Practical Issues in the Analysis of Univariate GARCH Models*

Eric Zivot†

April 18, 2008

Abstract

This paper gives a tour through the empirical analysis of univariate GARCH models for financial time series with stops along the way to discuss various practical issues associated with model specification, estimation, diagnostic evaluation and forecasting.

1 Introduction

There are many very good surveys covering the mathematical and statistical properties of GARCH models. See, for example, [9], [14], [74], [76], [27] and [83]. There are also several comprehensive surveys that focus on the forecasting performance of GARCH models including [78], [77], and [3]. However, there are relatively few surveys that focus on the practical econometric issues associated with estimating GARCH models and forecasting volatility. This paper, which draws heavily from [88], gives a tour through the empirical analysis of univariate GARCH models for financial time series with stops along the way to discuss various practical issues. Multivariate GARCH models are discussed in the paper by [80]. The plan of this paper is as follows. Section 2 reviews some stylized facts of asset returns using example data on Microsoft and S&P 500 index returns. Section 3 reviews the basic univariate GARCH model. Testing for GARCH effects and estimation of GARCH models are covered in Sections 4 and 5. Asymmetric and non-Gaussian GARCH models are discussed in Section 6, and long memory GARCH models are briefly discussed in Section 7. Section 8 discusses volatility forecasting, and final remarks are given Section 9[1].

Asset	Mean	Med	Min	Max	Std. Dev	Skew	Kurt	JB
				Daily Returns				
MSFT	0.0016	0.0000	-0.3012	0.1957	0.0253	-0.2457	11.66	13693
S&P 500	0.0004	0.0005	-0.2047	0.0909	0.0113	-1.486	32.59	160848
				Monthly Returns				
MSFT	0.0336	0.0336	-0.3861	0.4384	0.1145	0.1845	4.004	9.922
S&P 500	0.0082	0.0122	-0.2066	0.1250	0.0459	-0.8377	5.186	65.75

Notes: Sample period is 03/14/86 - 06/30/03 giving 4365 daily observations.

Table 1: Summary Statistics for Daily and Monthly Stock Returns.

2 Some Stylized Facts of Asset Returns

Let P_t denote the price of an asset at the end of trading day t. The continuously compounded or log return is defined as $r_t = \ln(P_t/P_{t-1})$. Figure 1 plots the daily log returns, squared returns, and absolute value of returns of Microsoft stock and the S&P 500 index over the period March 14, 1986 through June 30, 2003. There is no clear discernible pattern of behavior in the log returns, but there is some persistence indicated in the plots of the squared and absolute returns which represent the volatility of returns. In particular, the plots show evidence of volatility clustering - low values of volatility followed by low values and high values of volatility followed by high values. This behavior is confirmed in Figure 2 which shows the sample autocorrelations of the six series. The log returns show no evidence of serial correlation, but the squared and absolute returns are positively autocorrelated. Also, the decay rates of the sample autocorrelations of r_t^2 and $|r_t|$ appear much slower, especially for the S&P 500 index, than the exponential rate of a covariance stationary autoregressive-moving average (ARMA) process suggesting possible long memory behavior. Monthly returns, defined as the sum of daily returns over the month, are illustrated in Figure 3. The monthly returns display much less volatility clustering than the daily returns.

Table 1 gives some standard summary statistics along with the Jarque-Bera test for normality. The latter is computed as

$$\mathrm{JB} = \frac{T}{6}\left(\widehat{\mathrm{skew}}^2 + \frac{(\widehat{\mathrm{kurt}}-3)^2}{4}\right), \tag{1}$$

where $\widehat{\mathrm{skew}}$ denotes the sample skewness and $\widehat{\mathrm{kurt}}$ denotes the sample kurtosis. Under the null that the data are iid normal, JB is asymptotically distributed as chi-square

*The paper was prepared for the *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J-P Kreiss, and T. Mikosch. Thanks to Saraswata Chaudhuri, Richard Davis, Ron Schoenberg and Jiahui Wang for helpful comments and suggestions. Financial support from the Gary Waterman Distinguished Scholarship is greatly appreciated.

†Department of Economics, Box 353330, University of Washington. ezivot@u.washington.edu.

[1] All of the examples in the paper were constructed using S-PLUS 8.0 and S+FinMetrics 2.0. Script files for replicating the examples may be downloaded from http://faculty.washington.edu/ezivot.

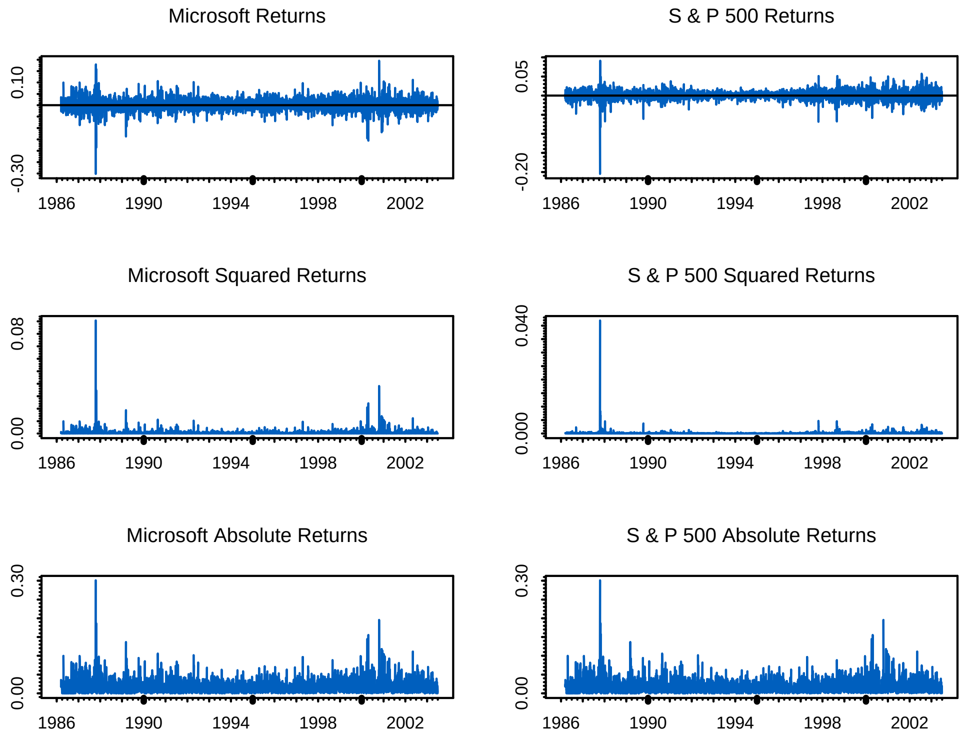

Figure 1: Daily returns, squared returns and absolute returns for Microsoft and the S&P 500 index.

with 2 degrees of freedom. The distribution of daily returns is clearly non-normal with negative skewness and pronounced excess kurtosis. Part of this non-normality is caused by some large outliers around the October 1987 stock market crash and during the bursting of the 2000 tech bubble. However, the distribution of the data still appears highly non-normal even after the removal of these outliers. Monthly returns have a distribution that is much closer to the normal than daily returns.

3 The ARCH and GARCH Model

[33] showed that the serial correlation in squared returns, or conditional heteroskedasticity, can be modeled using an autoregressive conditional heteroskedasticity (ARCH) model of the form

$$y_t = E_{t-1}[y_t] + \epsilon_t, \tag{2}$$

$$\epsilon_t = z_t \sigma_t, \tag{3}$$

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + \cdots + a_p \epsilon_{t-p}^2, \tag{4}$$

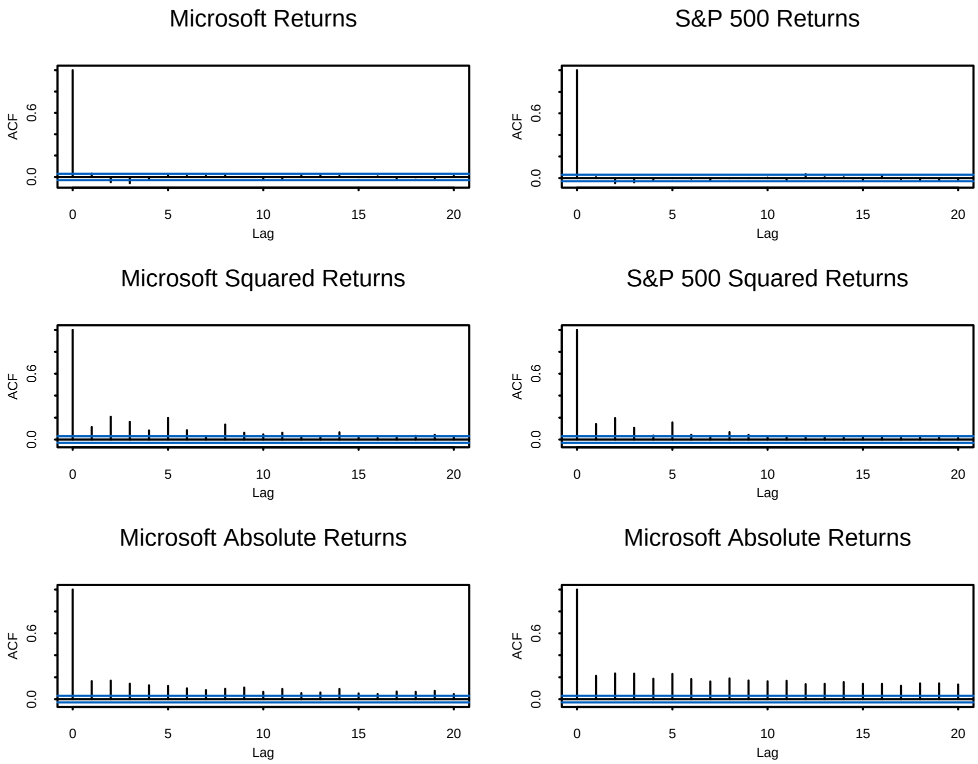

Figure 2: Sample autocorrelations of r_t, r_t^2 and $|r_t|$ for Microsoft and S&P 500 index.

