term out-of-sample horizons. Our results in Chapter 2 suggest that at the one day horizon the realized volatility model performs better than formulations of the ARCH class. However, as the realized volatility may be inefficient, this specification may perform worse than one that models the intradaily volatility explicitly. A third approach will therefore be considered, specifically, the direct application of ARCH formulations to intradaily returns. However, moving to higher frequency data, we face the problem of intraday seasonality, *i.e.* that the volatility changes systematically within the day. We explicitly address this problem by extending traditional ARCH models via nonparametric methods.

The empirical analysis of this study is based on five-minute intradaily stock returns on the Dow Jones Industrials Average (DJIA) portfolio over the period extending from January 4, 1993 to August 31, 1999. We consider the forecasting performance of various models for up to 40-days-ahead horizons. The sample is divided into two periods: the first spans from January 4, 1993 to May 29, 1998 and is used for data discovery, estimation and *ex ante* forecast evaluation, while the second runs from June 1, 1998 to August 31, 1999 and is reserved for *ex post* forecast evaluation only.

This chapter is organized as follows. Following this introduction, Section 3.2 presents a brief account of the theoretical underpinnings of the realized volatility measure. Section 3.3

details the construction of the data that provide the basis for the subsequent empirical analysis. Section 3.4 examines the properties of the five-minute data, where we pay particular attention to the seasonality of intraday volatility. In Section 3.5 we develop semi-parametric extensions to long-memory models of the ARCH class to account for both persistence and seasonality in volatility. Section 3.6 presents the properties of our daily return and realized volatility series. In Section 3.7 we apply ARCH models to daily data and in Section 3.8 we present the specifications that model the realized volatility directly. Finally, in Section 3.9 we turn to the *ex ante* and *ex post* multi-day horizon forecasting exercises. Section 3.10 concludes this chapter.

3.2 Realized Volatility Measurement

To set forth the notation, consider dividing a day (or trading period) into N evenly spaced intervals and let $p_{n,t}$ denote the time $n \geq 0$ logarithmic price at day t. The time series process of continuously compounded returns is then defined by:

$$r_{n,t} = p_{n,t} - p_{n-1,t}$$

where $n = 1, \ldots, N$ and $t = 1, \ldots, T$. If $N = 1$, for any series we ignore the first subscript n and thus r_t denotes the time series of daily returns.

We assume that the sequence $r_{n,t}$ has the following properties: (a) it has mean zero, $E[r_{n,t}] = 0$, (b) it is uncorrelated at all leads and lags, $E[r_{n,t}\, r_{m,s}] = 0 \quad \forall\, n, m, s, t$ except when $n = m$ and $s = t$, (c) the variances and covariances of the squares exists and are finite, $E[r_{n,t}^2\, r_{m,s}^2] < \infty$, and (d) it has the representation:

$$r_{n,t} = \sigma_{n,t}\, z_{n,t}$$

where $\sigma_{n,t}^2$ is the variance of $r_{n,t}$ and $z_{n,t}$ is independent of $\sigma_{n,t}^2$ and independent and identically distributed (*i.i.d.*) with $E[z_{n,t}] = 0$ and $E[z_{n,t}^2] = 1$.

For any N, an unbiased estimator of the daily return volatility is:

$$s_t^2 = \sum_{n=1}^{N} r_{n,t}^2$$

as:

$$E\big[s_t^2\big] = \sum_{n=1}^{N} \sigma_{n,t}^2 \, E\big[z_{n,t}^2\big] = \sum_{n=1}^{N} \sigma_{n,t}^2 = \sigma_t^2$$

where σ_t^2 is the return population variance measured at the daily unit.

The variance of s_t^2 is in Chapter 2 derived as:

$$V\big[s_t^2\big] = \frac{\sigma_t^4}{N}\Big(K_{N,t} - 1\Big)\Big(1 + 2\sum_{n=1}^{N-1} \frac{N-n}{N}\rho_{N,n,t}\Big) \tag{3.2.1}$$

where $K_{N,t}$ denotes the kurtosis of $\{r_{n,t}\}_1^N$, and $\rho_{N,n,t}$ the nth autocorrelation of $\{r_{n,t}^2\}_1^N$. The kurtosis and autocorrelations have subscript N as these may change with the number of intraday returns. For any particular value of N, measurement error increases with the daily population variance, with the kurtosis of intraday returns and with the autocorrelations of intraday squared returns.[2]

In Chapter 2 it is shown that s_t^2 converges in mean-square to σ_t^2, *i.e.* $\lim_{N\to\infty} V[s_t^2] = 0$, provided that $\sigma_{n,t}^2$ is stationary and that the discrete time kurtosis of returns is bounded. The latter assumption implies continuity of the sample paths of $\sigma_{n,t}^2$ by the Kolmogorov criterion (Revuz and Yor 1991, Theorem I.1.8).[3] Thus, by increasing the sampling frequency of intradaily returns, it is possible to construct volatility measures that are asymptotically free of any error and, as such, volatility can become an observed variable.

[2]These results have interesting implications for forecast evaluation when the daily squared return is used, $r_t^2 = s_t^2$ when $N = 1$. By (3.2.1), a volatility model should appear to forecast better when it is applied to assets with low return volatility and/or kurtosis.

[3]Consistency may alternatively be established under the assumption that the price process $p_{n,t}$ follows $dp_{n,t} = \sigma_{n,t}\, dW_{n,t}$, where $W_{n,t}$ denotes a Wiener process. Under the assumption that $\sigma_{n,t}$ is continuous, it follows from the results in Karatzas and Shreve (1988, Chapter 1.5) or Barndorff-Nielsen and Shephard (1999) that $\text{plim}_{N\to\infty} \sum_{i=1}^{N} r_{n,t}^2 = \int_1^\infty \sigma_{n,t}^2 \, dn = \sigma_t^2$. See also Andersen *et al.* (1999) who establish consistency in the context of special semi-martingales.

Unfortunately, this convergence result gives little guidance about the actual reduction in measurement error for finite N. However, one may consider some special cases of (3.2.1) that offer some insight. For instance, suppose that $r_{n,t}$ is *i.i.d.*, then equation 3.2.1 reduces to: $V[\,s_t^2\,] = (K_t - 1)\,\sigma_t^4\,/\,N$ (because returns are assumed *i.i.d.*, the kurtosis does not depend on N and this subscript is therefore omitted). Under this scenario, measurement error decreases at rate N. Nonetheless, for various assets it is well documented that the *i.i.d.* assumption does not hold as the squares of returns are found to be correlated (the ARCH effect). Therefore, this expression is expected to give the lower bound of measurement error.

Recall that the results reported thus far are contingent upon the twin assumptions that returns have mean zero and are uncorrelated. The latter assumption becomes particularly questionable when N is large, as serial correlation in returns is a common symptom of market micro-structure effects such as price discreteness, bid-ask bounces and non-synchronous trading (see for instance the textbook treatment by Campbell, Lo and MacKinlay 1997; Chapter 3). The violation of both assumptions can however be easily studied when considering the MA (q) (moving average) representation of $\{r_{n,t}\}_1^N$:

$$r_{n,t} = \mu_{n,t} + \epsilon_{n,t} + \sum_{i=1}^{q} \psi_{i,t}\,\epsilon_{n-i,t} \tag{3.2.2}$$

where the residual $\epsilon_{n,t}$ is assumed to be uncorrelated across all leads and lags. Note that we allow the mean $\mu_{n,t}$ to change either across intraday intervals n (to allow for seasonality within the day), or across days t (to permit periodicity across days), or both while the moving average representation may change across t. Upon squaring (3.2.2), taking expectations, and summing over $n = 1, \ldots, N$, it follows that:

$$E\left[\sum_{n=1}^{N} r_{n,t}^2\right] = E[\,s_t^2\,] = \sum_{n=1}^{N} \mu_{n,t}^2 + (1 + \sum_{1}^{q} \psi_{i,t}^2)\;E\left[\sum_{n=1}^{N} \epsilon_{n,t}^2\right] \tag{3.2.3}$$

where $E[\,\sum_{n=1}^{N} \epsilon_{n,t}^2\,] = \sigma_t^2$. At day t, the cumulative squared returns measure has therefore an additive bias given by the sum of the squared intraday means and a multiplicative bias given by the squared dynamic coefficients of the moving average representation. Under conditions of serial correlation and non-zero means $\mu_{n,t}$, the realized volatility measure therefore unambiguously overestimates actual volatility. One may, of course, test for the statistical significance of the parameters that are used to capture any temporal dependence in returns and use (3.2.3) to determine whether either bias is economically important.

Before proceeding, we would like to note that the cumulative squared returns measure does not require that the return intervals are of equal length throughout the day. This property might prove useful when there is concern about bias that is due to infrequent trading. Under these circumstances, it would be sensible to set the ending mark of each intraday interval once a specific number of trades are accumulated in it. Thus, rather than choosing N *a priori* for the calculation of the realized volatility, one could set some cut-off for the accumulated number of trades. Such a modification should also result in an overall improvement in measurement error provided that volatility is increasing in the number of trades. While the measurement error of the realized volatility increases with actual volatility, it decreases with the number of intervals during the day and this, in turn, increases as the number of trades rises. However, if one is interested in the volatility of an index, where trading frequencies differ across components of the index, such a method becomes more difficult to implement. We shall leave this issue for future research.

3.3 Data Source and Construction

The empirical analysis is based on data from the NYSE Transaction and Quote (TAQ) database which records all trades and quotations for the securities listed on the NYSE, AMEX, NASDAQ and the regional exchanges. Our sample consists of the Dow Jones Industrials Average (DJIA) index constructed from the transaction prices of the 30 stocks that are contained in it. The data span from January 4, 1993 to August 31, 1999 (1683 observations). Within each day, we consider the transaction record extending from 9:30 to 16:05, the time when the number of trades noticeably dropped. Next to transaction prices, volume and time (rounded to the nearest second), TAQ records various codes describing each trade. We used this information to filter trades that were recorded in error and out of time sequence. For a more detailed description of the filtering procedure we refer to Chapter 2.

Taking all 30 stocks together, over our entire sample we observe a trade every 0.9 seconds. Naturally, the trading frequency of the index components is lower. It also varies greatly across stocks: the median time between trades for a single stock ranges from a low of 7 seconds to a high of 40 seconds. This suggests that one should worry that non-synchronous trading induces spurious serial correlation in the return process which, in turn, would render the cumulative squared returns measure biased. Since we are focusing on an index, the market micro-structure effects that are due to price discreteness and bid-ask bounces are of less concern as these tend to wash out in the aggregate (see Gottlieb and Kalay 1985, Ball 1988 and Harris 1990, for instance).

To mitigate the problem of bias, following Andersen and Bollerslev (1998) and Andersen *et al.* (1999), we shall rely on five-minute returns. These are constructed from the logarithmic difference between the prices recorded at or immediately before the five-minute marks. When considering the transactions record extending from 9:30 to 16:05, this provides us with 79 returns for each of the 1683 days, or a total of $132,957$ observations over our sample period.

Of course, it remains an empirical question whether the five-minute cut-off is sufficient large enough so that the problem of bias due to market micro-structure effects and/or non-zero means is of no practical concern. In our next section we shall investigate the properties of our five-minute return series and raise the issue of bias again.

To have meaningful *ex post* forecast evaluations, we divide our sample into two periods. For model estimation, we employ data spanning from January 4, 1993 to May 29, 1998 (1366 observations). For ex post forecast evaluation, we reserve the data extending from June 1, 1998 to August 31, 1999 (317 observations).

3.4 Properties of Intradaily Returns and Volatility

From the results in Section 3.2 it follows that the cumulative squared returns measure will be biased if five-minute returns are correlated and have non-zero mean values that may vary within the day. In this section we shall first examine the properties of intradaily returns to determine whether and to what extent such bias is relevant. If there is indeed evidence for temporal dependence in the conditional mean of five-minute returns, this cannot be ignored when modeling the conditional variance. As such, the study of the level returns will help us to define the ARCH models that we shall apply to five-minute data.

In subsection 3.4.2, we turn to the properties of intraday volatility. As we shall see shortly, our data display strong seasonal patterns. This is a well-known characteristic of U.S. stock markets that has been documented in Wood *et al.* (1985), Harris (1986), Lockwood and Linn (1990) among others. Despite these findings, this characteristic has been largely ignored in the ARCH and Stochastic Volatility literature. Intuitively it is clear that one cannot properly account for the temporal dependence in volatility when one takes no notice of the seasonal relation of volatility within the day. Not surprisingly, the straightforward application of standard volatility models has given puzzling results. For instance, Kroner (1994) and Guillaume (1994) find that the GARCH model can imply less volatility persistence when applied to high frequency rather than low frequency returns. This defies the aggregation results developed in Nelson (1990, 1992), Drost and Nijman (1993) and Drost and Werker (1996).

There have nonetheless been attempts to correct for intraday seasonality. In the exchange rate literature, seasonality in intradaily data is considered by Müller *et al.* (1990) and Dacorogna *et al.* (1993). More recently, Andersen and Bollerslev (1997) provide a general procedure that is applicable to both exchange rates and stocks. Their approach uses a two-step filtering procedure for removing the intraday volatility seasonality from returns.

In the third step, volatility models are applied to the filtered data.

Unfortunately, however promising, filtering procedures inadvertently change the character of the data. For this reason, we shall in this paper follow a different approach, specifically, modeling simultaneously the persistence and seasonality of volatility. Such a one-step procedure is not only more economical but also leaves the returns series intact.

A critical feature of the one-step estimation method is that it assumes that the return process can be decomposed as a product of three components: deterministic intraday volatility, seasonal-free volatility and *i.i.d.* noise. To determine whether such decomposition is in fact warranted, we propose in subsection 3.4.3 a simple one-step filtering procedure that identifies the deterministic component using minimal assumptions. This allows us to construct a filtered return series and then study whether the volatility characteristics of it are free of any problems that may be related to seasonality. However, given the aforementioned problems of filtering data, we shall not apply volatility models to returns constructed in this manner. Nonetheless, as our one-step estimation procedure relies on the idea of a tripartite decomposition, it will be useful, as a pre-specification exercise, to undertake this filtering approach.