where $E_{t-1}[\cdot]$ represents expectation conditional on information available at time $t-1$, and z_t is a sequence of iid random variables with mean zero and unit variance. In the basic ARCH model z_t is assumed to be iid standard normal. The restrictions $a_0 > 0$ and $a_i \geq 0$ $(i = 1, \ldots, p)$ are required for $\sigma_t^2 > 0$. The representation (2) - (4) is convenient for deriving properties of the model as well as for specifying the likelihood function for estimation. The equation for σ_t^2 can be rewritten as an AR(p) process for ϵ_t^2

$$\epsilon_t^2 = a_0 + a_1\epsilon_{t-1}^2 + \cdots + a_p\epsilon_{t-p}^2 + u_t, \tag{5}$$

where $u_t = \epsilon_t^2 - \sigma_t^2$ is a martingale difference sequence (MDS) since $E_{t-1}[u_t] = 0$ and it is assumed that $E(\epsilon_t^2) < \infty$. If $a_1 + \cdots + a_p < 1$ then ϵ_t is covariance stationary, the persistence of ϵ_t^2 and σ_t^2 is measured by $a_1 + \cdots + a_p$ and $\bar{\sigma}^2 = \text{var}(\epsilon_t) = E(\epsilon_t^2) = a_0/(1 - a_1 - \cdots - a_p)$.

An important extension of the ARCH model proposed by [12] replaces the AR(p) representation in (4) with an ARMA(p, q) formulation

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i\epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j\sigma_{t-j}^2, \tag{6}$$

where the coefficients a_i $(i = 0, \cdots, p)$ and b_j $(j = 1, \cdots, q)$ are all assumed to be

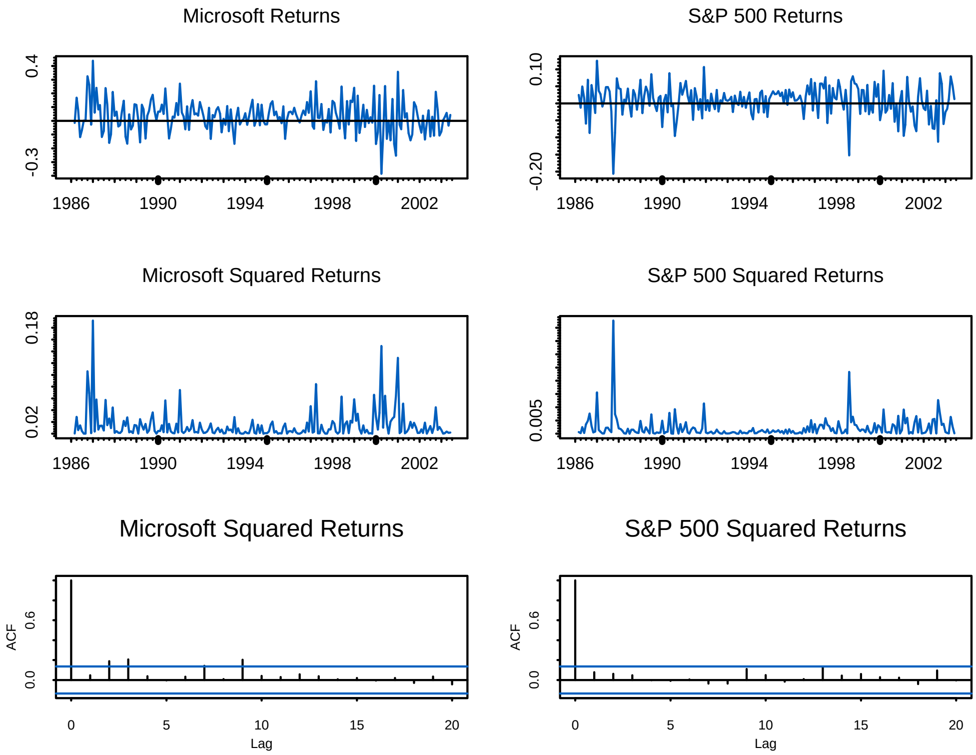

Figure 3: Monthly Returns, Squared Returns and Sample Autocorrelations of Squared Returns for Microsoft and the S&P 500.

positive to ensure that the conditional variance σ_t^2 is always positive.[2] The model in (6) together with (2)-(3) is known as the generalized ARCH or GARCH(p,q) model. The GARCH(p,q) model can be shown to be equivalent to a particular ARCH(∞) model. When $q = 0$, the GARCH model reduces to the ARCH model. In order for the GARCH parameters, b_j $(j = 1, \cdots, q)$, to be identified at least one of the ARCH coefficients a_i $(i > 0)$ must be nonzero. Usually a GARCH(1,1) model with only three parameters in the conditional variance equation is adequate to obtain a good model fit for financial time series. Indeed, [49] provided compelling evidence that is difficult to find a volatility model that outperforms the simple GARCH(1,1).

Just as an ARCH model can be expressed as an AR model of squared residuals, a GARCH model can be expressed as an ARMA model of squared residuals. Consider the GARCH(1,1) model

$$\sigma_t^2 = a_0 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2. \tag{7}$$

Since $E_{t-1}(\epsilon_t^2) = \sigma_t^2$, (7) can be rewritten as

$$\epsilon_t^2 = a_0 + (a_1 + b_1)\epsilon_{t-1}^2 + u_t - b_1 u_{t-1}, \tag{8}$$

[2]Positive coefficients are sufficient but not necessary conditions for the positivity of conditional variance. See [72] and [23] for more general conditions.

which is an ARMA(1,1) model with $u_t = \epsilon_t^2 - E_{t-1}(\epsilon_t^2)$ being the MDS disturbance term.

Given the ARMA(1,1) representation of the GARCH(1,1) model, many of its properties follow easily from those of the corresponding ARMA(1,1) process for ϵ_t^2. For example, the persistence of σ_t^2 is captured by $a_1 + b_1$ and covariance stationarity requires that $a_1 + b_1 < 1$. The covariance stationary GARCH(1,1) model has an ARCH(∞) representation with $a_i = a_1 b_1^{i-1}$, and the unconditional variance of ϵ_t is $\bar{\sigma}^2 = a_0/(1 - a_1 - b_1)$.

For the general GARCH(p, q) model (6), the squared residuals ϵ_t behave like an ARMA($\max(p, q), q$) process. Covariance stationarity requires $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_i < 1$ and the unconditional variance of ϵ_t is

$$\bar{\sigma}^2 = \text{var}(\epsilon_t) = \frac{a_0}{1 - \left(\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_i\right)}. \tag{9}$$

3.1 Conditional Mean Specification

Depending on the frequency of the data and the type of asset, the conditional mean $E_{t-1}[y_t]$ is typically specified as a constant or possibly a low order autoregressive-moving average (ARMA) process to capture autocorrelation caused by market microstructure effects (e.g., bid-ask bounce) or non-trading effects. If extreme or unusual market events have happened during sample period, then dummy variables associated with these events are often added to the conditional mean specification to remove these effects. Therefore, the typical conditional mean specification is of the form

$$E_{t-1}[y_t] = c + \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{j=1}^{s} \theta_j \epsilon_{t-j} + \sum_{l=0}^{L} \boldsymbol{\beta}_l' \mathbf{x}_{t-l} + \epsilon_t, \tag{10}$$

where $\mathbf{x}_t$ is a $k \times 1$ vector of exogenous explanatory variables.

In financial investment, high risk is often expected to lead to high returns. Although modern capital asset pricing theory does not imply such a simple relationship, it does suggest that there are some interactions between expected returns and risk as measured by volatility. Engle, Lilien and Robins (1987) proposed to extend the basic GARCH model so that the conditional volatility can generate a risk premium which is part of the expected returns. This extended GARCH model is often referred to as GARCH-in-the-mean or GARCH-M model. The GARCH-M model extends the conditional mean equation (10) to include the additional regressor $g(\sigma_t)$, which can be an arbitrary function of conditional volatility σ_t. The most common specifications are $g(\sigma_t) = \sigma_t^2$, σ_t, or $\ln(\sigma_t^2)$.

3.2 Explanatory Variables in the Conditional Variance Equation

Just as exogenous variables may be added to the conditional mean equation, exogenous explanatory variables may also be added to the conditional variance formula (6) in a straightforward way giving

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 + \sum_{k=1}^{K} \boldsymbol{\delta}_k' \mathbf{z}_{t-k},$$

where $\mathbf{z}_t$ is a $m \times 1$ vector of variables, and $\boldsymbol{\delta}$ is a $m \times 1$ vector of positive coefficients. Variables that have been shown to help predict volatility are trading volume, macroeconomic news announcements ([58], [43], [17]), implied volatility from option prices and realized volatility ([82], [11]), overnight returns ([46], [68]), and after hours realized volatility ([21])

3.3 The GARCH Model and Stylized Facts of Asset Returns

Previously it was shown that the daily returns on Microsoft and the S&P 500 exhibited the "stylized facts" of volatility clustering as well as a non-normal empirical distribution. Researchers have documented these and many other stylized facts about the volatility of economic and financial time series. [14] gave a complete account of these facts. Using the ARMA representation of GARCH models shows that the GARCH model is capable of explaining many of those stylized facts. The four most important ones are: volatility clustering, fat tails, volatility mean reversion, and asymmetry.

To understand volatility clustering, consider the GARCH$(1,1)$ model in (7). Usually the GARCH coefficient b_1 is found to be around 0.9 for many daily or weekly financial time series. Given this value of b_1, it is obvious that large values of σ_{t-1}^2 will be followed by large values of σ_t^2, and small values of σ_{t-1}^2 will be followed by small values of σ_t^2. The same reasoning can be obtained from the ARMA representation in (8), where large/small changes in ϵ_{t-1}^2 will be followed by large/small changes in ϵ_t^2.

It is well known that the distribution of many high frequency financial time series usually have fatter tails than a normal distribution. That is, extreme values occur more often than implied by a normal distribution. [12] gave the condition for the existence of the fourth order moment of a GARCH$(1,1)$ process. Assuming the fourth order moment exists, [12] showed that the kurtosis implied by a GARCH$(1,1)$ process with normal errors is greater than 3, the kurtosis of a normal distribution. [51] and [52] extended these results to general GARCH(p,q) models. Thus a GARCH model with normal errors can replicate some of the fat-tailed behavior observed in financial time series. A more thorough discussion of extreme value theory for GARCH is given by [24]. Most often a GARCH model with a non-normal error distribution is required to fully capture the observed fat-tailed behavior in returns. These models are reviewed in sub-Section 6.2.

Although financial markets may experience excessive volatility from time to time, it appears that volatility will eventually settle down to a long run level. Recall, the unconditional variance of ϵ_t for the stationary GARCH$(1,1)$ model is $\bar{\sigma}^2 = a_0/(1 - a_1 - b_1)$. To see that the volatility is always pulled toward this long run, the ARMA representation in (8) may be rewritten in mean-adjusted form as:

$$(\epsilon_t^2 - \bar{\sigma}^2) = (a_1 + b_1)(\epsilon_{t-1}^2 - \bar{\sigma}^2) + u_t - b_1 u_{t-1}. \tag{11}$$

If the above equation is iterated k times, it follows that

$$(\epsilon_{t+k}^2 - \bar{\sigma}^2) = (a_1 + b_1)^k (\epsilon_t^2 - \bar{\sigma}^2) + \eta_{t+k},$$

where η_t is a moving average process. Since $a_1 + b_1 < 1$ for a covariance stationary GARCH(1,1) model, $(a_1 + b_1)^k \to 0$ as $k \to \infty$. Although at time t there may be a large deviation between ϵ_t^2 and the long run variance, $\epsilon_{t+k}^2 - \bar{\sigma}^2$ will approach zero "on average" as k gets large; i.e., the volatility "mean reverts" to its long run level $\bar{\sigma}^2$. The magnitude of $a_1 + b_1$ controls the speed of mean reversion. The so-called half-life of a volatility shock, defined as $\ln(0.5)/\ln(a_1 + b_1)$, measures the average time it takes for $|\epsilon_t^2 - \bar{\sigma}^2|$ to decrease by one half. Obviously, the closer $a_1 + b_1$ is to one the longer is the half-life of a volatility shock. If $a_1 + b_1 > 1$, the GARCH model is non-stationary and the volatility will eventually explode to infinity as $k \to \infty$. Similar arguments can be easily constructed for a GARCH(p, q) model.

The standard GARCH(p, q) model with Gaussian errors implies a symmetric distribution for y_t and so cannot account for the observed asymmetry in the distribution of returns. However, as shown in Section 6, asymmetry can easily be built into the GARCH model by allowing ϵ_t to have an asymmetric distribution or by explicitly modeling asymmetric behavior in the conditional variance equation (6).

3.4 Temporal Aggregation

Volatility clustering and non-Gaussian behavior in financial returns is typically seen in weekly, daily or intraday data. The persistence of conditional volatility tends to increase with the sampling frequency[3]. However, as shown in [32], for GARCH models there is no simple aggregation principle that links the parameters of the model at one sampling frequency to the parameters at another frequency. This occurs because GARCH models imply that the squared residual process follows an ARMA type process with MDS innovations which is not closed under temporal aggregation. The practical result is that GARCH models tend to be fit to the frequency at hand. This strategy, however, may not provide the best out-of-sample volatility forecasts. For example, [68] showed that a GARCH model fit to S&P 500 daily returns produces better forecasts of weekly and monthly volatility than GARCH models fit to weekly or monthly returns, respectively.

4 Testing for ARCH/GARCH effects

The stylized fact of volatility clustering in returns manifests itself as autocorrelation in squared and absolute returns or in the residuals from the estimated conditional mean equation (10). The significance of these autocorrelations may be tested using

[3] The empirical result that aggregated returns exhibit smaller GARCH effects and approach Gaussian behavior can be explained by the results of [26] who showed that a central limit theorem holds for standardized sums of random variables that follow covariance stationary GARCH processes.

the Ljung-Box or modified Q-statistic

$$\mathrm{MQ}(p) = T(T+2)\sum_{j=1}^{p}\frac{\hat{\rho}_j^2}{T-j}, \tag{12}$$

where $\hat{\rho}_j$ denotes the j-lag sample autocorrelation of the squared or absolute returns. If the data are white noise then the MQ(p) statistic has an asymptotic chi-square distribution with p degrees of freedom. A significant value for MQ(p) provides evidence for time varying conditional volatility.

To test for autocorrelation in the raw returns when it is suspected that there are GARCH effects present, [27] suggested using the following heteroskedasticity robust version of (12)

$$\mathrm{MQ}^{HC}(p) = T(T+2)\sum_{j=1}^{p}\frac{1}{T-j}\left(\frac{\hat{\sigma}^4}{\hat{\sigma}^4+\hat{\gamma}_j}\right)\hat{\rho}_j^2,$$

where $\hat{\sigma}^4$ is a consistent estimate of the squared unconditional variance of returns, and $\hat{\gamma}_j$ is the sample autocovariance of squared returns.

Since an ARCH model implies an AR model for the squared residuals ϵ_t^2, [33] showed that a simple Lagrange multiplier (LM) test for ARCH effects can be constructed based on the auxiliary regression (5). Under the null hypothesis that there are no ARCH effects, $a_1 = a_2 = \cdots = a_p = 0$, the test statistic

$$\mathrm{LM} = T \cdot R^2 \tag{13}$$

has an asymptotic chi-square distribution with p degrees of freedom, where T is the sample size and R^2 is computed from the regression (5) using estimated residuals. Even though the LM test is constructed from an ARCH model, [61] show that it also has power against more general GARCH alternatives and so it can be used as a general specification test for GARCH effects.

[64], however, argued that the LM test (13) may reject if there is general misspecification in the conditional mean equation (10). They showed that such misspecification causes the estimated residuals $\hat{\epsilon}_t$ to be serially correlated which, in turn, causes $\hat{\epsilon}_t^2$ to be serially correlated. Therefore, care should be exercised in specifying the conditional mean equation (10) prior to testing for ARCH effects.

4.1 Testing for ARCH Effects in Daily and Monthly Returns

Table 2 shows values of MQ(p) computed from daily and monthly squared returns and the LM test for ARCH, for various values of p, for Microsoft and the S&P 500. There is clear evidence of volatility clustering in the daily returns, but less evidence for monthly returns especially for the S&P 500.

5 Estimation of GARCH Models

The general GARCH(p, q) model with normal errors is (2), (3) and (6) with $z_t \sim$ iid $N(0, 1)$. For simplicity, assume that $E_{t-1}[y_t] = c$. Given that ϵ_t follows Gaussian

		MQ(p) r_t^2			LM		
Asset	p	1	5	10	1	5	10
				Daily Returns			
MSFT		56.81	562.1	206.8	56.76	377.9	416.6
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
S&P 500		87.59	415.5	456.1	87.52	311.4	329.8
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
				Monthly Returns			
MSFT		0.463	17.48	31.59	0.455	16.74	33.34
		(0.496)	(0.003)	(0.000)	(0.496)	(0.005)	(0.000)
S&P 500		1.296	2.590	6.344	1.273	2.229	5.931
		(0.255)	(0.763)	(0.786)	(0.259)	(0.817)	(0.821)

Notes: p-values are in parentheses.