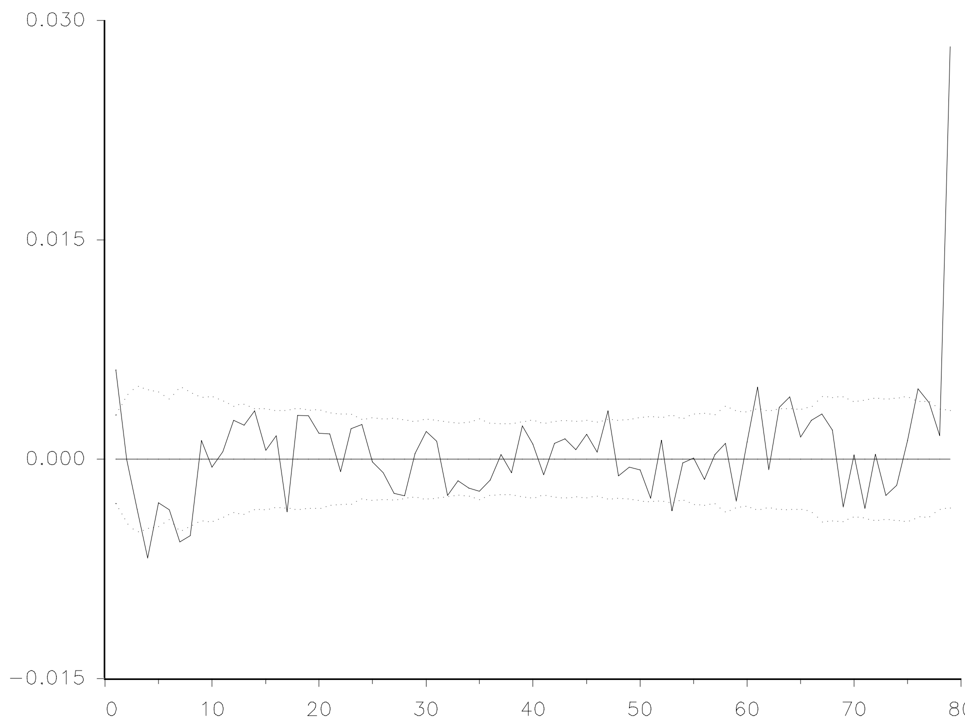
3.4.1 Seasonality of Intraday Returns

To investigate the intraday seasonality of returns, Figure 3.1 plots the average sample mean of five-minute returns, $\bar{r}_n = T^{-1}\sum_{t=1}^{T} r_{n,t}$ for $n = 1, \ldots, 79$ and $T = 1366$, along with the 1.96 standard error bands of the estimates under the assumption that $\bar{r}_n = 0$.[4] If returns are distributed normal, these bands represent therefore the five percent confidence interval for the hypotheses that the mean estimates are zero. As the distribution of returns is known to be leptokurtic, the displayed intervals are likely too tight, however. Taking this into consideration – and ignoring the first and last five-minute mean observations – one would possibly not reject the hypotheses that the estimates are significantly different from zero

[4]Assuming that the sequence $\{r_{n,t}\}_{t=1}^{T}$ has mean zero and is uncorrelated, the variance of $\bar{r}_n$ equals $\bar{s}_n^2/T$ where $\bar{s}_n^2 = T^{-1}\sum_{t=1}^{T} r_{n,t}^2$ (see also equation 3.4.1 later in this section).

at the 5% level. Perhaps this conclusion would remain for the first mean estimate as well, but certainly not for the last value displayed in Figure 3.1. By Chebyshev's inequality, it is significantly different from zero with a probability not greater than 0.370%.[5]

Figure 3.1: Intraday Average Returns



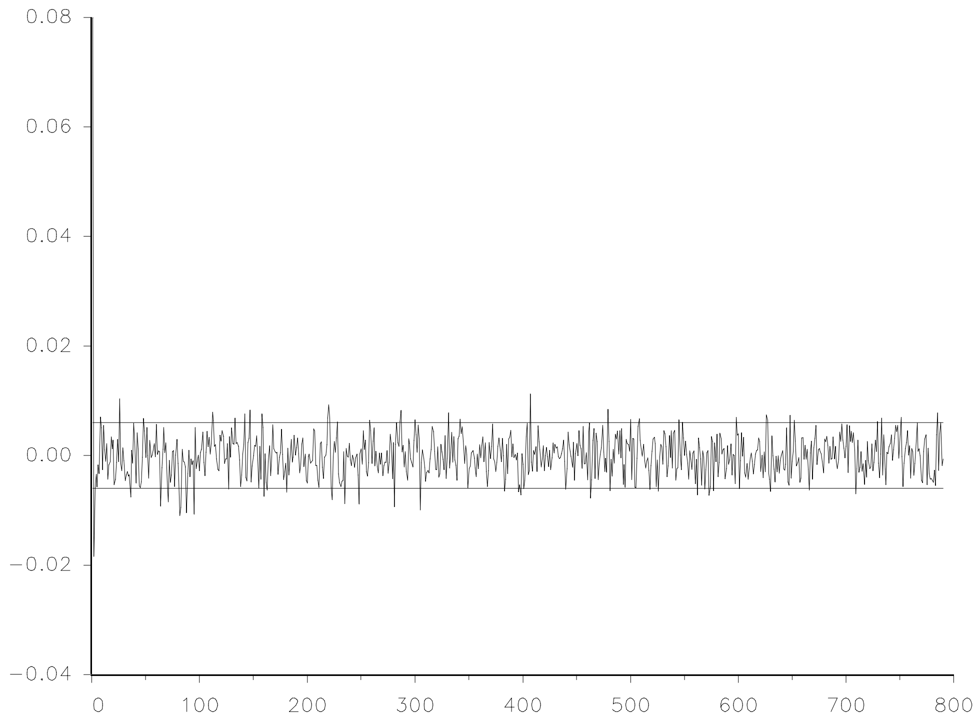
The solid line displays the 79 consecutive average five-minute returns over the time extending from 9:30 to 16:05. The doted lines represent the asymptotic five percent confidence interval for the hypotheses that the mean estimates are zero. The sample period is January 4, 1993 to May 29, 1998.

Figure 3.2 displays the 790 sample autocorrelations of the five-minute returns for up 10 days. The first two values equal 0.080 and −0.018 and are significant judged by the $\pm 1.96\sqrt{(1/N\,T)}$ 5% confidence interval. Consistent with the spurious dependencies that would be induced in an index by non-synchronous trading, the first order autocorrelation is positive. Beyond the first few lags however, the realizations resemble white noise. Excluding

[5]Using the intraday mean return estimates, we found it interesting to determine by what percentage the DJIA increased over certain times of day. In the period extending from January 4, 1993 to May 29, 1998 it increased between 9:30 and 9:35 by 9.393%, between 9:35 and 16:00 by 14.257% and between 16:00 and 16:05 by 39.685%. Thus the index increased, on average, more during the last five minutes of trading than during the previous six and a half hours.

the first two values, 8.73% of the autocorrelations are significant at the 5% level. Although not all violations may be attributed to Type II error of the test, one should keep in mind that the chosen confidence interval is asymptotic when the innovations are not *i.i.d.*

Figure 3.2: Intradaily Return Sample Autocorrelation Function



The graph displays the first 790 sample autocorrelations of five-minute returns. The first and second autocorrelations are 0.080 and −0.018, respectively. The horizontal lines are the 95% confidence intervals, $\pm 1.96/\sqrt{NT}$. The sample period is January 4, 1993 to May 29, 1998.

As the modeling and measurement of volatility is concerned, the main findings of this subsection are that the mean estimates of 9:30 to 9:35 and 16:00 to 16:05 returns are sizable and that the return process is serially correlated. However, for the realized volatility calculations the consequences are minimal. When estimating the MA model defined by equation 3.2.2, allowing for a non-zero mean in the first and last five-minute returns and the first two moving average parameters, we obtain $\hat{\mu}_1 = 0.0048$, $\hat{\mu}_{79} = 0.0293$, $\hat{\psi}_1 = 0.0426$ and $\hat{\psi}_2 = -0.0321$. From (3.2.2) it follows that the bias resulting from non-zero mean estimates is only 0.0009, while the bias emerging from serial correlation scales the volatility estimates upward by a factor of merely 1.0028.[6] Considering that the mean realized variance equals 0.4166, we view these biases too small to be of any economic significance. Thus, we shall not employ any corrections by imposing a structure on the conditional mean dynamics of returns.

As the estimation of ARCH models is concerned, we shall explicitly allow for the possibility that returns are correlated and that some of the five-minute returns may have a mean different from zero by modeling simultaneously the conditional mean and volatility of returns. From these results, as we shall see, it will follow that the expected bias in the cumulative squared returns measure is even smaller than the above analysis suggests.

3.4.2 Seasonality of Intraday Volatility

The fluctuations in intraday volatility become evident in Figure 3.3 where we plot average five-minute variances, defined as:

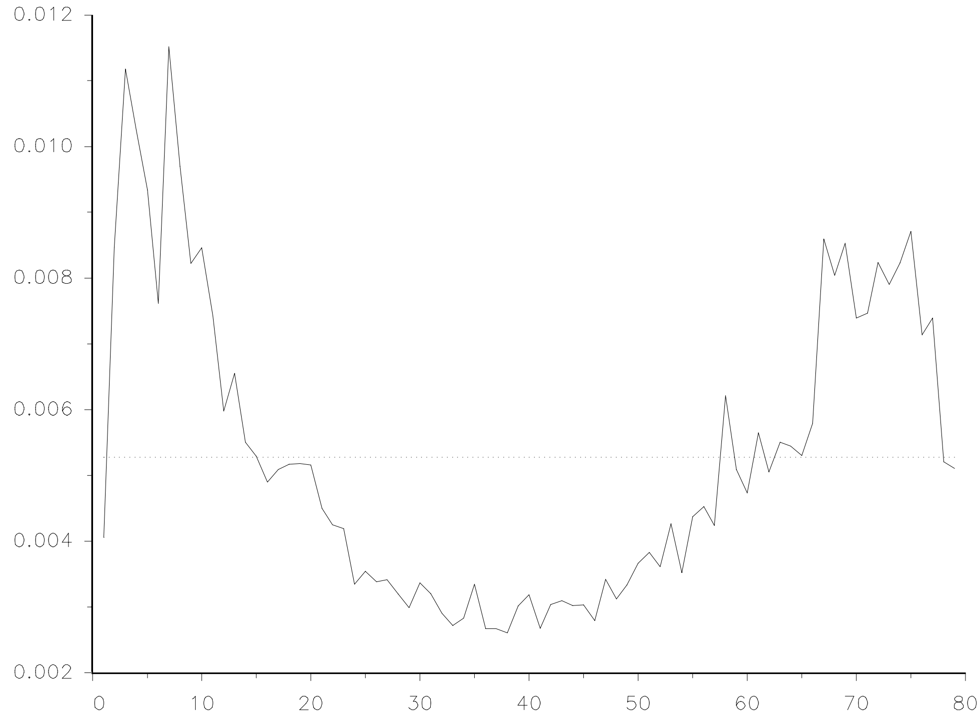
$$\bar{s}_n^2 = T^{-1} \sum_{t=1}^{T} r_{n,t}^2 \qquad n = 1, \ldots, N \tag{3.4.1}$$

for $N = 79$ and $T = 1366$. The horizontal line gives the average value of (3.4.1) over all N.

Each estimate $\bar{s}_n^2$ may be viewed as the *realized* average five-minute variance, as we take the mean of nth day squared returns. The properties of the estimator follow therefore readily from the results in Section 3.2. It will be unbiased if the nth value of the five-minute returns has mean zero and if all T of these values are uncorrelated. Measurement error will now depend on T, the autocorrelations of the nth five-minute squared returns and the kurtosis of nth five-minute returns. As above, the degree of biases stemming from the violations of the assumptions made for the conditional mean can be assessed using (3.2.3). We shall however ignore this issue as any bias is, again, found to be negligible.

[6] We obtained these results by conditioning the residual innovations on Student's t density to account for the excess kurtosis in returns. When the normal density is used instead, bias resulting from non-zero five-minute means is 0.0008 while bias due to serial correlation is 1.0069.

Figure 3.3: Intraday Average Variances



The solid line displays the 79 average variance estimates of five-minute returns over the time extending from 9:30 to 16:05 (see equation 3.4.1). The dotted horizontal line gives the mean of these values. The sample period is January 4, 1993 to May 29, 1998.

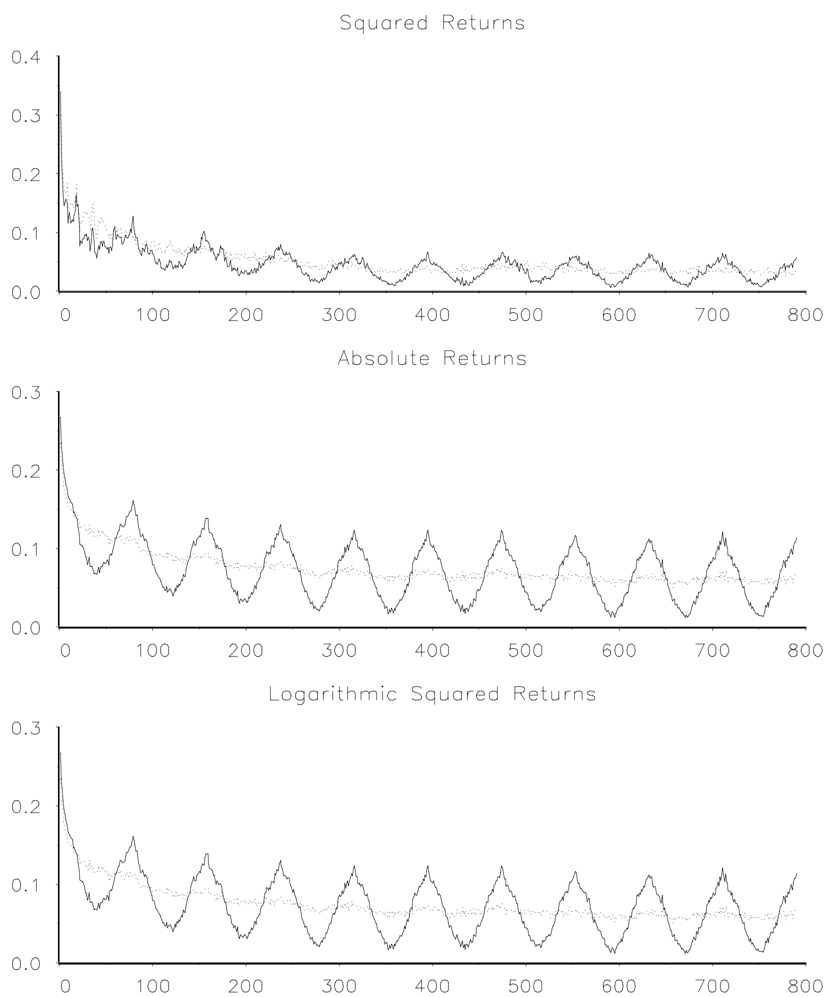
As it has been documented by Wood *et al.* (1985), Harris (1986) and others, volatility follows *roughly* a U-shaped pattern: it is high during the first hour of trading, reaches its low around mid-day and is again high during the last trading hour. Noticeable exceptions do however occur: there is an increase in volatility at the very beginning of the day, volatility dips at about 10:00 and 15:30 and one observes a fairly steady decline of volatility starting at 15:45 (the 75th value).

The solid line in Figure 3.4 plots the first 790 sample autocorrelations of five-minute squared returns, $r_{n,t}^2$ (upper panel), absolute returns, $|r_{n,t}|$ (middle panel), and logarithmic squared returns, $\ln(r_{n,t}^2)$ (lower panel).[7] We can see that over every subsequent set of 79 correlations

[7] To circumvent the *inlier* problem that arises when taking the logarithm of near zero or zero squared returns, we employ the transformation suggested by Fuller (1996): $\tilde{\ln}(r_{n,t}^2) = \ln(r_{n,t}^2 + \gamma\, s^2) - \gamma\, s^2/(r_{n,t}^2 + \gamma\, s^2)$, where s^2 is the sample variance of $r_{n,t}^2$ and, as in Fuller, $\gamma = 0.02$.

a U-shaped pattern reoccurs. When looking at the first 79 autocorrelations, for instance, we can notice a peak at the first five-minute offset, a low at roughly the 40th five-minute offset, and again a peak at 79th five-minute offset. Although the amplitude of each cycle appears to remain the same, the decay in the autocorrelations becomes evident in that subsequent cycles are at lower levels.