Table 2: Tests for ARCH Effects in Daily Stock Returns

distribution conditional on past history, the prediction error decomposition of the log-likelihood function of the GARCH model conditional on initial values is

$$\log L = \sum_{t=1}^{T} l_t = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{\epsilon_t^2}{\sigma_t^2}, \tag{14}$$

where $l_t = -\frac{1}{2}(\log(2\pi) + \log\sigma_t^2) - \frac{1}{2}\frac{\epsilon_t^2}{\sigma_t^2}$. The conditional loglikelihood (14) is used in practice since the unconditional distribution of the initial values is not known in closed form[4]. As discussed in [69] and [20], there are several practical issues to consider in the maximization of (14). Starting values for the model parameters c, a_i $(i = 0, \cdots, p)$ and b_j $(j = 1, \cdots, q)$ need to be chosen and an initialization of ϵ_t^2 and σ_t^2 must be supplied. The sample mean of y_t is usually used as the starting value for c, zero values are often given for the conditional variance parameters other than a_0 and a_1, and a_0 is set equal to the unconditional variance of y_t[5]. For the initial values of σ_t^2, a popular choice is

$$\sigma_t^2 = \epsilon_t^2 = \frac{1}{T}\sum_{s=1}^{T}\epsilon_s^2, \ t \leq 0,$$

where the initial values for ϵ_s are computed as the residuals from a regression of y_t on a constant.

Once the log-likelihood is initialized, it can be maximized using numerical optimization techniques. The most common method is based on a Newton-Raphson iteration of the form

$$\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_n - \lambda_n \mathbf{H}(\hat{\boldsymbol{\theta}}_n)^{-1}\mathbf{s}(\hat{\boldsymbol{\theta}}_n),$$

[4] [29] gave a computationally intensive numerical procedure for approximating the exact log-likelihood.

[5] Setting the starting values for all of the ARCH coefficients a_i $(i = 1, \ldots, p)$ to zero may create an ill-behaved likelihood and lead to a local minimum since the remaining GARCH parameters are not identified.

where $\boldsymbol{\theta}_n$ denotes the vector of estimated model parameters at iteration n, λ_n is a scalar step-length parameter, and $\mathbf{s}(\boldsymbol{\theta}_n)$ and $\mathbf{H}(\boldsymbol{\theta}_n)$ denote the gradient (or score) vector and Hessian matrix of the log-likelihood at iteration n, respectively. The step length parameter λ_n is chosen such that $\ln L(\boldsymbol{\theta}_{n+1}) \geq \ln L(\boldsymbol{\theta}_n)$. For GARCH models, the BHHH algorithm is often used. This algorithm approximates the Hessian matrix using only first derivative information

$$-\mathbf{H}(\boldsymbol{\theta}) \approx \mathbf{B}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \frac{\partial l_t}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \boldsymbol{\theta}'}.$$

In the application of the Newton-Raphson algorithm, analytic or numerical derivatives may be used. [41] provided algorithms for computing analytic derivatives for GARCH models.

The estimates that maximize the conditional log-likelihood (14) are called the maximum likelihood (ML) estimates. Under suitable regularity conditions, the ML estimates are consistent and asymptotically normally distributed and an estimate of the asymptotic covariance matrix of the ML estimates is constructed from an estimate of the final Hessian matrix from the optimization algorithm used. Unfortunately, verification of the appropriate regularity conditions has only been done for a limited number of simple GARCH models, see [63], [60], [55], [56] and [81]. In practice, it is generally assumed that the necessary regularity conditions are satisfied.

In GARCH models for which the distribution of z_t is symmetric and the parameters of the conditional mean and variance equations are variation free, the information matrix of the log-likelihood is block diagonal. The implication of this is that the parameters of the conditional mean equation can be estimated separately from those of the conditional variance equation without loss of asymptotic efficiency. This can greatly simplify estimation. An common model for which block diagonality of the information matrix fails is the GARCH-M model.

5.1 Numerical Accuracy of GARCH Estimates

GARCH estimation is widely available in a number of commercial software packages (e.g. EVIEWS, GAUSS, MATLAB, Ox, RATS, S-PLUS, TSP) and there are also a few free open source implementations. [41], [69], and [20] discussed numerical accuracy issues associated with maximizing the GARCH log-likelihood. They found that starting values, optimization algorithm choice, and use of analytic or numerical derivatives, and convergence criteria all influence the resulting numerical estimates of the GARCH parameters. [69] and [20] studied estimation of a GARCH(1,1) model from a variety of commercial statistical packages using the exchange rate data of [15] as a benchmark. They found that it is often difficult to compare competing software since the exact construction of the GARCH likelihood is not always adequately described. In general, they found that use of analytic derivatives leads to more accurate estimation than procedures based on purely numerical evaluations.

In practice, the GARCH log-likelihood function is not always well behaved, especially in complicated models with many parameters, and reaching a global maximum of the log-likelihood function is not guaranteed using standard optimization techniques. Also, the positive variance and stationarity constraints are not straightforward to implement with common optimization software and are often ignored in practice. Poor choice of starting values can lead to an ill-behaved log-likelihood and cause convergence problems. Therefore, it is always a good idea to explore the surface of the log-likelihood by perturbing the starting values and re-estimating the GARCH parameters.

In many empirical applications of the GARCH(1,1) model, the estimate of a_1 is close to zero and the estimate of b_1 is close to unity. This situation is of some concern since the GARCH parameter b_1 becomes unidentified if $a_1 = 0$, and it is well known that the distribution of ML estimates can become ill-behaved in models with nearly unidentified parameters. [66] studied the accuracy of ML estimates of the GARCH parameters a_0, a_1 and b_1 when a_1 is close to zero. They found that the estimated standard error for b_1 is spuriously small and that the t-statistics for testing hypotheses about the true value of b_1 are severely size distorted. They also showed that the concentrated loglikelihood as a function of b_1 exhibits multiple maxima. To guard against spurious inference they recommended comparing estimates from pure ARCH(p) models, which do not suffer from the identification problem, with estimates from the GARCH(1,1). If the volatility dynamics from these models are similar then the spurious inference problem is not likely to be present.

5.2 Quasi-Maximum Likelihood Estimation

Another practical issue associated with GARCH estimation concerns the correct choice of the error distribution. In particular, the assumption of conditional normality is not always appropriate. However, as shown by [86] and [16], even when normality is inappropriately assumed, maximizing the Gaussian log-likelihood (14) results in quasi-maximum likelihood estimates (QMLEs) that are consistent and asymptotically normally distributed provided the conditional mean and variance functions of the GARCH model are correctly specified. In addition, [16] derived an asymptotic covariance matrix for the QMLEs that is robust to conditional non-normality. This matrix is estimated using

$$\mathbf{H}(\hat{\boldsymbol{\theta}}_{QML})^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}}_{QML})\mathbf{H}(\hat{\boldsymbol{\theta}}_{QML})^{-1}, \tag{15}$$

where $\hat{\boldsymbol{\theta}}_{QML}$ denotes the QMLE of $\boldsymbol{\theta}$, and is often called the "sandwich" estimator. The coefficient standard errors computed from the square roots of the diagonal elements of (15) are sometimes called "Bollerslev-Wooldridge" standard errors. Of course, the QMLEs will be less efficient than the true MLEs based on the correct error distribution. However, if the normality assumption is correct then the sandwich covariance is asymptotically equivalent to the inverse of the Hessian. As a result, it is good practice to routinely use the sandwich covariance for inference purposes.

[35] and [16] evaluated the accuracy of the quasi-maximum likelihood estimation of GARCH(1,1) models. They found that if the distribution of z_t in (3) is symmetric, then QMLE is often close to the MLE. However, if z_t has a skewed distribution then the QMLE can be quite different from the MLE.

5.3 Model Selection

An important practical problem is the determination of the ARCH order p and the GARCH order q for a particular series. Since GARCH models can be treated as ARMA models for squared residuals, traditional model selection criteria such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) can be used for selecting models. For daily returns, if attention is restricted to pure ARCH(p) models it is typically found that large values of p are selected by AIC and BIC. For GARCH(p, q) models, those with $p, q \leq 2$ are typically selected by AIC and BIC. Low order GARCH(p,q) models are generally preferred to a high order ARCH(p) for reasons of parsimony and better numerical stability of estimation (high order GARCH(p, q) processes often have many local maxima and minima). For many applications, it is hard to beat the simple GARCH(1,1) model.

5.4 Evaluation of Estimated GARCH models

After a GARCH model has been fit to the data, the adequacy of the fit can be evaluated using a number of graphical and statistical diagnostics. If the GARCH model is correctly specified, then the estimated standardized residuals $\hat{\epsilon}_t/\hat{\sigma}_t$ should behave like classical regression residuals; i.e., they should not display serial correlation, conditional heteroskedasticity or any type of nonlinear dependence. In addition, the distribution of the standardized residuals $\hat{\epsilon}_t/\hat{\sigma}_t$ should match the specified error distribution used in the estimation.

Graphically, ARCH effects reflected by serial correlation in $\hat{\epsilon}_t^2/\hat{\sigma}_t^2$ can be uncovered by plotting its SACF. The modified Ljung-Box statistic (12) can be used to test the null of no autocorrelation up to a specific lag, and Engle's LM statistic (13) can be used to test the null of no remaining ARCH effects[6]. If it is assumed that the errors are Gaussian, then a plot of $\hat{\epsilon}_t/\hat{\sigma}_t$ against time should have roughly ninety five percent of its values between ± 2; a normal qq-plot of $\hat{\epsilon}_t/\hat{\sigma}_t$ should look roughly linear[7]; and the JB statistic should not be too much larger than six.

5.5 Estimation of GARCH Models for Daily and Monthly Returns

Table 3 gives model selection criteria for a variety of GARCH(p, q) fitted to the daily returns on Microsoft and the S&P 500. For pure ARCH(p) models, an ARCH(5) is chosen by all criteria for both series. For GARCH(p, q) models, AIC picks a GARCH(2,1) for both series and BIC picks a GARCH(1,1) for both series[8].

Table 4 gives QMLEs of the GARCH(1,1) model assuming normal errors for the Microsoft and S&P 500 daily returns. For both series, the estimates of a_1 are around

---

[6] These tests should be viewed as indicative, since the distribution of the tests are influenced by the estimation of the GARCH model. For valid LM tests, the partial derivatives of σ_t^2 with respect to the conditional volatility parameters should be added as additional regressors in the auxiliary regression (5) based on estimated residuals.

[7] If an error distribution other than the Gaussian is assumed, then the qq-plot should be constructed using the quantiles of the assumed distribution.

[8] The low log-likelihood values for the GARCH(2,2) models indicate that a local maximum was reached.

(p,q)	Asset	AIC	BIC	Likelihood
(1,0)	MSFT	-19977	-19958	9992
	S&P 500	-27337	-27318	13671
(2,0)	MSFT	-20086	-20060	10047
	S&P 500	-27584	-27558	13796
(3,0)	MSFT	-20175	-20143	10092
	S&P 500	-27713	-27681	13861
(4,0)	MSFT	-20196	-20158	10104
	S&P 500	-27883	-27845	13947
(5,0)	MSFT	-20211	-20166	10113
	S&P 500	-27932	-27887	13973
(1,1)	MSFT	-20290	-20264	10149
	S&P 500	-28134	-28109	14071
(1,2)	MSFT	-20290	-20258	10150
	S&P 500	-28135	-28103	14072
(2,1)	MSFT	-20292	-20260	10151
	S&P 500	-28140	-28108	14075
(2,2)	MSFT	-20288	-20249	10150
	S&P 500	-27858	-27820	13935

Table 3: Model Selection Criteria for Estimated GARCH(p,q) Models.

0.09 and the estimates of b_1 are around 0.9. Using both ML and QML standard errors, these estimates are statistically different from zero. However, the QML standard errors are considerably larger than the ML standard errors. The estimated volatility persistence, $a_1 + b_1$, is very high for both series and implies half-lives of shocks to volatility to Microsoft and the S&P 500 of 15.5 days and 76 days, respectively. The unconditional standard deviation of returns, $\bar{\sigma} = \sqrt{a_0/(1 - a_1 - b_1)}$, for Microsoft and the S&P 500 implied by the GARCH(1,1) models are 0.0253 and 0.0138, respectively, and are very close to the sample standard deviations of returns reported in Table 1.

Estimates of GARCH-M(1,1) models for Microsoft and the S&P 500, where σ_t is added as a regressor to the mean equation, show small positive coefficients on σ_t and essentially the same estimates for the remaining parameters as the GARCH(1,1) models.

Figure 4 shows the first differences of returns along with the fitted one-step-ahead volatilities, $\hat{\sigma}_t$, computed from the GARCH(1,1) and ARCH(5) models. The ARCH(5) and GARCH(1,1) models do a good job of capturing the observed volatility clustering in returns. The GARCH(1,1) volatilities, however, are smoother and display more persistence than the ARCH(5) volatilities.

Graphical diagnostics from the fitted GARCH(1,1) models are illustrated in Figure 5. The SACF of $\hat{\epsilon}_t^2/\hat{\sigma}_t^2$ does not indicate any significant autocorrelation, but the normal qq-plot of $\hat{\epsilon}_t/\hat{\sigma}_t$ shows strong departures from normality. The last three columns of Table 4 give the standard statistical diagnostics of the fitted GARCH

	GARCH Parameters			Residual Diagnostics		
Asset	a_0	a_1	b_1	MQ(12)	LM(12)	JB
			Daily Returns			
MSFT	$2.80e^{-5}$ $(3.42e^{-6})$ $[1.10e^{-5}]$	0.0904 (0.0059) [0.0245]	0.8658 (0.0102) [0.0371]	4.787 (0.965)	4.764 (0.965)	1751 (0.000)
S&P 500	$1.72e^{-6}$ $(2.00e^{-7})$ $[1.25e^{-6}]$	0.0919 (0.0029) [0.0041]	0.8990 (0.0046) [0.0436]	5.154 (0.953)	5.082 (0.955)	5067 (0.000)
			Monthly Returns			
MSFT	0.0006 [0.0006]	0.1004 [0.0614]	0.8525 [0.0869]	8.649 (0.733)	6.643 (0.880)	3.587 (0.167)
S&P 500	$3.7e^{-5}$ $[9.6e^{-5}]$	0.0675 [0.0248]	0.9179 [0.0490]	3.594 (0.000)	3.660 (0.988)	72.05 (0.000)

Notes: QML standard errors are in brackets.

Table 4: Estimates of GARCH(1,1) Model with Diagnostics.

models. Consistent with the SACF, the MQ statistic and Engle's LM statistic do not indicate remaining ARCH effects. Furthermore, the extremely large JB statistic confirms nonnormality.

Table 4 also shows estimates of GARCH(1,1) models fit to the monthly returns. The GARCH(1,1) models fit to the monthly returns are remarkable similar to those fit to the daily returns. There are, however, some important differences. The monthly standardized residuals are much closer to the normal distribution, especially for Microsoft. Also, the GARCH estimates for the S&P 500 reflect some of the characteristics of spurious GARCH effects as discussed in [66]. In particular, the estimate of a_1 is close to zero, and has a relatively large QML standard error, and the estimate of b_1 is close to one and has a very small standard error.

6 GARCH Model Extensions

In many cases, the basic GARCH conditional variance equation (6) under normality provides a reasonably good model for analyzing financial time series and estimating conditional volatility. However, in some cases there are aspects of the model which can be improved so that it can better capture the characteristics and dynamics of a particular time series. For example, the empirical analysis in the previous Section showed that for the daily returns on Microsoft and the S&P 500, the normality assumption may not be appropriate and there is evidence of nonlinear behavior in the standardized residuals from the fitted GARCH(1,1) model. This Section discusses several extensions to the basic GARCH model that make GARCH modeling more flexible.

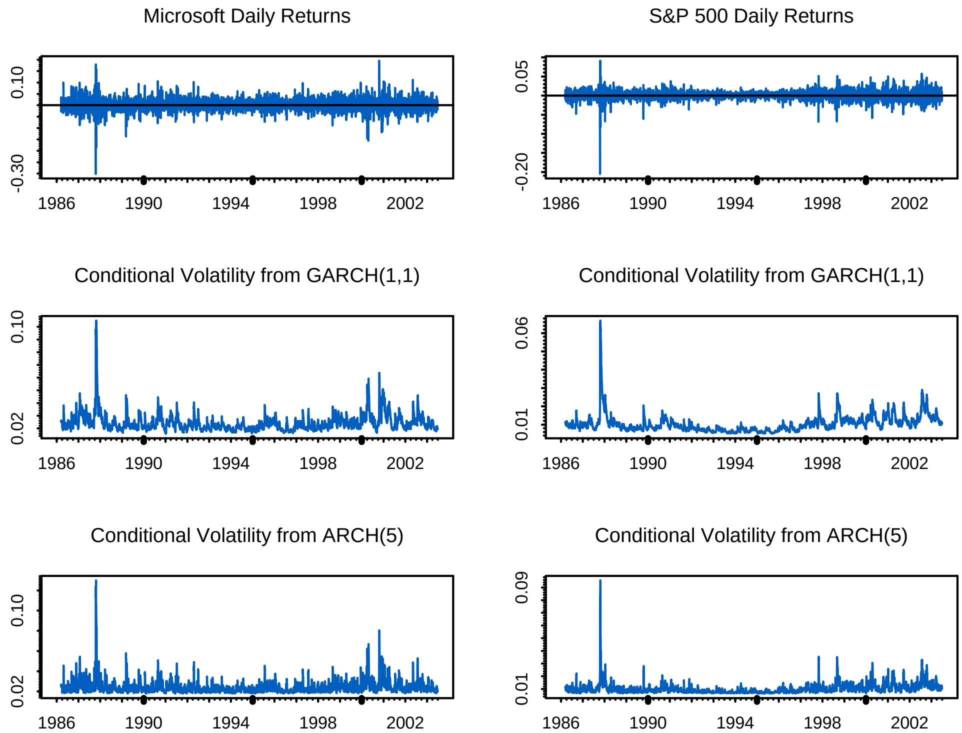

Figure 4: One-step ahead volatilities from fitted ARCH(5) and GARCH(1,1) models for Microsoft and S&P 500 index.

6.1 Asymmetric Leverage Effects and News Impact

In the basic GARCH model (6), since only squared residuals ϵ_{t-i}^2 enter the conditional variance equation, the signs of the residuals or shocks have no effect on conditional volatility. However, a stylized fact of financial volatility is that bad news (negative shocks) tends to have a larger impact on volatility than good news (positive shocks). That is, volatility tends to be higher in a falling market than in a rising market. [10] attributed this effect to the fact that bad news tends to drive down the stock price, thus increasing the leverage (i.e., the debt-equity ratio) of the stock and causing the stock to be more volatile. Based on this conjecture, the asymmetric news impact on volatility is commonly referred to as the leverage effect.