Figure 3.4: Intradaily Volatility Sample Autocorrelation Functions



The solid lines display the first 790 sample autocorrelations of five-minute 'raw' squared returns, absolute returns and logarithmic squared returns. The dotted lines give the sample autocorrelations when the returns are filtered according to the procedure described in Section 3.4.3. The sample period is January 4, 1993 to May 29, 1998.

As the measurement of the realized volatility is concerned, the intraday volatility pattern documented above is of little concern. While measurement error depends on the autocorrelations in squared returns, the quality of our measure is not compromised when these correlations follow specific patterns. The pronounced systematic fluctuations in the sample autocorrelation function of intradaily squared, absolute, or logarithmic squared returns give however an initial indication that it would probably be hazardous to apply the standard volatility models to intradaily data. Neither models of geometric decay nor the ones that allow for hyperbolic decay can accommodate the strong cyclical patterns that the sample autocorrelation functions display.

3.4.3 Correcting for Intraday Volatility Seasonality

We shall assume that intradaily returns have the representation:

$$r_{n,t} = \sigma_{n,t}\, z_{n,t} = b_n\, \tilde{\sigma}_{n,t}\, z_{n,t} \tag{3.4.2}$$

where $z_{n,t}$ is *i.i.d.*, mean zero and unit variance and the volatility process $\sigma_{n,t}$ is the product of a deterministic seasonal component b_n (that varies within the day, but not from day to day) and the intradaily volatility $\tilde{\sigma}_{n,t}$ that is seasonal-free.[8,9] All three return components are assumed to be independent and the volatility variables are nonnegative, *i.e.* $\sigma_{t,n} \geq 0$ and $b_n \geq 0$.

In our next section we develop models that simultaneously identify all three components in (3.4.2). In this section we shall propose a simple procedure that filters from returns the seasonal effects that we observe in intraday volatility. This allows us to study, at an initial stage, whether the above setup is useful.

[8]Note that b_n is a vector of length N x T where the first N elements are stacked T times.

[9]Andersen and Bollerslev (1997) consider the more general setup that specifies $b_{n,t}$ rather than b_n – thus the seasonal component is allowed to vary from day to day. However, at the filtering stage it is assumed that: $r_{n,t} = b_{n,t}\, \sigma_t\, z_{n,t}$, where the daily volatility σ_t has to be estimated (in their application using a GARCH (1,1) model).

Upon squaring (3.4.2), averaging over all t, and taking expectations it follows that:

$$E\left[\frac{1}{T}\sum_{t=1}^{T} r_{n,t}^2\right] = \frac{1}{T}\sum_{t=1}^{T}\sigma_{n,t}^2 = b_n^2\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\sigma}_{n,t}^2\right] \qquad n = 1,\ldots,N \tag{3.4.3}$$

where the term on the left equals, by (3.4.1), $E[\bar{s}_n^2]$: the expectation of the average nth day interval variance. The factor in brackets on the right is the average seasonal-free nth interval variance. Because it is seasonal-free, we shall *assume* that it will equal the unconditional variance at the five-minute unit, *i.e.* $(N\,T)^{-1}\sum_{t=1}^{T}\sigma_t^2$. Graphically, the seasonal-free representation of intraday volatility in Figure 3.3 would be a horizontal line at $(N\,T)^{-1}\sum_{t=1}^{T}\sigma_t^2$.

As the sample analogue to $(N\,T)^{-1}\sum_{t=1}^{T}\sigma_t^2$ is $(N\,T)^{-1}\sum_{t=1}^{T}s_t^2 = N^{-1}\bar{s}^2$, this then suggests the following estimator for b_n:

$$\hat{b}_n = \sqrt{\frac{\bar{s}_n^2}{\bar{s}^2/N}} \qquad n = 1,\ldots,N \tag{3.4.4}$$

Thus the deterministic seasonal components b_n may be estimated as the square-root of the average nth day interval variance over the average variance at the nth day unit.

Having obtained all N estimates for the deterministic seasonal component b_n, one might then construct the pseudo-return series:

$$\tilde{r}_{n,t} = \frac{r_{n,t}}{\hat{b}_n} \tag{3.4.5}$$

which, if the representation of returns in (3.4.2) is correct, should not display any seasonal volatility patterns.

Note that by construction, $N^{-1}\sum_{n=1}^{N}\hat{b}_n^2 = (\sum_{n=1}^{N}\bar{s}_n^2)/\bar{s}^2 = 1$ (the sum of average five-minute variances equals the daily mean variance) and therefore it holds that on *on average* the sum of squared pseudo-returns, $\sum_{n=1}^{N} r_{n,t}^2/\hat{b}_n$, will equal the sum of squared returns

$\sum_{n=1}^{N} r_{n,t}^2$ (the realized volatility). Unless, trivially, $b_n = 1\ \forall N$, this however is not necessary true for each of the T days.[10] It will also not hold that the sum of returns, absolute or logarithmic squared pseudo-returns are, on average, equal their unadjusted counterparts. This arises since $N^{-1}\sum_{n=1}^{N} b_n \neq 1$, $N^{-1}\sum_{n=1}^{N} |b_n| \neq 1$, and $N^{-1}\sum_{n=1}^{N} ln(b_n^2) \neq 1$, unless the trivial case holds. Of course, one may always find some normalization of b_n so that the above three equations hold with equality, but this again does not imply that the intraday aggregates will be equal for each day. As such, it becomes evident that this type of filtering procedure – and all others for that matter – ultimately changes the return process in various ways. Clearly then, one could not comfortably compare the accumulated variance predictions of a model applied to filtered returns to the realized volatility measure which is obtained from raw data. It is mainly for this reason that we in our next section employ a formulation that leaves returns unaltered.

To determine, at an initial stage, whether the above representation of returns in (3.4.2) is useful, we plot in Figure 3.4 in dotted lines the autocorrelation functions of five-minute squared pseudo-returns, $\tilde{r}_{n,t}^2$, absolute returns, $|\tilde{r}_{n,t}|$, and logarithmic squared returns, $\ln(\tilde{r}_{n,t}^2)$. Upon comparing the results from filtered to unfiltered returns, it becomes evident that the correlation structure of five-minute volatilities display a much more coherent pattern than before. Specifically, the slow hyperbolic decay indicates the presence of long-memory. Using daily return data, this phenomenon has been documented by Ding, Engle and Granger (1993) and Crato and de Lima (1994), among others. Close inspection of the filter-based autocorrelations indicates, however, that perhaps some seasonality remains. Nonetheless, the cyclical behavior that remains is relatively small to what is observed when the raw returns are used.

[10]Consider, for instance, the following $N = 2$ and $T = 2$ senario: $r_{1,1}^2 = 1$, $r_{2,1}^2 = 2$, $r_{1,2}^2 = 3$ and $r_{2,2}^2 = 4$. Thus, $\bar{s}_1^2 = 2$, $\bar{s}_2^2 = 3$ and $\bar{s} = 5$. Form these values one obtains $\hat{b}_1^2 = 2/5$ and $\hat{b}_2^2 = 3/5$. But, $r_{1,1}^2 + r_{2,1}^2 = 3 \neq 15/2 = \tilde{r}_{1,1}^2 + \tilde{r}_{2,1}^2$.

3.5 Intradaily ARCH Modeling

The most popular approach to measuring and modeling volatility is to fit specifications of the ARCH class to returns. Since the introduction of the ARCH model by Engle (1982), various extensions have been proposed. While it has been extensively documented that the volatility of asset returns displays long-range dependence, only two ARCH-type specifications have been put forward that attempt to properly account for this. One is the FIGARCH model by Baille, Bollerslev and Mikkelsen (1996) and the other is the FIEGARCH model by Bollerslev and Mikkelsen (1996). As the correct specification of volatility dependencies become especially important when one focuses on long-term horizon forecasts, we concentrate on these two models. Nonetheless, we will also consider the short-memory specifications that are nested in them (Bollerslev's (1986) GARCH model and the EGARCH model in Nelson (1991)) in order to determine what is gained.

In this section we consider the FIGARCH and FIEGARCH specifications in the context that they are applied to intradaily data. Our previous section indicated that it would likely be hazardous to apply the standard formulation of these models directly to our five-minute data. We therefore propose extensions that account for the inherent intraday seasonality.

In order to accommodate temporal dependencies in the conditional mean of returns, we shall consider the return representation:

$$\big(1+\phi(L_{pl})\big)\, r_{n,t} = \sum_{i=1}^{N} \mu_n \, \mathrm{I}_n + \big(1+\psi(L_{ql})\big)\, \varepsilon_{n,t}$$

$$\varepsilon_{n,t} = \sigma_{n,t}\, z_{n,t} \tag{3.5.1}$$

where it is assumed that $E[z_{n,t}] = 0$ and $E[z_{n,t}^2] = 1$. The indicator I_n is one when $n = 1, \ldots, N$ and zero otherwise. Throughout this paper, L denotes the lag operator and a polynomial of the form $a(L_l)$ is defined by $\sum_{i=1}^{l} a_i\, L^i$. The above representation of returns thus allows for non-zero nth five-minute means and for serial correlation in the intradaily return process.

For the FIGARCH$\,(p,d,q)$ model, we consider the conditional variance process:

$$\sigma^2_{n,t} = \frac{\exp\big(\sum_{j=1}^{J} \omega_j \, x_{j,(n,t)}\big)}{a} \, \tilde{\sigma}^2_{n,t}$$

$$\tilde{\sigma}^2_{n,t} = \frac{\omega_0 + \Big[\big(1-\beta(L_p)\big) - \big(1-\alpha(L_q)-\beta(L_p)\big)\,\big(1-L\big)^d\Big]\,\varepsilon^2_{n,t}}{\big(1-\beta(L_p)\big)} \tag{3.5.2}$$

where a is a normalizing constant and $x_{j,(n,t)}$, $j = 1, ..., J$ is a set of exogenous variables to be defined later. The FIGARCH model is covariance stationary only in the special case when $d = 0$, whereby it reduces to the GARCH specification. The model displays however the important property of having a bounded cumulative impulse-response function for any $0 \leq d < 1$. As in Bollerslev (1987), we condition the innovations $z_{n,t}$ on the standardized Student's t density, *i.e.* $z_{n,t} \sim T(0, 1, \eta_1)$. This density has thicker tails than the normal when $\eta_1 < \infty$.

We consider for the FIEGARCH (p,d,q) model the following conditional variance process:

$$\ln(\sigma^2_{n,t}) = \omega_0 + \sum_{j=1}^{J} \omega_j \, x_{j,(n,t)} + \frac{\alpha(L_q)\,\big(\gamma\, z_{n,t} + |z_{n,t}| - \mathrm{E}\big[\,|z_{n,t}|\,\big]\big)}{\big(1-L\big)^d\,\big(1-\beta(L_p)\big)} \tag{3.5.3}$$

where, again, $x_{j,(n,t)}$, $j = 1, ...J$ is a set of exogenous variables. As in the standard ARFIMA model that allows for fractional integration in the conditional mean, for $-0.5 < d < 0.5$, $\ln(\sigma^2_{n,t})$ is covariance stationary and for $0 < d < 1$, the process is mean-reverting and said to have long-memory. When $d = 0$, the model reduces to the EGARCH specification of Nelson (1991). As in that formulation, we condition the innovations $z_{n,t}$ on the density of the generalized error distribution, *i.e.* $z_{n,t} \sim GED(0, 1, \eta_2)$. The density is normal when $\eta_2 = 2$, while it displays heavy tails for $\eta_2 < 2$.

A feature commonly found for equities is that past negative returns tend to have a bigger impact on future volatility than positive returns of the same magnitude (see for instance

Black 1976, Pagan and Schwert 1990 and Engle and Ng 1993). This phenomenon is often called the 'leverage' or 'news' effect. The EGARCH and FIEGARCH models allow for this asymmetry of volatility in past returns if $\gamma < 0$.

Note that the exponential of the exogenous variables in (3.5.2) and (3.5.3) will scale the conditional variances up or down. The exponential transformation was chosen so that the variance process will be strictly non-negative without imposing additional restrictions.

To account for the apparent seasonality that would exactly comply with the representation of returns set forth in equation 3.4.2, one could specify that $x_{j,(n,t)} = \mathrm{I}_j$ for $j = 1, ..., N-1$ where I_j would take value one at the jth five-minute mark and zero otherwise (the Nth value is left out as the models contain the constant ω_0). As the estimation of $N-1$ additional variables is computationally demanding, we instead account for the intraday seasonality using the following set of variables:

$$\sum_{j=1}^{J} \omega_j \, x_{j,(n,t)} = \omega_1 \left(\mathrm{I}_1 - \frac{1}{N} \right) + \omega_2 \left(\mathrm{I}_N - \frac{1}{N} \right) + \omega_3 \left(\frac{n\,N}{\sum_{i=1}^{N} i} - 1 \right) + \sum_{k=1}^{K} \omega_{4+2(k-1)} \sin\left(\frac{2\,k\,\pi\,n}{N} \right) + \omega_{5+2(k-1)} \cos\left(\frac{2\,k\,\pi\,n}{N} \right) \tag{3.5.4}$$

where $n = 1, \ldots, N$ and $N = 79$ for our five-minute data. The various normalizations in equation 3.5.4 guarantee that this expression has mean zero over $n = 1, \ldots, N$.

The sum of trigonometric terms in (3.5.4) is the non-parametric Flexible Fourier Form (FFF) of Gallant (1981).[11] In theory, K must go to infinity to approximate any functional form, however we found that it was not worthwhile (in terms of significance) to go above $K = 2$ once we allowed for the additional terms in (3.5.4): two deterministic variables set to one at either the first or last intraday mark (w_1 and w_2) as well as trend component (w_3).

[11] Andersen and Bollerslev (1997) use as well the FFF specification to approximate the intraday volatility pattern while Dacorogna *et al.* (1993) using polynomials that correspond to the geographical locations of exchange markets. In a different context, Pagan and Schwert (1990) apply the FFF formulation to daily returns in order to account for the leverage effect.

Whether this extension to the long-memory ARCH models will indeed help us to better identify the volatility process is a question that we shall answer in our forecasting exercises.

In the context of forecasting volatility over multiple days, some comments and modifications to the FIGARCH model are in order. From Baille, Bollerslev and Mikkelsen (1996) it follows that when $0 < d \leq 1$, non-stationarity of the model stems from the fact that that the coefficients ψ_i in the infinite binominal expansion of $(1-L)^d = 1 - \sum_{i=1}^{\infty} \psi_i L^i$ sum to unity. When $0 < d < 0$ it is in practice necessary to truncate $(1-L)^d$, say at lag m. In this case, however, $\sum_{i=1}^{\infty} \psi_i < 1$. For instance, when $d = 0.375$ one obtains $\sum_{i=1}^{1,000} \psi_i = 0.948$, $\sum_{i=1}^{2,000} \psi_i = 0.960$, and $\sum_{i=1}^{10,000} \psi_i = 0.978$. The process one identifies when estimating the model is therefore stationary. Let $a_1, ..., a_u$ denote the dynamic coefficients in the denominator of the conditional variance process (3.5.2), where $u = \max\{m, q\}$. As in the standard GARCH model, it therefore should hold that $w_0/(1-\beta_1, \ldots, -\beta_p - a_1 - \ldots, -a_u) = (TN)^{-1} \sum_{t=1}^{T} \sum_{n=1}^{N} r_{n,t}$. That is, the constant estimate over one minus the sum of dynamic coefficients should equal the unconditional variance of the process. This however was not the case in our application and as such the forecasts of the model converged to a constant that was *substantially* different from the unconditional variance. To circumvent the problem, we therefore set w_0 such that the implied unconditional variance equals the mean variance of returns. Interestingly, this hardly altered the maximized log-likelihood, yet improved the forecasting performance of the model in a pronounced manner. We should note that on the basis of our forecast evaluation criteria, this model would otherwise have been the worst performing specification.

To forecast variances over s days, *i.e.* $s\,N$ five-minute intervals, it is for the FIEGARCH model necessary to obtain:

$$E\Big[e^{a_1\, g(z_{t+s,N-1}) + a_2\, g(z_{t+s,N-2}) + \cdots + a_{sN-1}\, g(z_{t+1,1})} \,|\, t\,N\Big] \tag{3.5.5}$$

where $g(z_{n,t}) = \gamma\, z_{n,t} + |z_{n,t}| - \mathrm{E}[\,|z_{n,t}|\,]$ and a_j is the jth coefficient of the moving average representation. The solution to (3.5.5) is in Nelson (1991; Appendix I) and thus not reproduced here. We obtain better in-sample forecasting results when calculating the required moments using the sample analogs to (3.5.5), *i.e.* the mean value of the term in brackets using the in-sample $z_{n,t}$ residuals. We therefore report these results.

Finally, as we shall see shortly, for some models we obtaine significant estimates for the conditional mean parameters. As our realized volatility measure ignores these, we adjusted the ARCH model forecast accordingly using the results in equation 3.2.3. For instance, when we find that μ_n and ψ_1 are significant, we calculate the variance forecast for time n, t as $(1 + \hat{\psi}_1^2)\, \hat{\sigma}_{n,t}^2 + \hat{\mu}_n^2$, where $\hat{\sigma}_{n,t}^2$ is the unadjusted variance forecast.

Maximum likelihood estimates of the GARCH (1,2), FIGARCH (1,d,1), EGARCH (1,2) and FIEGARCH (1,d,2) models are presented in Table 3.1. We employ the acronym 'SP' (Semi-Parametric) as a prefix to distinguish those versions of these models that account for intraday seasonality. Standard errors, based on the matrix of second derivatives of the log-likelihood function, are in parentheses. All reported estimates are significant at the 5% level on the basis of either Wald or log-likelihood ratio tests. $\mathfrak{L}^*$ reports the maximized log-likelihood.

We can see that the size of the conditional mean estimates ($\hat{\mu}_1$, $\hat{\mu}_{79}$ and $\hat{\psi}_1$) varies across models. While we universally obtaine large and significant estimates for the 79th intraday mean, μ_{79}, we find that the first mean, μ_1, is significant only in the GARCH, FIGARCH and EGARCH models (and hence in none of the semi-parametric specifications). In all models we do not obtain significant moving average parameters beyond the first order, while in the SP-FIEGARCH model no moving average coefficient is found significant. In Section 3.4 we reported $\hat{\mu}_1 = 0.0048$, $\hat{\mu}_{79} = 0.0293$, $\hat{\psi}_1 = 0.0426$ and $\hat{\psi}_2 = -0.0321$ under the assumption that the variance process is homoskedastic. All of these coefficients exceed the size of the ones reported in Table 3.1. Thus, one may conclude that some of the conditional mean dependence that was apparent earlier is in fact driven by changing variances.

Table 3.1: Intradaily ARCH Model Estimates

	GARCH		*FIGARCH*		*EGARCH*		*FIEGARCH*	
	—	*SP*	—	*SP*	—	*SP*	—	*SP*
$\hat{\mu}_1$	0.004		0.003		0.003			
	(0.002)		(0.002)		(0.002)			
$\hat{\mu}_{79}$	0.029	0.023	0.031	0.028	0.027	0.021	0.026	0.019
	(0.008)	(0.009)	(0.009)	(0.012)	(0.008)	(0.009)	(0.010)	(0.010)
$\hat{\psi}_1$	0.041	0.039	0.039	0.029	0.042	0.043	0.033	
	(0.009)	(0.011)	(0.009)	(0.009)	(0.008)	(0.018)	(0.016)	
$\hat{\omega}_1$		-0.893		-0.906		-0.861		-0.884
		(0.052)		(0.055)		(0.051)		(0.051)
$\hat{\omega}_2$		0.164		0.183		0.129		0.144
		(0.049)		(0.096)		(0.051)		(0.051)
$\hat{\omega}_3$		-0.351		-0.357		-0.347		-0.352
		(0.020)		(0.023)		(0.023)		(0.023)
$\hat{\omega}_4$		0.454		0.410		0.570		0.566
		(0.010)		(0.009)		(0.009)		(0.010)
$\hat{\omega}_5$		-0.371		-0.319		-0.166		-0.169
		(0.015)		(0.016)		(0.017)		(0.018)
$\hat{\omega}_6$		0.066		0.066		0.065		0.067
		(0.008)		(0.008)		(0.008)		(0.009)
$\hat{\omega}_7$		-0.070		-0.079		-0.051		-0.051
		(0.009)		(0.010)		(0.011)		(0.011)
$\hat{\omega}_0$					-5.513	-5.543	-5.509	-5.519
					(0.018)	(0.028)	(0.058)	(0.116)
$\hat{\beta}_1$	0.856	0.948	0.420	0.568	0.960	0.991	0.856	0.785
	(0.003)	(0.002)	(0.023)	(0.020)	(0.001)	(0.001)	(0.008)	(0.031)
$\hat{d}$			0.326	0.300			0.319	0.499
			(0.007)	(0.007)			(0.013)	(0.018)
$\hat{\alpha}_1$	0.150	0.119	-0.180	-0.179	0.308	0.238	0.296	0.229
	(0.005)	(0.005)	(0.009)	(0.009)	(0.008)	(0.008	(0.008)	(0.008)
$\hat{\alpha}_2$	-0.034	-0.074			-0.055	-0.130)	-0.123	-0.137
	(0.005)	(0.005)			(0.008)	(0.008	(0.009)	(0.013)
$\hat{\gamma}$					-0.096	-0.166	-0.105	-0.178
					(0.009)	(0.013)	(0.010)	(0.013)
$\hat{\eta}_{1,2}$	5.856	6.486	5.977	6.448	1.287	1.371	1.315	1.390
	(0.086)	(0.111)	(0.091)	(0.111)	(0.007)	(0.008)	(0.008)	(0.008)
$\mathfrak{L}^*$	147604	149031	147744	149220	147294	148790	147668	149096

The table reports conditional maximum likelihood estimates of the models defined by equations 3.5.1 to 3.5.4. The acronym 'SP' distinguishes the models that account for intraday seasonality. Estimates for η carry suffix 1 when the innovations $z_{n,t}$ are conditioned on Student's t density and suffix 2 when the generalized error density is used instead. Standard errors, based on the second derivatives of the log-likelihood function, are reported in parentheses. $\mathfrak{L}^*$ reports the maximized log-likelihood. The $(1-L)^d$ polynomial, when used, is truncated at lag 7900. The sample period is January 4, 1993 to May 29, 1998.

The seven coefficients accounting for intraday seasonality, ω_1 to ω_7, are jointly overwhelmingly significant on the basis of the log-likelihood ratio test. Comparing each pair of GARCH, FIGARCH, EGARCH and FIEGARCH specifications, the maximized log-likelihood, $\mathfrak{L}^*$, increases by 1428, 1475, 1495 and 1427, respectively.

The dynamic properties of the models change greatly when accounting for the cyclical behavior of intraday volatility. This is perhaps best illustrated, in a natural way, when considering the autoregressive representation of the EGARCH and FIEGARCH models. For the FIEGARCH model we find that the implied end-of-day autocorrelation for the first day (79th offset) equals 0.4936, for the fifth day (395th offset) 0.0284 and for tenth day (790th offset) 0.0008. This model therefore clearly accounts for some of the persistence found earlier in the sample autocorrelation function of logarithmic intradaily squared returns (see Figure 3.4). The corresponding autocorrelations of the SP-FIEGARCH are however much higher: 0.6255, 0.4536 and 0.3783, respectively. Thus, while the FIEGARCH model implies that volatility at the five-minute unit is essentially uncorrelated with the volatility 19 days earlier, the SP-FIEGARCH model finds a sizeable correlation. For the EGARCH model, we obtain an autocorrelation of 0.0418 for the first day, 0.0000 for the fifth day and 0.0000 for the tenth day. In contrast, the SP-EGARCH specification displays more persistence, yielding 0.3271 for the first day, 0.0031 for the fifth and 0.0001 for the tenth. When comparing these autocorrelations with those of the FIEGARCH and SP-FIEGARCH specifications, it becomes however evident that the EGARCH and SP-EGARCH formulations cannot account for the long-range dependency of volatility we noticed in the autocorrelation functions in Figure 3.4.[12]

Notice also that the estimates for the fractional integration parameter d for the FIEGARCH and SP-FIEGARCH specifications are below 0.5, suggesting that the volatility process is covariance-stationary. The estimates for the distribution parameter η_1 and η_2 suggest that

[12] Unfortunately, the structures of the GARCH and FIGARCH models do not permit a 'natural' interpretation of the volatility autocorrelation functions. One may, however, attempt to quantify the persistence in GARCH or FIGARCH models with the accumulated impulse response functions. See Baillie, Bollerslev and Mikkelsen (1996).

the residual innovations $z_{n,t}$ have excess kurtosis: the estimates for η_1 are clearly not 'close to infinity' while the estiamtes for η_2 are below 2.

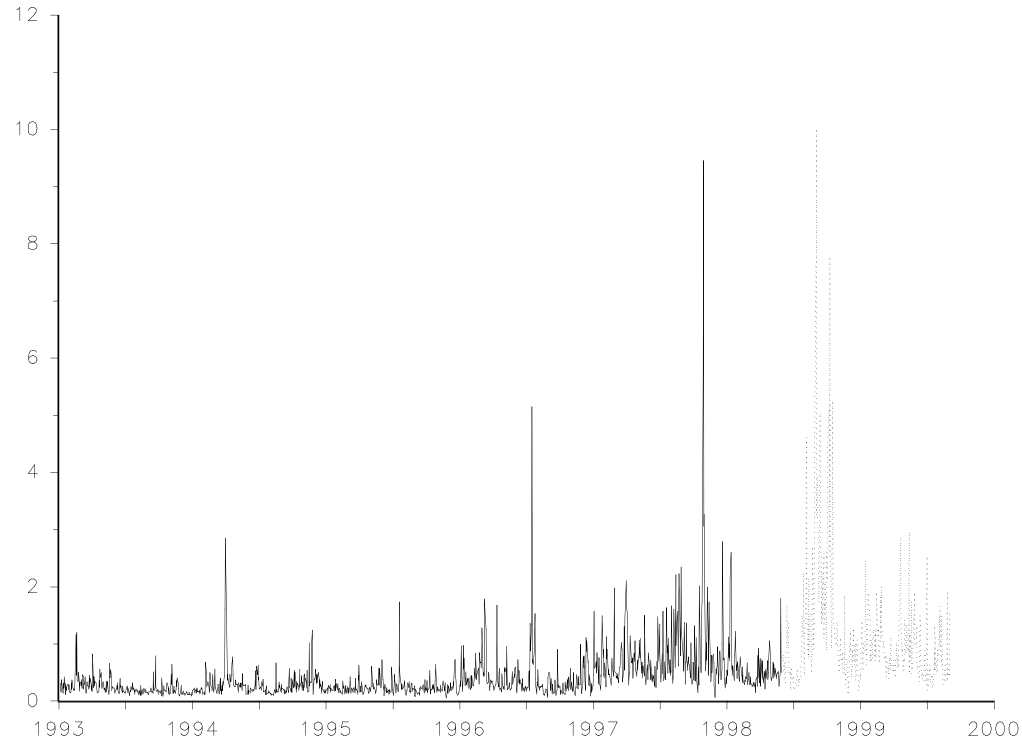
The importance of the long-range dependence also becomes evident when comparing the maximized log-likelihood statistics, $\mathfrak{L}^*$. The fractionally-integrated models do better. For instance, while the EGARCH yields 147294, the FIEGARCH model yields a considerably higher 147668. Similarly, the GARCH gives us 17604 while the FIGARCH model obtains the higher 147744. Across all models, we can see that the SP-FIGARCH model yields the highest statistic.

3.6 Properties of Daily Returns and Daily Realized Volatilities

The properties of the realized volatility measure underlying this study are documented in great detail in Chapter 2. Here we shall briefly summarize our findings as well as present summary statistics for our daily return series. This will help for the specification of the daily ARCH models in our next section and the realized volatility models discussed in Section 3.8. As we employ realized volatilities for forecast evaluation, the understanding the properties of this measure will also proof useful to interpret the performance of the various model we consider.

The time series of realized variances is displayed in Figure 3.5. The solid line portion of the figure spans from January 4, 1993 to May 29, 1998, the time period used for model estimation, while the broken line of the figure stretches from June 1, 1998 to August 31, 1999, the time reserved for *ex post* forecast evaluation. It becomes evident that the return volatility of the DJIA index changes considerably over time and that volatilities cluster, *i.e.* when volatility is high (low), subsequent volatilities tend to be high (low) for a prolonged period.

Figure 3.5: Daily Realized Variances



The solid line displays the time series of realized variances over the period extending from January 4, 1993 to May 29, 1998 while the dotted line gives the realized variances over the period from June 1, 1998 to August 31, 1999.

Using the first sub-sample, Table 3.2 presents a menu of summary statistics describing the properties of daily returns, r_t, and standardized returns, $(r_t - \bar{r})/s_t$ as well the characteristics of daily realized variances, s_t^2, standard deviations, s_t and logarithmic variances, $\ln(s_t^2)$. The numbers in parentheses below the Ljung-Box portmanteau statistic, Q_{20}, report the probability that the 20 autocorrelations are not significant. All other numbers in parenthesis report standard error estimates.