6.1.1 Testing for Asymmetric Effects on Conditional Volatility

A simple diagnostic for uncovering possible asymmetric leverage effects is the sample correlation between r_t^2 and r_{t-1}. A negative value of this correlation provides some evidence for potential leverage effects. Other simple diagnostics, suggested by [39],

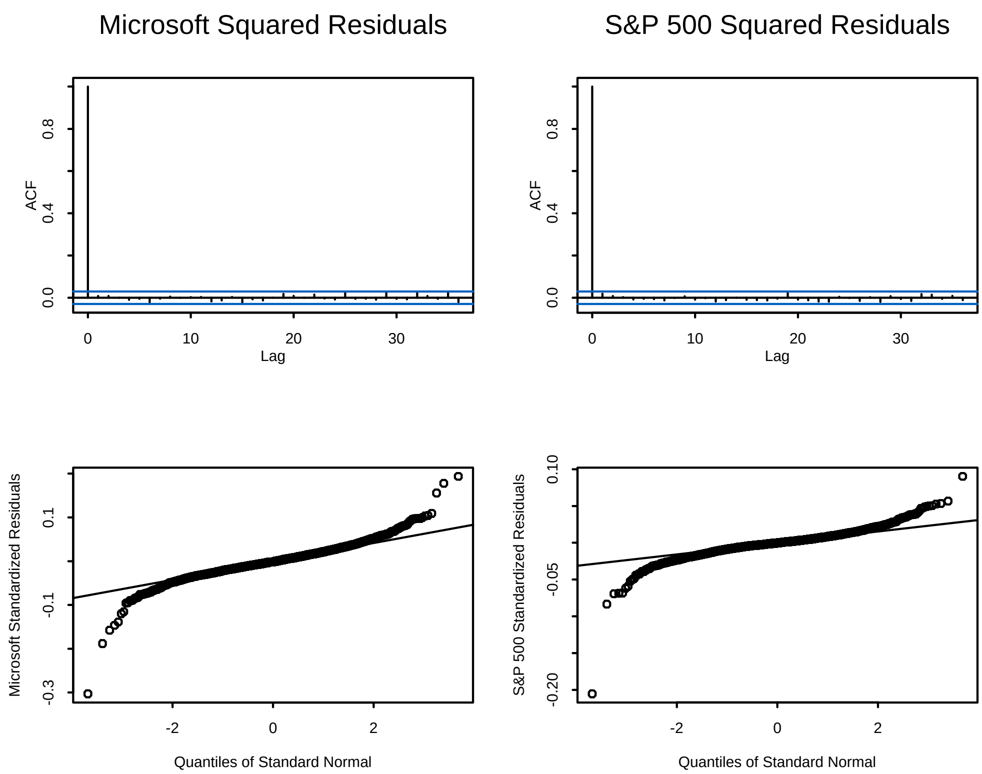

Figure 5: Graphical residual diagnostics from fitted GARCH(1,1) models to Microsoft and S&P 500 returns.

result from estimating the following test regression

$$\hat{\varepsilon}_t^2 = \beta_0 + \beta_1 \hat{w}_{t-1} + \xi_t,$$

where $\hat{\varepsilon}_t$ is the estimated residual from the conditional mean equation (10), and $\hat{w}_{t-1}$ is a variable constructed from $\hat{\varepsilon}_{t-1}$ and the sign of $\hat{\varepsilon}_{t-1}$. A significant value of β_1 indicates evidence for asymmetric effects on conditional volatility. Let S_{t-1}^- denote a dummy variable equal to unity when $\hat{\varepsilon}_{t-1}$ is negative, and zero otherwise. Engle and Ng consider three tests for asymmetry. Setting $\hat{w}_{t-1} = S_{t-1}^-$ gives the Sign Bias test; setting $\hat{w}_{t-1} = S_{t-1}^- \hat{\varepsilon}_{t-1}$ gives the Negative Size Bias test; and setting $\hat{w}_{t-1} = S_{t-1}^+ \hat{\varepsilon}_{t-1}$ gives the Positive Size Bias test.

6.1.2 Asymmetric GARCH Models

The leverage effect can be incorporated into a GARCH model in several ways. [71] proposed the following exponential GARCH (EGARCH) model to allow for leverage effects

$$h_t = a_0 + \sum_{i=1}^{p} a_i \frac{|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} b_j h_{t-j}, \tag{16}$$

where $h_t = \log \sigma_t^2$. Note that when ϵ_{t-i} is positive or there is "good news", the total effect of ϵ_{t-i} is $(1+\gamma_i)|\epsilon_{t-i}|$; in contrast, when ϵ_{t-i} is negative or there is "bad news", the total effect of ϵ_{t-i} is $(1-\gamma_i)|\epsilon_{t-i}|$. Bad news can have a larger impact on volatility, and the value of γ_i would be expected to be negative. An advantage of the EGARCH model over the basic GARCH model is that the conditional variance σ_t^2 is guaranteed to be positive regardless of the values of the coefficients in (16), because the logarithm of σ_t^2 instead of σ_t^2 itself is modeled. Also, the EGARCH is covariance stationary provided $\sum_{j=1}^q b_j < 1$.

Another GARCH variant that is capable of modeling leverage effects is the threshold GARCH (TGARCH) model,[9] which has the following form

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i \epsilon_{t-i}^2 + \sum_{i=1}^p \gamma_i S_{t-i} \epsilon_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2, \tag{17}$$

where

$$S_{t-i} = \begin{cases} 1 & \text{if} \quad \epsilon_{t-i} < 0 \\ 0 & \text{if} \quad \epsilon_{t-i} \geq 0 \end{cases} .$$

That is, depending on whether ϵ_{t-i} is above or below the threshold value of zero, ϵ_{t-i}^2 has different effects on the conditional variance σ_t^2: when ϵ_{t-i} is positive, the total effects are given by $a_i \epsilon_{t-i}^2$; when ϵ_{t-i} is negative, the total effects are given by $(a_i + \gamma_i)\epsilon_{t-i}^2$. So one would expect γ_i to be positive for bad news to have larger impacts.

[31] extended the basic GARCH model to allow for leverage effects. Their *power* GARCH (PGARCH(p, d, q)) model has the form

$$\sigma_t^d = a_0 + \sum_{i=1}^p a_i (|\epsilon_{t-i}| + \gamma_i \epsilon_{t-i})^d + \sum_{j=1}^q b_j \sigma_{t-j}^d, \tag{18}$$

where d is a positive exponent, and γ_i denotes the coefficient of leverage effects. When $d = 2$, (18) reduces to the basic GARCH model with leverage effects. When $d = 1$, the PGARCH model is specified in terms of σ_t which tends to be less sensitive to outliers than when $d = 2$. The exponent d may also be estimated as an additional parameter which increases the flexibility of the model. [31] showed that the PGARCH model also includes many other GARCH variants as special cases.

Many other asymmetric GARCH models have been proposed based on smooth transition and Markov switching models. See [44] and [83] for excellent surveys of these models.

6.1.3 News Impact Curve

The GARCH, EGARCH, TGARCH and PGARCH models are all capable of modeling leverage effects. To clearly see the impact of leverage effects in these models, [75], and [39] advocated the use of the so-called news impact curve. They defined the news

[9] The original TGARCH model proposed by [87] models σ_t instead of σ_t^2. The TGARCH model is also known as the GJR model because [47] proposed essentially the same model.

GARCH(1,1)	$\sigma_t^2 = A + a_1(\lvert\epsilon_{t-1}\rvert + \gamma_1\epsilon_{t-1})^2$ $A = a_0 + b_1\bar{\sigma}^2$ $\bar{\sigma}^2 = a_0/[1 - a_1(1+\gamma_1^2) - b_1]$
TGARCH(1,1)	$\sigma_t^2 = A + (a_1 + \gamma_1 S_{t-1})\epsilon_{t-1}^2$ $A = a_0 + b_1\bar{\sigma}^2$ $\bar{\sigma}^2 = a_0/[1 - (a_1 + \gamma_1/2) - b_1]$
PGARCH(1,1,1)	$\sigma_t^2 = A + 2\sqrt{A}a_1(\lvert\epsilon_{t-1}\rvert + \gamma_1\epsilon_{t-1})$ $+a_1^2(\lvert\epsilon_{t-1}\rvert + \gamma_1\epsilon_{t-1})^2,\ A = (a_0 + b_1\bar{\sigma})^2$ $\bar{\sigma}^2 = a_0^2/[1 - a_1/\sqrt{2/\pi} - b_1]^2$
EGARCH(1,1)	$\sigma_t^2 = A\exp\{a_1(\lvert\epsilon_{t-1}\rvert + \gamma_1\epsilon_{t-1})/\bar{\sigma}\}$ $A = \bar{\sigma}^{2b_1}\exp\{a_0\}$ $\bar{\sigma}^2 = \exp\{(a_0 + a_1\sqrt{2/\pi})/(1-b_1)\}$

Table 5: News impact curves for asymmetric GARCH processes. $\bar{\sigma}^2$ denotes the unconditional variance.