Paying attention to the three volatility measures first, notice that the skewness and kurtosis estimates of logarithmic variances are close to the values hypothesized under normality (zero and three, respectively). As logarithmic variances are distributed approximately normal, it follows that variances and standard deviations are distributed (approximately) lognormal. Accordingly, we find that the distributions of variances and standard deviations are skewed

Table 3.2: Properties of Daily Returns and Daily Realized Volatilities

	mean	*var*	*skew*	*kurt*	Q_{20}	d	$r_{t-1}I$
r_t	0.046	0.514	0.025	14.241	29.980	-0.058	
	(0.018)	(0.078)	(0.066)	(0.133)	(0.070)	(0.041)	
$(r_t - \bar{r})/s_t$	0.045	1.078	0.026	2.751	20.001	0.001	
	(0.029)	(0.037)	(0.066)	(0.133)	(0.458)	(0.041)	
s_t^2	0.417	0.223	8.188	122.588	2117.997	0.379	-0.956
	(0.034)	(0.099)	(0.066)	(0.133)	(0.000)	(0.043)	(0.258)
s_t	0.599	0.058	2.574	16.784	4669.983	0.395	-0.447
	(0.020)	(0.012)	(0.066)	(0.133)	(0.000)	(0.041)	(0.052)
$\ln(s_t^2)$	-1.151	0.456	0.748	3.784	6082.372	0.396	-1.093
	(0.060)	(0.045)	(0.066)	(0.133)	(0.000)	(0.039)	(0.071)

The table reports summary statistics for daily returns, r_t, standardized returns, $(r_t - \bar{r})/s_t$, as well as daily realized variances, s_t^2, standard deviations, s_t, and logarithmic variances, $\ln(s_t^2)$. The first four columns report the mean, variance ('var'), skewness ('skew') and kurtosis ('kurt') estimates. Column five gives the Ljung-Box portmanteau statistic for up to 20th-order serial correlation, Q_{20}. Geweke and Porter-Hudak (1993) estimates for the fractional integration parameter d are given in column 6. The last column reports for a series v_t the estimate of ω_3 in the regression $v_t = \omega_1 + \omega_2 \, r_{t-1} + \omega_3 \, r_{t-1} \, \mathrm{I} + \epsilon_t$, where the indicator function I takes value one when $r_{t-1} < 0$. The numbers in parentheses below the Ljung-Box portmanteau statistic report the probability that the 20 autocorrelations are not significant. All other numbers in parenthesis report standard error estimates. For columns one, two and seven we report the autocorrelation consistent Newly-West (1987) standard errors involving 20 lags. For the skewness and kurtosis estimates we calculate the standard errors as $\sqrt{6/T}$ and $\sqrt{24/T}$ respectively. The standard errors for the Geweke and Porter-Hudak estimates are obtained using the usual OLS regression formula. The sample period is January 4, 1993 to May 29, 1998.

right and leptokurtic.

All three volatility measures display a high degree of temporal dependency. The Ljung-Box portmanteau statistics, Q_{20}, reject clearly the hypotheses of no joint significance of the first 20 autocorrelations. Next to the portmanteau statistics, we report the Geweke and Porter-Hudauk (1993) log-periodogram regression estimates for the fractional integration parameter d to assess the long-run behavior of volatility. The estimated values are several standard errors away from both zero and 0.5 and this suggests that the realized volatility process has long-memory and is covariance-stationary. This is consistent with what we found in the context of the intradaily FIEGARCH specifications.

Recall that the EGARCH and FIEGARCH model estimates suggested the presence of the leverage effect, *i.e.* that negative lagged returns lead to greater future volatility than lagged positive returns of the same magnitude. To determine whether the asymmetry of volatility in past returns remains important when the model-free daily realized volatility is considered, the last column in Table 3.2 reports the estimates of ω_3 in the regression $v_t = \omega_1 + \omega_2\, r_{t-1} + \omega_3\, r_{t-1}\, \mathrm{I} + \epsilon_t$, where v_t denotes one of our three volatility series and the indicator function I takes value one when $r_{t-1} < 0$. As the estimates for ω_3 are significantly smaller than zero at the conventional levels, the leverage effect is clearly an important characteristic of the realized volatility process.

Turning our attention to the results for daily returns, r_t, we can see that the estimated mean return is two standard errors away from zero, suggesting that one can reject the non-zero mean hypotheses at the 5% level. This is of course to be expected from our analysis of intraday returns where we found that the first and particularly the last five-minute means are sizeable.

Note that the series $(r_t - \bar{r})/s_t$ is the sample analog to the standard ARCH and SV model representation of returns: $r_t - \mu = z_t\, \sigma_t$, where μ is the mean and σ_t the (time t) standard deviation of r_t. In such a representation, it is assumed that z_t is mean zero, unit variance and *i.i.d.* noise. Note that $(r_t - \bar{r})/s_t$ has indeed the desired property that it has a mean close to zero and a variance of about unity.

Upon comparing the kurtosis estimates for $(r_t - \bar{r})/s_t$ and r_t we can see that changing volatilities can fully account for the excess kurtosis in the distribution of returns. While the kurtosis estimate for returns is 14.241, it is only 2.751 for standardized returns, close to the value hypothesized under normality. This then suggests that the distribution of returns is $(r_t - \mu) \sim N(0, \sigma_t)$. In applied work it is often found that when one uses the conditional time $t-1$ standard deviation estimates from ARCH and Stochastic Volatility models, that the innovations z_t have excess kurtosis (see our previous section for instance). This perhaps suggests model misspecification or that the conditional time $t-1$ standard deviation process is different from the unconditional one (or both, of course).

Note finally that the hypotheses that the first 20 autocorrelations of returns and standardized returns, $(r_t - \bar{r})/s_t$, are jointly insignificant cannot be rejected at the 5% level. In Section 3.4, the significant sample autocorrelations in the five-minute returns series were likely driven by market micro-structure effects. Thus the absence of serial correlation in daily returns is to be expected.

3.7 Daily ARCH Models

The vast majority of all empirical studies related to the ARCH approach to modeling volatility employ daily data. In our forecast comparison we shall therefore consider these models at the daily return frequency as well. As in Section 3.5, we consider the FIGARCH and FIEGARCH models to accommodate long-run volatility dependencies. Since we discussed the basic features of these models at some length previously, we shall in this section only highlight aspects that arise when daily data are used for estimation.

Recall that we in our previous section concluded that the daily DJIA returns have a positive mean, but are uncorrelated over the sample extending from January 4, 1993 to May 29. We therefore consider the following return representation:

$$
\begin{aligned}
r_t &= \mu + \varepsilon_t \\
\varepsilon_t &= \sigma_t \, z_t
\end{aligned}
\tag{3.7.1}
$$

where $E[z_t] = 0$ and $E[z_t^2] = 1$.

When using the FIGARCH (p,d,q) model, we consider the conditional variance process:

$$
\sigma_t^2 = \frac{\omega + \Big[\big(1 - \beta(L_p)\big) - \big(1 - \alpha(L_q) - \beta(L_p)\big) \big(1 - L\big)^d \Big] \varepsilon_t^2}{\big(1 - \beta(L_p)\big)} \tag{3.7.2}
$$

As in Section 3.5, we condition the innovations $z_t = \epsilon_t/\sigma_t$ on the density of the Student's t distribution, *i.e.* $z_t \sim T(0, 1, \eta_1)$.

The conditional variance process in the FIEGARCH (p,d,q) model is defined by:

$$\ln(\sigma_t^2) = \omega + \frac{\alpha(L_q)\,\big(\,\gamma\, z_t + |z_t| - \mathrm{E}\big[\,|z_t|\,\big]\,\big)}{\big(\,1 - L\,\big)^d\,\big(\,1 - \beta(L_p)\,\big)} \tag{3.7.3}$$

where we shall assume that $z_t \sim GED(0, 1, \eta_2)$, *i.e.* we condition the residual innovations on the density of the generalized error distribution.

We noted in Section 3.5 that the constant estimate in the conditional variance process of the interdaily FIGARCH model implied an unconditional variance that did not match what we observed in our data. Consequently, we obtained long-horizon variance forecasts that were substantially biased. Instead of estimating the constant, we solved this problem by setting it so that the implied unconditional variance was equal to what was found in the data. In our application of the FIGARCH model to daily data, as we shall see in Section 3.9, this quandary does not occur and therefore no adjustment is made. Unlike in Section 3.5, estimates for ω in (3.7.2) are therefore reported. Admittedly, we do not have a convincing explanation for this discrepancy between the intradaily and daily model.

As we allow for a constant term in the return process given by (3.8.1), we calculate the variance forecast for day t as: $\hat{\mu}^2 + \hat{\sigma}_t^2$, where $\hat{\mu}$ is the estimated mean return and $\hat{\sigma}_t^2$ is the unadjusted variance forecast. This adjustment is made so that we do not penalize a model unfairly, if it should indeed be true that the realized volatility is overestimated.

Maximum likelihood estimates for the GARCH (1,1), FIGARCH (1,d,1), EGARCH (1,2) and FIEGARCH (0,1,1) models are reported in Table 3.3. Standard errors are given in parentheses and $\mathfrak{L}^*$ reports the maximized log-likelihood. With the exception of the fractional integration parameter d in the FIGARCH model, all estimates are significant at the 5% level on the basis of either Wald or log-likelihood ratio tests.

Consistent with prior empirical work using ARCH models, the innovations z_t are heavy-tailed, the implied volatility processes are highly persistent and, when we allow for asym-

Table 3.3: Daily ARCH Model Estimates

	$\hat{\mu}$	$\hat{\omega}$	$\hat{\beta}_1$	$\hat{d}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}$	$\hat{\eta}_{1,2}$	$\mathfrak{L}^*$
GARCH	0.063	0.008	0.930		0.054			6.250_1	-1340.0
	(0.016)	(0.005)	(0.024)		(0.018)			(1.043)	
FIGARCH	0.063	0.021	0.652	0.375	-0.285			6.536_1	-1338.6
	(0.016)	(0.012)	(0.105)	(0.108)	(0.108)			(1.150)	
EGARCH	0.050	-0.884	0.972		0.231	-0.117	-0.596	1.425_2	-1329.1
	(0.015)	(0.125)	(0.014)		(0.043)	(0.047)	(0.158)	(0.075)	
FIEGARCH	0.065	-1.245		0.585	0.227		-0.668	1.418_2	-1326.7
	(0.014)	(0.274)		(0.056)	(0.041)		(0.039)	(0.072)	

The table reports conditional maximum likelihood estimates of the models defined by equations 3.7.1 to 3.7.3. Estimates for η carry suffix 1 when the innovations z_t are conditioned on Student's t density and suffix 2 when the generalized error density is used instead. Standard errors, based on the second derivatives of the log-likelihood function, are reported in parentheses. $\mathfrak{L}^*$ reports the maximized log-likelihood. The $(1-L)^d$ polynomial, when used, is truncated at lag 1000. The sample period is January 4, 1993 to May 29, 1998.

metry in returns, the news parameters suggest the presence of the leverage effect. Our FIEGARCH estimate of $\hat{d} = 0.585$ is in line with the one reported by Bollerslev *et al.* (1996), who found $\hat{d} = 0.633$ for the S&P 500 composite stock index. However, this estimate is not consistent with what we found in the context of the two intradaily FIEGARCH models in Section 3.5. In that application, we obtained $\hat{d} = 0.345$ when we did not account for seasonality while we found 0.499 when intraday patterns were taken into consideration. In either case, the estimates of the higher frequency models suggest that the logarithmic variance process is stationary while the estimate for the daily model does not.

Based on maximized log-likelihood, the FIEGARCH model is the most promising ARCH specification for characterizing changing variances. Thus, this model should perform well in the forecasting exercises.

3.8 Daily Realized Volatility Models

Models of the ARCH class have been instrumental in capturing the temporal dependencies of volatility. In Chapter 2 we suggested another approach outside this family. Our method consists of applying conventional time series models directly to the realized volatility. Thus, rather than modeling the conditional variance of returns, we treat the realized volatility as observed and model its conditional mean. The most obvious advantage of this approach is its simplicity. On the basis of *ex ante* one-day-ahead prediction criteria we also found that it leads to important improvements compared to the traditional ARCH method. One focus of this study is to determine whether this superiority continues to hold when forecasting volatility for horizons beyond one-day and out-of-sample.

To capture the time series characteristics of the daily realized volatility, we find that the following specification, which we denote as RV-ARFIMAX1 (p,d,q), yields good results:

$$\big(1 - L\big)^d \, \big(1 - \beta(L_p)\big) \ln(s_t^2) = \omega_0 + \omega_1 \, r_{t-1} \, \mathrm{I}^- + \omega_2 \, r_{t-1} \, \mathrm{I}^+ + \big(1 + \alpha(L_q)\big) \, \varepsilon_t \qquad (3.8.1)$$

where I^- (I^+) is one when $r_{t-1} < 0$ ($r_{t-1} \geq 0$) and zero otherwise. It is assumed that $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2)$. Next to the constant and the standard two ARMA (p,q) polynomials, the above specification contains the fractional differencing operator to capture the long-run dependencies of volatility and lagged negative and positive returns to account for the leverage effect.

As the realized volatility is assumed to be stochastic, the above formulation is in the spirit of Stochastic Volatility (SV) models. The obvious difference though is that SV models are applied to (transforms of) returns and consequently the volatility process is treated as unobserved. If $\omega_1 = \omega_2 = 0$, the functional form we propose is however identical to the one of the SV-ARFIMA model proposed by Breidt, Crato and de Lima (1998). However, instead of the realized logarithmic volatility, they model the logarithm of daily squared returns.

Aside from stochastic character, the functional form of our specification also closely resembles the FIEGARCH model of Bollerslev and Mikkelsen (1996) described earlier. In that specification the logarithmic variance process is allowed to be fractionally integrated as well. In the FIEGARCH model the leverage effect is accounted for in that the logarithm of variances is a function of all past positive and negative *standardized* returns, $(r_t - \bar{r})/\sigma_t$. In our formulation, it is the history of *raw* returns that matters.

As we shall see in our forecasting exercises, the specification of the news-impact using raw returns leads to important improvements at the one day horizon. However, such specification has the considerable disadvantage that for multi-step-ahead variance forecasts, one requires the solution to:

$$E\big[\, e^{a\, r_{t+m|t}\, I^-} \,|\, t \,\big] = \int_{-\infty}^{0} e^{a\, r_{t+m}}\, f(r_{t+m} \,|\, t)$$

where a is a number determined by the moving average representation of the process and $f(r_{t+m} \,|\, t)$ is the conditional time t density of future returns. Although we know from our previous section that $r_t \sim N(\bar{r}, s_t^2)$, this does not imply that r_{t+s}, conditional upon time t, is distributed normal as well. Assuming that s_t^2 is mean reverting, at least for $m = \infty$ we have $E[\, s_{t+m}^2 \,|\, t\,] = E[\, s_t^2\,]$, *i.e.* the conditional and unconditional expectations are eventually equal. In that case we know that returns will display excess kurtosis and thus normality will not hold.

In the EGARCH and FIEGARCH model, this problem is circumvented by considering standardized returns instead of raw returns. We shall consequently also consider a modified version of our model, denoted RV-ARFIMAX2 (p,d,q), given by:

$$\big(\,1 - L\,\big)^d \,\big(\,1 - \beta(L_p)\,\big) \ln(s_t^2) = w_0 + w_1\, z_{t-1}\, \mathrm{I}^- + w_2\, z_{t-1}\, \mathrm{I}^+ + \big(\,1 + \alpha(L_q)\,\big)\, \varepsilon_t \qquad (3.8.2)$$

where $z_{t-1} = (r_{t-1} - \bar{r})/s_{t-1}$, I^- (I^+) is one when $z_{t-1} < 0$ ($z_{t-1} \geq 0$) and zero otherwise and we shall assume that $z_{t-1} \sim i.i.d.\, N(0,1)$.[13] As it now hold that $f(z_{t+m} \,|\, t\,)$ is normal

[13]The *i.i.d.* assumption is perhaps questionable. Although we do not find any evidence for serial correlation in z_t (see Section 3.7) and z_t^2, the BDS test (Brock, Dechert, LeBaron and Scheinkman 1987) for independence of z_t yields test statistics of $W_2 = -2.721$, $W_3 = -2.839$ and $W_4 = -2.089$. As these statistics are distributed standard normal, we therefore have to reject independence at the 5% level.

for all m, we obtain:

$$E\left[\, e^{a\, z_{t+m}\, I^-} \,|\, t \,\right] = \int_{-\infty}^{0} e^{a\, z_{t+m}} \, \frac{1}{\sqrt{2\,\pi}} \, e^{-0.5\, z_{t+m}} = e^{0.5\, a^2} \left(-\frac{a}{|a|} \, \mathrm{Erf}\,(\sqrt{0.5\, a^2}\,) + 0.5 \right) \quad (3.8.3)$$

where Erf is the error function.