Asset	$\text{corr}(r_t^2, r_{t-1})$	Sign Bias	Negative Size Bias	Positive Size Bias
Microsoft	−0.0315	−0.4417 (0.6587)	−6.816 (0.000)	3.174 (0.001)
S&P 500	−0.098	2.457 (0.014)	−11.185 (0.000)	1.356 (0.175)

Notes: p-values are in parentheses.

Table 6: Tests for Asymmetric GARCH Effects.

impact curve as the functional relationship between conditional variance at time t and the shock term (error term) at time $t-1$, holding constant the information dated $t-2$ and earlier, and with all lagged conditional variance evaluated at the level of the unconditional variance. Table 5 summarizes the expressions defining the news impact curves, which include expressions for the unconditional variances, for the asymmetric GARCH(1,1) models.

6.1.4 Asymmetric GARCH Models for Daily Returns

Table 6 shows diagnostics and tests for asymmetric effects in the daily returns on Microsoft and the S&P 500. The correlation between r_t^2 and r_{t-1} is negative and fairly small for both series indicating weak evidence for asymmetry. However, the Size Bias tests clearly indicate asymmetric effects with the Negative Size Bias test giving the most significant results.

Table 7 gives the estimation results for EGARCH(1,1), TGARCH(1,1) and PGARCH(1,d,1) models for $d = 1, 2$. All of the asymmetric models show statistically significant lever-

Model	a_0	a_1	b_1	γ_1	BIC
			Microsoft		
EGARCH	-0.7273 [0.4064]	0.2144 [0.0594]	0.9247 [0.0489]	-0.2417 [0.0758]	-20265
TGARCH	$3.01e^{-5}$ [$1.02e^{-5}$]	0.0564 [0.0141]	0.8581 [0.0342]	0.0771 [0.0306]	-20291
PGARCH 2	$2.87e^{-5}$ [$9.27e^{-6}$]	0.0853 [0.0206]	0.8672 [0.0313]	-0.2164 [0.0579]	-20290
PGARCH 1	0.0010 [0.0006]	0.0921 [0.0236]	0.8876 [0.0401]	-0.2397 [0.0813]	-20268
			S&P 500		
EGARCH	-0.2602 [0.3699]	0.0720 [0.0397]	0.9781 [0.0389]	-0.3985 [0.4607]	-28051
TGARCH	$1.7e^{-6}$ [$7.93e^{-7}$]	0.0157 [0.0081]	0.9169 [0.0239]	0.1056 0.0357	-28200
PGARCH 2	$1.78e^{-6}$ [$8.74e^{-7}$]	0.0578 [0.0165]	0.9138 [0.0253]	-0.4783 [0.0910]	-28202
PGARCH 1	0.0002 [$2.56e^{-6}$]	0.0723 [0.0003]	0.9251 [$8.26e^{-6}$]	-0.7290 [0.0020]	-28253

Notes: QML standard errors are in brackets.

Table 7: Estimates of Asymmetric GARCH(1,1) Models.

age effects, and lower BIC values than the symmetric GARCH models. Model selection criteria indicate that the TGARCH(1,1) is the best fitting model for Microsoft, and the PGARCH(1,1,1) is the best fitting model for the S&P 500.

Figure 6 shows the estimated news impact curves based on these models. In this plot, the range of ϵ_t is determined by the residuals from the fitted models. The TGARCH and PGARCH(1,2,1) models have very similar NICs and show much larger responses to negative shocks than to positive shocks. Since the EGARCH(1,1) and PGARCH(1,1,1) models are more robust to extreme shocks, impacts of small (large) shocks for these model are larger (smaller) compared to those from the other models and the leverage effect is less pronounced.

6.2 Non-Gaussian Error Distributions

In all the examples illustrated so far, a normal error distribution has been exclusively used. However, given the well known fat tails in financial time series, it may be more appropriate to use a distribution which has fatter tails than the normal distribution. The most common fat-tailed error distributions for fitting GARCH models are: the Student's t distribution; the double exponential distribution; and the generalized error distribution.

[13] proposed fitting a GARCH model with a Student's t distribution for the standardized residual. If a random variable u_t has a Student's t distribution with ν degrees of freedom and a scale parameter s_t, the probability density function (pdf)

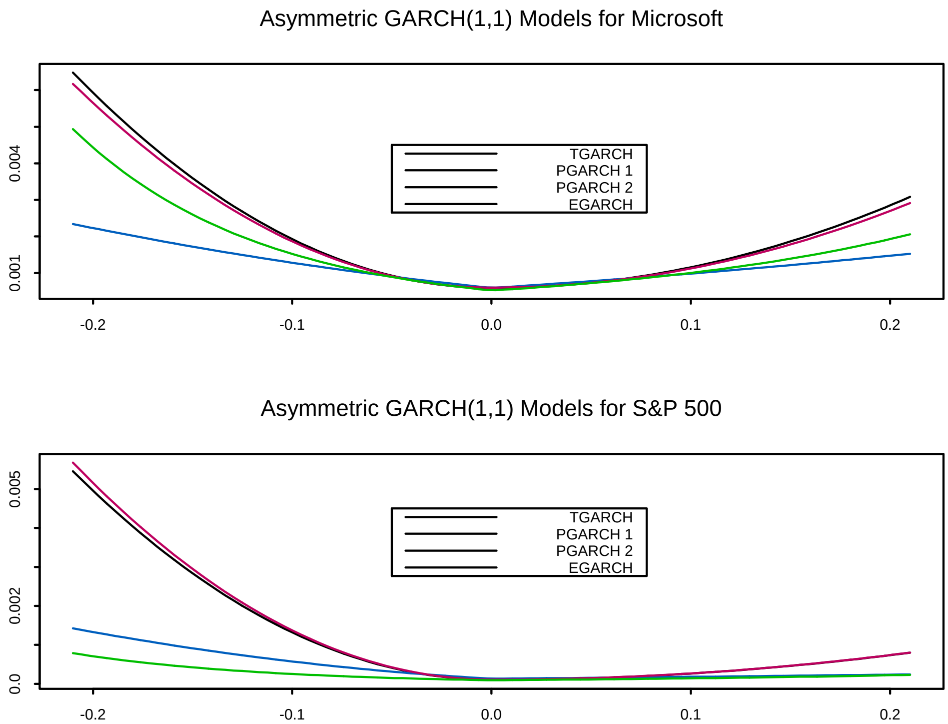

Figure 6: News impact curves from fitted asymmetric GARCH(1,1) models for Microsoft and S&P 500 index.

of u_t is given by

$$f(u_t) = \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} \frac{s_t^{-1/2}}{[1+u_t^2/(s_t\nu)]^{(\nu+1)/2}},$$

where $\Gamma(\cdot)$ is the gamma function. The variance of u_t is given by

$$\text{var}(u_t) = \frac{s_t\nu}{\nu-2},\ v > 2.$$

If the error term ϵ_t in a GARCH model follows a Student's t distribution with ν degrees of freedom and $\text{var}_{t-1}(\epsilon_t) = \sigma_t^2$, the scale parameter s_t should be chosen to be

$$s_t = \frac{\sigma_t^2(\nu-2)}{\nu}.$$

Thus the log-likelihood function of a GARCH model with Student's t distributed errors can be easily constructed based on the above pdf.

[71] proposed to use the generalized error distribution (GED) to capture the fat tails usually observed in the distribution of financial time series. If a random variable

u_t has a GED with mean zero and unit variance, the pdf of u_t is given by

$$f(u_t) = \frac{\nu \exp[-(1/2)|u_t/\lambda|^\nu]}{\lambda \cdot 2^{(\nu+1)/\nu}\Gamma(1/\nu)},$$

where

$$\lambda = \left[\frac{2^{-2/\nu}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2},$$

and ν is a positive parameter governing the thickness of the tail behavior of the distribution. When $\nu = 2$ the above pdf reduces to the standard normal pdf; when $\nu < 2$, the density has thicker tails than the normal density; when $\nu > 2$, the density has thinner tails than the normal density.

When the tail thickness parameter $\nu = 1$, the pdf of GED reduces to the pdf of double exponential distribution:

$$f(u_t) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|u_t|}.$$

Based on the above pdf, the log-likelihood function of GARCH models with GED or double exponential distributed errors can be easily constructed. See to [48] for an example.

Several other non-Gaussian error distribution have been proposed. [42] introduced the asymmetric Student's t distribution to capture both skewness and excess kurtosis in the standardized residuals. [85] proposed the normal inverse Gaussian distribution. [45] provided a very flexible seminonparametric innovation distribution based on a Hermite expansion of a Gaussian density. Their expansion is capable of capturing general shape departures from Gaussian behavior in the standardized residuals of the GARCH model.

6.2.1 Non-Gaussian GARCH Models for Daily Returns

Table 8 gives estimates of the GARCH(1,1) and best fitting asymmetric GARCH(1,1) models using Student's t innovations for the Microsoft and S&P 500 returns. Model selection criteria indicated that models using the Student's t distribution fit better than the models using the GED distribution. The estimated degrees of freedom for Microsoft is about 7, and for the S&P 500 about 6. The use of t-distributed errors clearly improves the fit of the GARCH(1,1) models. Indeed, the BIC values are even lower than the values for the asymmetric GARCH(1,1) models based on Gaussian errors (see Table 7). Overall, the asymmetric GARCH(1,1) models with t-distributed errors are the best fitting models. The qq-plots in Figure 7 shows that the Student's t distribution adequately captures the fat-tailed behavior in the standardized residuals for Microsoft but not for the S&P 500 index.