For both the RV-ARFIMAX1 and RV-ARFIMAX2 models, we will require for variance forecasts an adjustment due to the stochastic component ϵ_t. From our assumption that $\epsilon_t \sim N(0, \sigma^2)$, it follows:

$$E\left[\, e^{a\, \epsilon_{t+m}} \,|\, t \,\right] = \int_{-\infty}^{\infty} e^{a\, \epsilon_{t+m}} \, \frac{1}{\sqrt{2\,\pi\, \sigma^2}} \, e^{-0.5\, \epsilon_{t+m}^2 / \sigma^2} = e^{0.5\, a^2\, \sigma^2} \quad (3.8.4)$$

where a is, again, given by the moving average representation.

Note that the adjustment terms given by (3.8.3) and (3.8.4) will scale the variance forecasts and the exact amount of scaling will critically depend upon the assumptions imposed to derive these two expressions. In practice, as in the EGARCH and FIEGARCH models, one may alternatively calculate the expectations using the data employed for estimation. For consistency, this is the approach we shall take.

For our forecasting exercises we shall consider the following five realized volatility (RV) models with corresponding labels: RV-AR an ARFIMA (7,0,0), RV-FI an ARFIMA (0,d,0), RV-ARX2 an ARFIMAX2 (7,0,0), RV-FIX2 an ARFIMAX2 (0,d,0) and RV-FIX1 an AFRIMAX1 (0,d,0). Parameter estimates of these specifications are given in Table 3.4. Standard errors are reported in parentheses under the coefficient estimates. All of the estimates are statistically significant at the 5% level on the basis of either Wald or likelihood ratio tests. The table also reports the maximized log-likelihood $\mathfrak{L}^*$ and the Schwartz Bayesian Information Criterion (SBC) which was used to determine the 'optimal' length of autoregressive and moving average polynomials. By the SBC we found that moving average terms were not warranted in any of the specifications. Furthermore, once we allowed for fractional integration of the volatility process, none of the autoregressive terms were found important either. When we allowed for the news effect, we found that only negative returns or negative standardized returns matter.

Table 3.4: Daily Realized Volatility Model Estimates

	RV-AR	*RV-FI*	*RV-ARX2*	*RV-FIX2*	*RV-FIX1*
$\hat{\omega}_0$	-0.170	-0.043	-0.253	-0.125	-0.153
	(0.030)	(0.015)	(0.031)	(0.018)	(0.020)
$\hat{\omega}_1$			-0.196	-0.198	-0.316
			(0.023)	(0.023)	(0.030)
$\hat{\beta}_1$	0.419		0.368		
	(0.027)		(0.027)		
$\hat{\beta}_2$	0.106		0.128		
	(0.029)		(0.029)		
$\hat{\beta}_1$	0.093		0.095		
	(0.029)		(0.029)		
$\hat{\beta}_1$	0.063		0.071		
	(0.029)		(0.029)		
$\hat{\beta}_1$	0.068		0.077		
	(0.029)		(0.029)		
$\hat{\beta}_1$	0.102		0.100		
	(0.027)		(0.026)		
$\hat{d}$		0.392		0.364	0.324
		(0.020)		(0.018)	(0.017)
$\hat{\sigma}^2$	0.227	0.221	0.215	0.210	0.205
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
$\mathfrak{L}^*$	-923.4	-908.0	-889.5	-871.3	-856.1
SBC	-952.3	-918.8	-922.0	-885.8	-870.6

The table reports conditional maximum likelihood estimates of the models defined by equations 3.8.1 and 3.8.2. Standard errors, based on the second derivatives of the log-likelihood function, are reported in parentheses. $\mathfrak{L}^*$ reports the maximized log-likelihood and *SBC* reports the Schwarz Bayesian Information Criterion, $SBC = \mathfrak{L}^* - 0.5\,k\,\ln(T)$, where k is the number of estimated coeeficients and $T = 1366$. The $(1-L)^d$ polynomial, when used, is truncated at lag 1000. The sample period is January 4, 1993 to May 29, 1998.

The estimates for the fractional integration parameter d are several standard errors away from both one-half and zero – indicating that the logarithmic variance process is covariance-stationary and highly persistent. It is noticeable that the estimates for d are lower in the models that account for the leverage effect. Specifically, we obtain $\hat{d} = 0.392$ for the RV-FI model (which does not account for the leverage effect), while the RV-FIX2 model yields $\hat{d} = 0.364$ and the RV-FIX1 model gives $\hat{d} = 0.324$. One might be tempted to conclude that the two models with lower estimates for d display less persistence. This however may not be necessarily correct as in these models the current volatility depends not only on the entire history of past volatilities but also on past negative returns. As such, a negative return innovation leads to persistent changes in future volatility as does an innovation in the noise component ϵ_t.

When looking at the estimates for the autoregressive coefficients in the RV-AR and RV-ARX2 mode, it becomes evident that these decay until lag four, but then increase again. This increase is not consistent with what one would expect when the process is fractionally integrated with $d > 0$, as the coefficients in the autoregressive representation of such models strictly decay. Our forecasting exercises should help us to determine whether the RV-AR and RV-ARX models find important short-run dynamics or whether the coefficients are just estimated with error (notice that one can easily find patterns of strict decay when looking at the two standard error bands of the autoregressive coefficients).

All coefficients estimated for ω_1 are negative and highly significant (compare, for instance, the maximized log-likelihood for the RV-FI model and the RV-FIX1 model). The negative sign suggests that large negative returns increase future volatility and this is consistent with the leverage effect.

By the SBC criterion, the models are ranked in decreasing order as follows: RV-FIX2, RV-FIX1, RV-FI, RV-AR, RV-ARX1. From this one may conclude that (a) long-memory models perform better than short-memory models, (b) models that account for the leverage effect outperform those that ignore it, and (c) that the leverage effect is best captured using negative returns rather than negative standardized returns. In the next section, we will determine whether these conclusions remain valid in the setting of forecasting.

3.9 Multi-Step Volatility Forecasting

In this section we shall evaluate how well the various models forecast the variance of the DJIA index *ex ante* and *ex post* over various time horizons, up to 40-days-ahead. In our first subsection, we shall detail the tools we employ to evaluate the variance forecasts. Next we discuss the *ex ante* forecast results, using data from January 4, 1993 to May 29, 1998. Finally, in our last subsection we turn to the *ex post* forecasting exercises employing the sample we have reserved thus far, the data extending from June 1, 1998 through August 31, 1999.

3.9.1 Volatility Forecast Evaluation

We shall denote the conditional time $t-1$, M days variance forecasts over $t, \ldots, t-1+M$ by $\hat{\sigma}^2_{t,M}$. It is obtained from either M accumulated daily or $M\,N$ accumulated intradaily volatility forecasts:

$$\hat{\sigma}^2_{t,M} = \frac{1}{M}\sum_{m=1}^{M}\hat{\sigma}^2_{t-1+m|t-1} = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}\hat{\sigma}^2_{n,t-1+m|N(t-1)}$$

where $\hat{\sigma}^2_{t-1+m|t-1}$ denotes the conditional time $t-1$ variance forecast for day $t-1+m$ and $\hat{\sigma}^2_{n,t-1+m|N(t-1)}$ denotes the conditional time N, $t-1$ (end of day $t-1$) variance forecast for the nth intraday time interval at day $t-1+m$. Upon dividing the horizon forecasts by M, we ensure that $\hat{\sigma}^2_{t,M}$ is measured at the daily unit.

In this setting of volatility forecasts, it has been traditional to compare $\hat{\sigma}^2_{t,1}$ to the daily squared return. As made clear in Section 3.2, we know this is a very noisy measure and consequently any variation found in the forecasts will likely not correspond well to the variation in squared returns. As in Andersen and Bollerslev (1998) and Ebens (1999), we shall therefore employ our realized volatility measure for forecast evaluation. Specifically, we shall use the accumulated M days realized variance $s^2_{t,M}$ defined by:

$$s^2_{t,M} = \frac{1}{M}\sum_{m=1}^{M} s^2_{t-1+m} = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N} r_{n,t-1+m}$$

A central property of optimal forecasts is that these are unbiased. We shall test whether this holds using the mean error criterion:

$$\mathrm{ME} = \frac{1}{T}\sum_{t=1}^{T}\left[s^2_{t,M} - \hat{\sigma}^2_{t,M}\right] \tag{3.9.1}$$

where the period used for forecast evaluation runs from $t = 1, \ldots, T$. We therefore evaluate the forecasts over days: $\{1, \ldots, M\}$, $\{2, \ldots, 1+\mathrm{M}\}, \ldots, \{t-M, \ldots, T\}$.

The ME, of course, depends only on the first moment structure of the joint distribution of the actual and forecasted series. Of considerable interest is the second moment (or some transform thereof) which reveals variations in forecasts correspond to those found in actual data.

Many criteria have been suggested that penalize for the mean and variance of the forecast errors. By far, the most commonly used criterion is the mean squared error (MSE) defined as: $T^{-1}\sum_{t=1}^{T}(\,s^2_{t,M} - \hat{\sigma}^2_{t,M}\,)^2$. Note however that the MSE depends on the fourth moment of estimated and forecasted returns, *i.e.* , MSE $= V[\,s^2_{t,M} - \hat{\sigma}^2_{t,M}\,] + (\,E[\,s^2_{t,M} - \hat{\sigma}^2_{t,M}\,]\,)^2$. Some of our models, however, imply that these moments do not exist (the GARCH models in Section 3.5, for instance). Even if they do exist, it is likely that they will be obtained with great error and this, in turn, would induce large error in the MSE as well. Hence, it would be difficult to clearly identify the better performing specifications.

An alternative measure that does not depend on the fourth moment of returns is the logarithmic loss function (see Pagan and Schwert 1990, Diebold and Lopez 1996): MSLE $= T^{-1}\sum_{t=1}^{T}(\,\ln(s^2_{t,M}) - \ln(\hat{\sigma}^2_{t,M})\,)^2$. This measure is in principle attractive since we from our study of the realized volatility know that logarithmic variances are nearly normally distributed. For a well-specified model one would therefore expect that the logarithmic forecast error is approximately normal as well and thus symmetric. In the case of symmetry, one would equally penalize negative and positive deviations from the mean error. However, for the MSLE it holds that MSLE $= V[\ln(s^2_{t,M}) - \ln(\hat{\sigma}^2_{t,M})] + (E[\ln(s^2_{t,M}) - \ln(\hat{\sigma}^2_{t,M}])^2$. As we can see, loss depends on the bias of logarithmic variance forecasts. In the setting of multi-period ARCH forecasts, it will generally not be true that the logarithmic transform of an unbiased variance forecast will yield an unbiased logarithmic variance forecast. As our realized volatility specifications are stochastic, this will not even be true for the one-step ahead forecast. Consequently, such a loss function would embed an automatic bias and thus an unwarranted penalty.

Considering the difficulties with the MSE and MSLE statistics, we select the mean absolute error (MAE) criterion as the loss measure. It is defined by:

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}[\,|\, s^2_{t,M} - \hat{\sigma}^2_{t,M}\,|\,]$$

We find it informative to compare the MAE of the various forecasting models to the one of a naive competitor, the MAE of the homoskedastic forecast:

$$\bar{s}^2 = \frac{1}{T}\sum_{t=1}^{T} s^2_t$$

where $T = 1366$ (the mean realized in-sample variance). Specifically, we shall report the relative MAE:

$$\text{RMAE}_v = \ln(\,\text{MAE}_{\bar{s}^2}\,) - \ln(\,\text{MAE}_v\,) \tag{3.9.2}$$

In the spirit of Theil's (1961) U statistic, the RMAE$_v$ thus gives for a model v the percentage improvement in 'forecastability' relative to the naive forecast. Since we calculate the percentage improvement using logarithms, one may also easily compare the percentage improvement of model v_1 over v_2 using RMAE_{v_1}- RMAE_{v_2}.

3.9.2 Ex Ante Volatility Horizon Forecast

It is instructive to first consider our in-sample forecasting results as these are based on a larger number of observations. Furthermore, if a model should perform poorly within sample, then there is little hope for a much better performance out of sample (otherwise, the better performance it is likely to be explained by problems related to forecast evaluation).

The results are presented in Table 3.5 which is organized as follows. There are four sets of models: intradaily ARCH specifications which do not account for seasonality (prefixed I), intradaily ARCH models which incorporate seasonality (prefixed SP), daily ARCH models (prefixed D) and, finally, the daily realized volatility specifications (prefixed RV). For convenience, we refer to these four groups of models as the I group, D group, SP group and RV group, respectively. For each model, we consider nine different horizons: one-day-ahead to five-day-ahead, then ten-day, twenty-day, thirty-day and finally the forty-day-ahead forecasting horizon. The exception is the RV-FIX model, for which we can only obtain one-day-ahead forecasts (see Section 3.8).

For each model and horizon, we present the mean forecast error, ME, and the relative mean absolute forecast error, RMAE. The numbers in parentheses below the ME statistic are the autocorrelation consistent Newey-West (1987) standard errors involving 20 lags. The standard errors carry an asterisk (in bold) if we reject the hypothesis that the ME is significantly different from zero at the 5% level. For the SP-GARCH model we can see that forecasts are biased for the one-day horizon. For the D-EGARCH specification, forecasts are biased for one-day to five-day horizons, while they are always biased for the D-FIEGARCH model. Note, however, that these biases are, on the whole, quite small (the mean daily realized variance is 0.417).

Table 3.5: *Ex Ante* Volatility Horizon Forecasts

	1-Day		*2-Day*		*3-Day*		*4-Day*		*5-Day*		*10-Day*		*20-Day*		*30-Day*		*40-Day*	
	ME	*RMAE*																
I-GARCH	-0.036	15.7	-0.020	12.7	-0.013	9.9	-0.010	8.1	-0.008	6.7	-0.004	3.7	-0.002	1.9	-0.001	1.2	-0.001	0.9
	(0.022)		(0.027)		(0.029)		(0.030)		(0.031)		(0.032)		(0.032)		(0.031)		(0.030)	
I-FIGARCH	-0.017	35.5	-0.010	**40.9**[3]	-0.007	**42.5**[3]	-0.006	**44.4**[3]	-0.005	**45.8**[3]	-0.003	**47.4**[3]	-0.001	**43.0**[3]	0.001	**40.2**[3]	0.001	**38.3**[3]
	(0.010)		(0.011)		(0.013)		(0.014)		(0.015)		(0.018)		(0.020)		(0.020)		(0.020)	
I-EGARCH	-0.006	13.8	0.003	9.8	0.006	7.6	0.008	6.3	0.008	5.5	0.010	3.8	0.011	3.0	0.012	2.6	0.012	2.4
	(0.028)		(0.031)		(0.032)		(0.032)		(0.033)		(0.033)		(0.032)		(0.031)		(0.030)	
I-FIEGARCH	0.009	**36.4**[4]	0.017	40.5	0.020	42.2	0.022	43.8	0.022	44.3	0.024	46.0	0.026	42.1	0.027	38.4	0.027	34.5
	(0.016)		(0.018)		(0.019)		(0.020)		(0.020)		(0.022)		(0.023)		(0.022)		(0.022)	
SP-GARCH	-0.032	24.1	-0.024	26.0	-0.018	24.4	-0.015	22.1	-0.013	20.4	-0.007	12.9	-0.003	6.7	-0.002	4.5	-0.002	3.4
	(0.012)*		(0.016)		(0.020)		(0.023)		(0.024)		(0.029)		(0.030)		(0.030)		(0.029)	
SP-FIGARCH	-0.016	35.6	-0.010	40.7	-0.007	42.2	-0.005	44.0	-0.004	45.2	-0.002	46.6	0.000	41.6	0.001	38.5	0.001	36.4
	(0.010)		(0.012)		(0.014)		(0.015)		(0.016)		(0.019)		(0.021)		(0.021)		(0.021)	
SP-EGARCH	0.015	32.9	0.020	31.5	0.020	27.8	0.019	24.0	0.018	21.0	0.015	12.3	0.013	7.1	0.013	5.3	0.012	4.4
	(0.017)		(0.022)		(0.025)		(0.026)		(0.028)		(0.030)		(0.031)		(0.030)		(0.030)	
SP-FIEGARCH	0.017	**40.9**[2]	0.017	**44.4**[2]	0.018	**47.0**[2]	0.018	**48.5**[2]	0.018	**50.2**[2]	0.019	**53.3**[2]	0.021	**51.2**[2]	0.023	**49.6**[2]	0.024	**49.2**[1]
	(0.012)		(0.014)		(0.015)		(0.015)		(0.016)		(0.018)		(0.019)		(0.019)		(0.019)	
D-GARCH	-0.003	16.9	-0.003	17.7	-0.003	17.0	-0.003	16.1	-0.003	16.4	-0.004	18.7	-0.005	18.4	-0.005	22.7	-0.006	25.9
	(0.027)		(0.027)		(0.028)		(0.028)		(0.029)		(0.030)		(0.031)		(0.031)		(0.029)	
D-FIGARCH	-0.005	20.4	-0.005	20.9	-0.005	20.3	-0.004	19.9	-0.004	20.4	-0.004	23.1	-0.004	23.5	-0.003	24.6	-0.003	25.0
	(0.025)		(0.025)		(0.026)		(0.026)		(0.026)		(0.027)		(0.029)		(0.028)		(0.027)	
D-EGARCH	0.039	32.9	0.039	34.0	0.039	34.2	0.039	33.9	0.038	33.8	0.038	35.0	0.039	32.6	0.041	34.6	0.042	34.1
	(0.016)*		(0.017)*		(0.018)*		(0.019)*		(0.019)*		(0.021)		(0.023)		(0.023)		(0.023)	
D-FIEGARCH	0.049	**36.5**[3]	0.049	**38.1**[4]	0.049	**39.0**[4]	0.049	**39.6**[4]	0.049	**39.5**[4]	0.049	**42.0**[4]	0.049	**39.3**[4]	0.050	**38.2**[4]	0.050	**37.2**[4]
	(0.014)*		(0.016)*		(0.016)		(0.017)*		(0.018)*		(0.020)*		(0.021)		(0.021)*		(0.020)*	
RV-AR	0.009	37.7	0.008	40.4	0.008	42.4	0.007	43.9	0.007	44.5	0.008	45.1	0.011	38.9	0.012	32.2	0.011	26.9
	(0.012)		(0.013)		(0.014)		(0.015)		(0.015)		(0.018)		(0.022)*		(0.023)		(0.024)	
RV-FI	0.011	39.0	0.011	43.2	0.011	46.0	0.011	47.9	0.011	49.8	0.011	53.7	0.013	**52.0**[1]	0.014	**49.8**[1]	0.015	**48.8**[2]
	(0.013)		(0.014)		(0.015)		(0.016)		(0.017)		(0.019)		(0.020)		(0.019)		(0.019)	
RV-ARX2	0.010	41.1	0.011	43.3	0.011	44.6	0.010	45.5	0.010	45.7	0.012	45.8	0.015	39.0	0.015	32.0	0.014	26.5
	(0.012)		(0.013)		(0.014)		(0.015)		(0.015)		(0.019)		(0.022)		(0.023)		(0.024)	
RV-FIX2	0.014	**42.2**[1]	0.015	**46.3**[1]	0.016	**48.5**[1]	0.016	**49.7**[1]	0.016	**50.7**[1]	0.016	**53.8**[1]	0.017	51.7	0.019	49.1	0.019	47.4
	(0.013)		(0.014)		(0.016)		(0.016)		(0.017)		(0.019)		(0.020)		(0.020)		(0.019)	
RV-FIX1	0.008	46.0																
	(0.008)																	

The table presents results on the in-sample forecasting performance of the models defined in Section 3.4, Section 3.6 and Section 3.7. The forecasts are over one-day to five-day horizons, then ten-day, twenty-day, thirty-day and finally the forty-day horizon. The forecast are evaluated using the mean forecast error criterion (*ME*) defined by equation 9.1 and the relative mean absolute error criterion (*RMAE*) defined by equation 3.9.2. The sample period for estimation and forecast evaluation is January 4, 1993 to May 29, 1998.

Recall that the RMAE gives the percentage improvement relative to the naive homoskedastic forecast. The reported numbers in Table 3.5 range from a low of 0.9% to a high of 53.8% improvement, suggesting that all models outperform the homoskedastic model, but also indicating that the performance among the models varies greatly.

For each group of models (excluding the anomalous RV-FIX), we select the best-performing model according to the RMAE criterion (in bold) and then rank these across groups by an integer ranging from 1 to 4 (1 being the best and 4 the worst of the best-performing models). For instance, for the one-day-ahead forecast, the D-FIEGARCH model is the best-performing model among the four daily ARCH specifications, but is only the third best in overall ranking across groups.

Comparing the two groups of intradaily ARCH specifications, we can see that accounting for seasonality improves always the forecasting performance for all but one model. The exception is the FIGARCH model, where the I-FIGARCH specification tends to perform better than the SP-FIGARCH model.

Within the I group, the I-FIGARCH model is the best-performing specifications, with the exception of the one-day-ahead horizon, where the I-FIEGARCH performs best. Within the SP group, the SP-FIEGARCH model performs always better than any other specification. Similarly, for the D group we find that the D-FIEGARCH model outperforms any other specification. Finally, for the RV group, the RV-FIX2 model is best-performing up to the ten-day horizon, while for longer horizons, the RV-FI performs best.

The overall ranking of groups according to the best-performing models is as follows: the RV group of models perform best, with the exception of the 40-day-horizon, where the SP class performs better. Notice also that the SP class always ranks second-best at all other horizons. Thus, overall, the RV and SP classes always take the two top slots, and thus are clearly better than the I or D groups. Furthermore, notice that the percentage difference in performance between the best-performing RV and SP models is minimal. At the opposite end, we can see that the D group contains the worst-performing specification for all horizons, with the exception of the one-day-horizon, when the performance of the I-group model falls below it.

Although not ranked according to the above scheme, the anomalous RV-FIX model performs best at the one-day-ahead horizon. This suggests that the leverage effect is best captured using lagged raw returns rather than lagged standardized returns. Overall, the importance of the leverage effect seems to diminish as the forecast horizon extends. While both the RV-FIX2 and RV-ARX2 models perform better when the horizon is short, the RV-FI and RV-AR models do better over long-horizons.

Note that the models which account for long-range volatility dependence (i.e. the fractionally integrated models) always outperform their short-memory counterparts. The improvements offered by the long-memory models become more substantial as the forecast horizon extends, however. For instance, in the SP group, the SP-FIEGARCH model improves upon the SP-EGARCH model by 44.8% at the 40-day horizon while the improvement is only 8% for the one day ahead forecasts.

3.9.3 Ex Post Volatility Horizon Forecast

The results for the *ex post* forecasts are presented in Table 3.6, organized in a manner identical to Table 3.5. As fewer observations are available in this forecasting exercise, the overall picture becomes a bit more blurred – particularly for the longer-range horizons. Notice that we now obtain for many models biased forecasts and that the degree of bias tends to become more severe as the horizon is extended. As a general tendency, however, it is evident that the long-memory models are less affected by bias. The reported percentage improvements of mean absolute error ranges from a low of -1.1% to a high of 68.5%. We find only three cases where the naive homoskedastic forecast does better.

Table 3.6: *Ex Post* Volatility Horizon Forecasts

	1-Day		*2-Day*		*3-Day*		*4-Day*		*5-Day*		*10-Day*		*20-Day*		*30-Day*		*40-Day*	
	ME	*RMAE*																
I-GARCH	0.422	31.6	0.564	20.1	0.621	14.3	0.652	11.1	0.671	9.3	0.714	4.7	0.746	2.4	0.769	1.6	0.793	1.2
	(0.108)*		(0.143)*		(0.157)*		(0.164)*		(0.169)*		(0.177)*		(0.179)*		(0.176)*		(0.170)*	
I-FIGARCH	0.067	47.6	0.105	**55.4**^{3}	0.130	**58.4**^{3}	0.148	**63.6**^{3}	0.164	**65.3**^{3}	0.219	**60.9**^{3}	0.287	**55.5**^{4}	0.336	**53.1**^{4}	0.381	**51.2**^{4}
	(0.054)		(0.069)		(0.080)		(0.088)		(0.095)		(0.119)		(0.147)		(0.163)*		(0.173)*	
I-EGARCH	0.613	15.5	0.679	7.4	0.704	4.6	0.717	3.1	0.725	2.3	0.748	0.3	0.770	-0.7	0.789	-1.0	0.811	-1.1
	(0.164)*		(0.175)*		(0.178)*		(0.180)*		(0.181)*		(0.183)*		(0.182)*		(0.178)*		(0.171)*	
I-FIEGARCH	0.236	**49.0**^{3}	0.284	53.0	0.313	54.4	0.333	54.8	0.350	54.4	0.406	49.1	0.472	44.5	0.520	41.2	0.562	37.7
	(0.105)*		(0.119)*		(0.128)*		(0.134)*		(0.139)*		(0.154)*		(0.167)*		(0.173)*		(0.173)*	
SP-GARCH	0.181	39.9	0.313	38.9	0.405	36.0	0.469	31.4	0.516	28.0	0.632	15.3	0.705	7.7	0.741	5.3	0.771	4.0
	(0.052)*		(0.083)*		(0.105)*		(0.121)*		(0.132)*		(0.158)*		(0.171)*		(0.171)*		(0.166)*	
SP-FIGARCH	0.080	48.0	0.124	55.7	0.150	60.8	0.170	65.4	0.187	65.8	0.244	59.9	0.313	53.8	0.363	51.3	0.408	48.9
	(0.057)		(0.073)		(0.084)		(0.093)		(0.099)		(0.123)*		(0.149)*		(0.164)*		(0.172)*	
SP-EGARCH	0.382	39.0	0.501	29.4	0.566	22.6	0.607	17.7	0.635	14.4	0.701	6.7	0.745	2.5	0.772	1.2	0.798	0.6
	(0.112)*		(0.138)*		(0.151)*		(0.159)*		(0.164)*		(0.175)*		(0.178)*		(0.175)*		(0.169)*	
SP-FIEGARCH	0.101	**51.4**^{2}	0.122	**58.0**^{1}	0.139	**63.1**^{1}	0.153	**67.4**^{1}	0.166	**68.5**^{1}	0.213	**62.3**^{1}	0.281	**57.3**^{2}	0.331	**55.6**^{3}	0.375	**54.7**^{3}
	(0.066)		(0.080)		(0.089)		(0.097)		(0.103)		(0.126)		(0.151)		(0.165)*		(0.173)*	
D-GARCH	0.254	33.2	0.257	38.4	0.260	40.9	0.264	43.9	0.268	44.8	0.285	43.3	0.317	47.2	0.349	47.8	0.385	38.8
	(0.076)*		(0.079)*		(0.081)*		(0.084)*		(0.086)*		(0.098)*		(0.119)*		(0.139)*		(0.155)*	
D-FIGARCH	0.175	37.4	0.177	43.5	0.179	46.9	0.181	49.8	0.183	50.2	0.192	48.4	0.208	**55.9**^{3}	0.224	**56.7**^{2}	0.245	59.2
	(0.077)*		(0.081)*		(0.084)*		(0.087)*		(0.090)*		(0.103)		(0.127)		(0.147)		(0.162)	
D-EGARCH	0.365	38.2	0.385	40.5	0.397	41.5	0.406	41.4	0.414	41.2	0.449	37.1	0.507	31.6	0.558	26.9	0.606	23.4
	(0.091)*		(0.098)*		(0.101)*		(0.104)*		(0.107)*		(0.117)*		(0.131)*		(0.140)*		(0.144)*	
D-FIEGARCH	0.194	**46.4**^{4}	0.200	**49.7**^{4}	0.204	**52.7**^{4}	0.206	**52.9**^{4}	0.208	**52.8**^{4}	0.217	**52.0**^{4}	0.231	52.4	0.247	55.6	0.264	**59.4**^{2}
	(0.099)		(0.107)		(0.112)		(0.116)		(0.119)		(0.131)		(0.146)		(0.156)		(0.162)	
RV-AR	0.184	47.3	0.211	53.1	0.232	58.2	0.251	62.0	0.270	62.0	0.356	54.4	0.482	40.6	0.564	31.1	0.625	24.7
	(0.070)*		(0.079)*		(0.086)*		(0.092)*		(0.097)*		(0.119)*		(0.144)*		(0.155)*		(0.158)*	
RV-FI	0.132	47.9	0.146	55.4	0.158	**60.0**^{2}	0.169	**64.6**^{2}	0.178	**65.7**^{2}	0.213	**61.9**^{2}	0.263	**59.8**^{1}	0.300	**59.4**^{1}	0.335	**59.7**^{1}
	(0.084)		(0.095)		(0.104)		(0.111)		(0.116)		(0.137)		(0.159)		(0.172)		(0.178)	
RV-ARX2	0.199	**52.6**^{1}	0.228	**55.9**^{2}	0.249	59.5	0.268	63.1	0.286	62.7	0.373	53.7	0.498	38.6	0.578	29.2	0.637	23.0
	(0.071)*		(0.081)*		(0.088)*		(0.093)*		(0.099)*		(0.121)*		(0.145)*		(0.155)*		(0.158)*	
RV-FIX2	0.146	49.6	0.166	55.2	0.180	59.6	0.191	63.7	0.201	64.6	0.237	61.0	0.288	59.0	0.325	58.7	0.360	59.3
	(0.089)		(0.100)		(0.109)		(0.116)		(0.121)		(0.141)		(0.162)		(0.173)		(0.178)*	
RV-FIX1	0.077	54.6																
	(0.053)																	

The table presents results on the out-of-sample forecasting performance of the models defined in Section 3.4, Section 3.6 and Section 3.7. The forecasts are over one-day to five-day horizons, then ten-day, twenty-day, thirty-day and finally the forty-day horizon. The forecast are evaluated using the mean forecast error criterion (ME) defined by equation 3.9.1 and the relative mean absolute error criterion ($RMAE$) defined by equation 3.9.2. The sample period for estimation is January 4, 1993 to May 29, 1998, the one for forecast evaluation is June 1, 1998 to August 31, 1999.

The conclusions that can be drawn from the out-of-sample exercises remain largely as before: the models that account for seasonality tend to improve upon the ones that do not, the SP and RV groups generally take the top two slots in the rankings; the D group tends to perform worst and the I-group is the next-to-worst; the fractionally integrated models tend to do better than the short-memory models; the exceptional RV-FI model is best at the one-day-horizon; accounting for the leverage effect improves short-horizon forecasts, but lowers the performance over longer time spans. The only notable difference from our previous results is that now the SP class tends to perform best while the RV runs a close second, rather than the other way around.

Group-by-group, the results are virtually identical. Within the I-group, the I-FIGARCH is best (again with the exception of the one-day-ahead forecasts). Within the SP group, the SP-FIEGARCH model is again always the best. Within the D-group, the D-FIEGARCH again does better everywhere, but now with the exception of the twenty-day and thirty-day horizons (where the D-FIGARCH does better). Finally, in the RV group, departing a bit from the previous results, the RV-ARX2 does better up to the two-day horizon, while, for longer horizons, RV-FI performs better again.

3.10 Conclusions

Using five-minute returns on the Dow Jones Industrials Average portfolio over the period from January 1993 to August 1999, this chapter investigated the short and long horizon in and out-of sample forecasting performance of ARCH models when these are applied to daily and intradaily returns as well as realized volatility specifications that model the time series of squared intraday returns. We proposed semi-parametric ARCH specifications that simultaneously model intraday seasonality and persistency of volatility.

Our results make a strong case for the temporal dependency in volatility – models that account for it tend to forecast much better than a naive homoskedastic specification. The degree of enhancement differs however greatly across the various specifications. Formulations that give proper consideration to long-range volatility dependencies tend to outperform their short-memory counterparts. Accounting for the leverage effect improves short-horizon forecasts, but tends to lower the quality of forecasts over longer time spans. The comparison of daily and interdaily ARCH models revealed that specifications of the latter type generally provide better forecasts once consideration to the seasonal dependency of volatility within the day is given. Overall, a long-memory daily realized volatility and long-memory interdaily semi-parametric specification gave the most accurate forecasts.

Chapter 4

Conclusions and Suggestions for Future Research

This dissertation concerned the measurement, modeling and forecasting of volatility. Our first chapter introduced the subject matter and outlined the difficulties when the existing approaches are employed. Specifically, volatility estimates given by statistical and economic methods are model driven and may therefore not be valid. Direct indicators of volatility, such as daily squared returns, are subject to substantial error.

In our second chapter, we proposed measuring daily volatility model-free by summing the squares of intraday returns. We first derived the theoretical properties of this realized volatility estimator and showed that measurement error – under quite general conditions – can be made arbitrary small when sufficiently many intraday return data are employed. Using the transaction price record of the Dow Jones Industrials Average portfolio we next documented the properties of our stock volatility sample. Our main results have been that variances are distributed lognormal, that the volatility process is covariance stationary and highly persistent and that volatility correlates more strongly with lagged negative than lagged positive returns.

These findings set the stage for the development of a time series model that captures the temporal dependency of our volatility variable. On the basis of *ex ante* one-day-ahead prediction criteria we found that our proposed specification yields unbiased and accurate volatility predictions and that these are better than the ones obtained by daily ARCH specifications, including those that closely match the properties of volatility we documented.

The central topic of our third chapter has been to identify volatility models that yield useful forecasts over long-term out-of-sample horizons. For this propose we have juxtaposed three classes of models: intradaily ARCH, daily ARCH and realized volatility models. We have found that there are clear advantages in using high-frequency data. Forecasts based on models which use intradaily data – intradaily ARCH and realized volatility models – tend to outperform daily ARCH specifications.

These results strengthened our findings in the second chapter on the realized volatility model. Specifically, we have shown that the superiority of the realized volatility specification over daily ARCH models continues to hold out-of-sample and over long horizons. However, the application of ARCH formulations to intradaily returns has shown that the realized volatility may be inefficient. Although we found that the direct application of these models is hazardous, specifications that model simultaneously intraday seasonality and long-term persistency of volatility performed effectively as well as the realized volatility specification.

Many promising research projects emerge from the approach we have taken to measure, model and forecast volatility. First, a key determinant of an option price is the asset return volatility until expiration of a contract. The forecasting models we developed outperformed any of the standard techniques and thus should help towards the more efficient pricing of these derivatives.

Second, our analysis has focused only on the second moments of daily returns and the relation of second moments to first moments. For modern financial theories, however, the entire distribution of returns is of interest. Measures of third and fourth moments are readily obtained using intradaily data – allowing one to study the relation of these measures over time and to each other.

Third, for portfolio selection problems considerable interest is on the joint distribution of returns, *e.g.* covariances and correlations. Using intradaily data, such measures may be computed at the daily frequency and thus analyzed, modeled and forecasted using either the time series techniques we employed or their multivariate extensions.

Fourth, the direct measurement of both daily variances and covariances allows the computation of daily realized CAPM betas. Their forecasted values may help to determine future risk factors and identify under priced assets.

Finally, the joint distribution of returns may be related to observed variables such as price volume, transaction volume, bid-ask spreads, news announcements and so on. Such relations may explain further characteristics of asset price dynamics.

Bibliography

Andersen, T. G. and T. Bollerslev. 1997. "Intraday Periodicity and Volatility Persistence." *Journal of Emperical Finance* 4:115–158.

———. 1998. "Answering the Skeptics: 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. 1999. "The Distribution of Exchange Rate Volatility." National Bureau of Economic Research. Working Paper 6961.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen. 1996. "Fractional Integrated Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 74:3–30.

Ball, C. A. 1988. "Estimation Bias Induced by Discrete Security Prices." *Journal of Finance* 43:841–865.

Barndorff-Nielsen, O. E. and N. Shephard. 1999. "Non-Gaussian OU based Models and some of their Uses in Financial Economics." Centre for Analytical Finance. University of Aarhus, Aarhus School of Business. Working Paper 37.

Bera, A. K. and M. L. Higgins. 1993. "ARCH Models: Properties, Estimation and Testing." *Journal of Economic Surveys* 3:15–102.

Beran, J. 1994. *Statistics for Long-Memory Processes.* Chapman and Hall.

Black, F. 1976. "Studies of Stock Market Volatility Changes." *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–327.

——— 1987. "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics* 69:542–547.

Bollerslev, T. and H. O. Mikkelsen. 1996. "Modeling and Pricing Long Memory in Stock Market Volatility." *Journal of Econometrics* 73:151–184.

Bollerslev, T., R. Y. Chou, and K. F. Kroner. 1992. "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics* 52:5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson. 1994. "ARCH Models." Chapter 49 of *Handbook of Econometrics*, edited by R. F. Engle and D. L. McFadden, Volume 4, 2959–3038. North-Holland.

Breidt, J., N. Crato, and P. J. F. de Lima. 1998. "The Detection and Estimation of Long-Memory in Stochastic Volatility." *Journal of Econometrics* 83:325–348.

Brock, W. A., W. D. Dechert, B. LeBaron, and J. A. Scheinkman. 1987. "A Test for Independence Based on the Correlation Dimension." Social Systems Research Institute. University of Wisconsin, Madison. Working Paper 9520.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton University Press.

Chung, C.-F. and R. T. Baillie. 1993. "Small Sample Bias in Conditional Sum-of-Squares Estimator of Fractionally Integrated ARMA Models." *Empirical Economics* 18:791–806.

Clark, P. K. 1973. "A Subordinated Stochastic Process Model with finite Variance for Speculative Prices." *Econometrica* 41:135–155.

Crato, N. and P. J. F. de Lima. 1994. "Long-Range Dependence in the Conditional Variance of Stock Returns." *Economics Letters* 45:281–285.

Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. N. Olsen, and O. V. Pictet. 1993. "A

Geographical Model for the Daily and Weekly Seasonal Volatility in the FX Market." *Journal of International Money and Finance* 12:413–438.

Diebold, F. X. and J. A. Lopez. 1996. "Forecast Evaluation and Combination." Chapter 5 of *Handbook of Statistics*, edited by G. S. Maddala and C. R. Rao, Volume 14, 241–268. North Holland.

Ding, Z., R. F. Engle, and C. W. J. Granger. 1993. "Long Memory Properties of Stock Market Returns and a new Model." *Journal of Empirical Finance* 1:83–106.

Drost, F. C. and T. E. Nijman. 1993. "Temporal Aggregation of GARCH Processes." *Econometrica* 61:909–927.

Drost, F. C. and B. J. M. Werker. 1996. "Closing the GARCH Gap: Continiouss Time GARCH modelling." *Journal of Econometrics* 74:31–57.

Drost, F. C., T. E. Nijman, and B. J. M. Werker. 1998. "Estimation and Testing in Models Containing Both Jumps and Conditional Heteroskedasticity." *Journal of Business & Economic Statistics* 16:237–243.

Ebens, H. 1999. "Realized Stock Volatility." Department of Economics. The Johns Hopkins University. Working Paper 420.

Engle, R. F. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance for U.K. Inflation." *Econometrica* 50:987–1008.

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

Engle, R. F., D. M. Lilien, and R. P. Robins. 1987. "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model." *Econometrica* 55:391–407.

Fama, E. 1965. "The Behavior of Stock Market Prices." *Journal of Business* 38:34–105.

Fuller, W. A. 1996. *Introduction to Statistical Time Series*. Second Edition. Wiley Series in Probability and Statistics. John Wiley and Sons.

Gallant, R. A. 1981. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form." *Journal of Econometrics* 15:211–245.

Geweke, J. and S. Porter-Hudak. 1983. "The Estimation and Application of Long Memory Time Series Models." *Journal of Time Series Analysis* 4:221–238.

Ghysels, E., A Harvey, and E. Renault. 1996. "Stochastic Volatility." Chapter 5 of *Handbook of Statistics*, edited by G. S. Maddala and C. R. Rao, Volume 14, 119–191. North Holland.

Gottlieb, G. and A. Kalay. 1985. "Implications of the Discreteness of Observed Stock Prices." *Journal of Finance* 40:135–153.

Härdle, W. and O. Linton. 1994. "Applied Nonparametric Methods." Chapter 38 of *Handbook of Econometrics*, edited by R. F. Engle and D. L. McFadden, Volume 4, 2293–2339. North-Holland.

Harris, L. 1986. "A Transactions Data Study of Weekly and Intradaily Patterns in Stock Returns." *Journal of Financial Economics* 16:99–117.

——— 1990. "Estimation of Stock Price Variances and Serial Covariances from Discrete Observations." *Journal of Financial and Quantitative Analysis* 25:291–306.

Hosking, J. R. M. 1981. "Fractional Differencing." *Biometrika* 68:165–176.

Hsieh, D. A. 1989. "Modeling Heteroskedasticity in Daily Foreign-Exchange Rates." *Journal of Business & Economic Statistics* 7:307–317.

——— 1991. "Chaos and Nonlinear Dynamics: Application to Financial Markets." *Journal of Finance* 46:1839–1877.

Hurvich, C. M., R Deo, and J. Brodsky. 1998. "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Long-Memory Parameter of a Long-Memory Time Series." *Journal of Time Series Analysis* 19:19–46.

Karatzas, I. and S. E. Shreve. 1988. *Brownian Motion and Stochastic Calculus.* Springer.

Kendall, M. G. and A. Stuart. 1963. *The Advanced Theory of Statistics*. 2nd Edition. Volume 1. Hafner Publishing Company.

Lockwood, L. J. and S. C. Linn. 1990. "An Examination of Stock Market Returns Volatility during Overnigh and Intraday Periods, 1964-1989." *Journal of Finance* 45:591–601.

Mandelbrot, B. B. 1963. "The Variation of Certain Speculative Prices." *Journal of Business* 36:394–419.

McDonald, J. B. and Y. J. Xu. 1995. "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics* 66:133–152.

Merton, R. C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica* 42:867–887.

Müller, U. A., R. B. Dacorogna, R. N. Olsen, O. V. Pictet, M. Schwartz, and C. Morgenegg. 1990. "Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis." *Journal of Banking and Finance* 14:1189–1208.

Nelson, D. B. 1990. "ARCH Models as Diffusion Approximations." *Journal of Econometrics* 74:31–57.

——— 1991. "Conditional Heteroskedasticity in Asset Returns: A new Approach." *Econometrica* 59:347–370.

——— 1992. "Filtering and Forecasting with Misspecified ARCH Models I" Getting the Right Variance with the Wrong Model." *Journal of Econometrics* 52:61–90.

Newey, W. K. and K. D. West. 1987. "A Simple, Positive-Definite, Heteroskedasticity Consistent Covariance Matrix." *Econometrica* 55:703–708.

Pagan, A. R. 1996. "The Econometrics of Financial Markets." *Journal of Empirical Finance* 3:15–102.

Pagan, A. R. and G. W. Schwert. 1990. "Alternative Models for Conditional Stock Volatility." *Journal of Econometrics* 45:267–290.

Palm, F. C. 1996. "GARCH Models of Volatility." Chapter 7 of *Handbook of Statistics*, edited by G. S. Maddala and C. R. Rao, Volume 14, 209–240. North-Holland.

Revuz, D. and M. Yor. 1991. *Continuous Martingales and Brownian Motion*. Springer.

Schwert, G. W. 1990. "Stock Market Volatility." *Financial Analyst Journal* 46:23–34.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Theil, H. 1961. *Economic Forecasts and Policy*. North Holland.

Tieslau, M. A., P. Schmidt, and R. T. Baillie. 1996. "The Minimum Distance Estimator for Long-Memory Processes." *Journal of Econometrics* 71:249–264.

West, K. D. and D. Cho. 1995. "The Predictive Ability of Several Models of Exchange Rate Volatility." *Journal of Econometrics* 69:367–391.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a direct Test for Heteroskedasticity." *Econometrica* 48:817–838.

Wood, R. A., T. H. McInish, and J. K. Ord. 1985. "An Investigation of Transaction Data for NYSE Stocks." *Journal of Finance* 25:723–738.

Vita

Heiko Ebens was born in Oldenburg, Germany on June 2, 1969. He received the Bachelor of Science degree in Business & Economics from the Georg-August-Universität Göttingen in 1992. After pursuing graduate studies in Finance at the Christian-Albrechts-Universität zu Kiel for one semester, he entered the University of Montana and obtained in 1994 his M.B.A. He next spent one year in the Economics Ph.D. program at the University of Kansas and then came to the Johns Hopkins University. Under guidance of Professor Pedro de Lima he conducted research in the fields of Macroeconomics, Econometrics and Financial Economics.