7 Long Memory GARCH Models

If returns follow a GARCH(p, q) model, then the autocorrelations of the squared and absolute returns should decay exponentially. However, the SACF of r_t^2 and $|r_t|$ for

Model	a_0	a_1	b_1	γ_1	v	BIC
			Microsoft			
GARCH	$3.39e^{-5}$	0.0939	0.8506		6.856	-20504
	$[1.52e^{-5}]$	[0.0241]	[0.0468]		[0.7121	
TGARCH	$3.44e^{-5}$	0.0613	0.8454	0.0769	7.070	-20511
	$[1.20e^{-5}]$	[0.0143]	[0.0380]	[0.0241]	[0.7023]	
			S&P 500			
GARCH	$5.41e^{-7}$	0.0540	0.0943		5.677	-28463
	$[2.15e^{-7}]$	[0.0095]	[0.0097]		[0.5571]	
PGARCH	0.0001	0.0624	0.9408	-0.7035	6.214	-28540
$d = 1$	[0.0002]	[0.0459]	[0.0564]	[0.0793]	[0.6369]	

Notes: QML standard errors are in brackets.

Table 8: Estimates of Non Gaussian GARCH(1,1) Models.

Microsoft and the S&P 500 in Figure 2 appear to decay much more slowly. This is evidence of so-called *long memory* behavior. Formally, a stationary process has long memory or long range dependence if its autocorrelation function behaves like

$$\rho(k) \to C_\rho k^{2d-1} \text{ as } k \to \infty,$$

where C_ρ is a positive constant, and d is a real number between 0 and $\frac{1}{2}$. Thus the autocorrelation function of a long memory process decays slowly at a hyperbolic rate. In fact, it decays so slowly that the autocorrelations are not summable:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty.$$

It is important to note that the scaling property of the autocorrelation function does not dictate the general behavior of the autocorrelation function. Instead, it only specifies the asymptotic behavior when $k \to \infty$. What this means is that for a long memory process, it is not necessary for the autocorrelation to remain significant at large lags as long as the autocorrelation function decays slowly. [8] gives an example to illustrate this property.

The following subSections describe testing for long memory and GARCH models that can capture long memory behavior in volatility. Explicit long memory GARCH models are discussed in [83].

7.1 Testing for Long Memory

One of the best-known and easiest to use tests for long memory or long range dependence is the rescaled range (R/S) statistic, which was originally proposed by [53], and later refined by [67] and his coauthors. The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. Specifically, consider a time series y_t, for $t = 1, \cdots, T$. The R/S statistic is defined

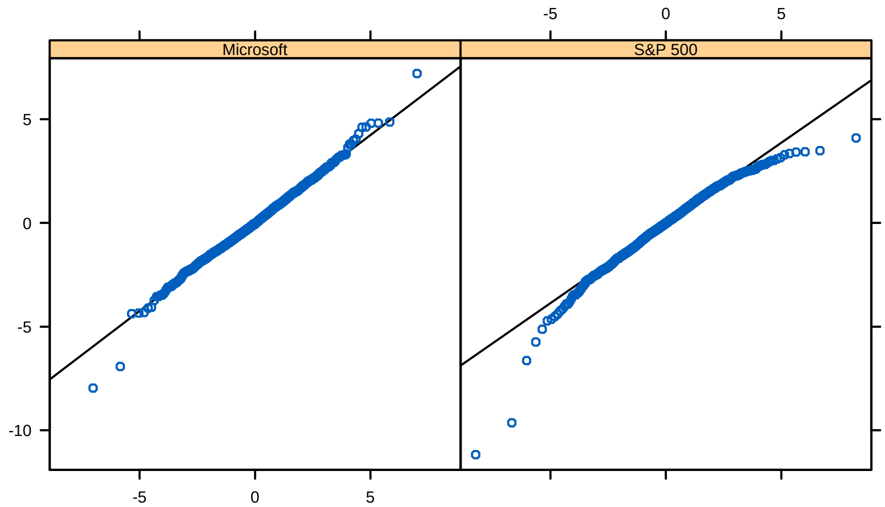

Figure 7: QQ-plots of Standardized Residuals from Asymmetric GARCH(1,1) models with Student's t errors.

as

$$Q_T = \frac{1}{s_T} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) \right], \tag{19}$$

where $\bar{y} = 1/T \sum_{i=1}^{T} y_i$ and $s_T = \sqrt{1/T \sum_{i=1}^{T} (y_i - \bar{y})^2}$. If y_t is iid with finite variance, then

$$\frac{1}{\sqrt{T}} Q_T \Rightarrow V,$$

where $\Rightarrow$ denotes weak convergence and V is the range of a Brownian bridge on the unit interval. [62] gives selected quantiles of V.

[62] pointed out that the R/S statistic is not robust to short range dependence. In particular, if y_t is autocorrelated (has short memory) then the limiting distribution of $Q_T/\sqrt{T}$ is V scaled by the square root of the long run variance of y_t. To allow for short range dependence in y_t, [62] modified the R/S statistic as follows

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) \right], \tag{20}$$

where the sample standard deviation is replaced by the square root of the Newey-West ([73]) estimate of the long run variance with bandwidth q.[10] [62] showed that if there is short memory but no long memory in y_t, $\tilde{Q}_T$ also converges to V, the range of a Brownian bridge. [18] found that (20) is effective for detecting long memory behavior in asset return volatility.

7.2 Two Component GARCH Model

In the covariance stationary GARCH model the conditional volatility will always mean revert to its long run level unconditional value. Recall the mean reverting form of the basic GARCH$(1,1)$ model in (11). In many empirical applications, the estimated mean reverting rate $\hat{a}_1 + \hat{b}_1$ is often very close to 1. For example, the estimated value of $a_1 + b_1$ from the GARCH(1,1) model for the S&P 500 index is 0.99 and the half life of a volatility shock implied by this mean reverting rate is $\ln(0.5)/\ln(0.956) = 76.5$ days. So the fitted GARCH(1,1) model implies that the conditional volatility is very persistent.

[37] suggested that the high persistence and long memory in volatility may be due to a time-varying long run volatility level. In particular, they suggested decomposing conditional variance into two components

$$\sigma_t^2 = q_t + s_t, \tag{21}$$

where q_t is a highly persistent long run component, and s_t is a transitory short run component. Long memory behavior can often be well approximated by a sum of two such components. A general form of the two components model that is based on a modified version of the PGARCH$(1, d, 1)$ is

$$\sigma_t^d = q_t^d + s_t^d, \tag{22}$$

$$q_t^d = \alpha_1|\epsilon_{t-1}|^d + \beta_1 q_{t-1}^d, \tag{23}$$

$$s_t^d = a_0 + \alpha_2|\epsilon_{t-1}|^d + \beta_2 s_{t-1}^d. \tag{24}$$

Here, the long run component q_t follows a highly persistent PGARCH$(1, d, 1)$ model and the transitory component s_t follows another PGARCH$(1, d, 1)$ model. For the two components to be separately identified the parameters should satisfy $1 < (\alpha_1 + \beta_1) < (\alpha_2 + \beta_2)$. It can be shown that the reduced form of the two components model is

$$\sigma_t^d = a_0 + (\alpha_1 + \alpha_2)|\epsilon_{t-1}|^d - (\alpha_1\beta_2 + \alpha_2\beta_1)|\epsilon_{t-2}|^d + (\beta_1 + \beta_2)\sigma_{t-1}^d - \beta_1\beta_2\sigma_{t-2}^d,$$

which is in the form of a constrained PGARCH$(2, d, 2)$ model. However, the two components model is not fully equivalent to the PGARCH$(2, d, 2)$ model because not all PGARCH$(2, d, 2)$ models have the component structure. Since the two components model is a constrained version of the PGARCH$(2, d, 2)$ model, the estimation of a two components model is often numerically more stable than the estimation of an unconstrained PGARCH$(2, d, 2)$ model.

[10] The long-run variance is the asymptotic variance of $\sqrt{T}(\bar{y} - \mu)$.

Asset	$\tilde{Q}_T$	
	r_t^2	$\|r_t\|$
Microsoft	2.3916	3.4557
S&P 500	2.3982	5.1232

Table 9: Modified R/S Tests for Long Memory.

a_0	α_1	β_1	α_2	β_2	v	BIC
			Microsoft			
$2.86e^{-6}$	0.0182	0.9494	0.0985	0.7025		-20262
$[1.65e^{-6}]$	[0.0102]	[0.0188]	[0.0344]	[0.2017]		
$1.75e^{-6}$	0.0121	0.9624	0.0963	0.7416	6.924	-20501
$5.11e^{-7}$	[0.0039]	[0.0098]	[0.0172]	[0.0526]	[0.6975]	
			S&P 500			
$3.2e^{-8}$	0.0059	0.9848	0.1014	0.8076		-28113
$[1.14e^{-8}]$	[0.0013]	[0.0000]	[0.0221]	[0.0001]		
$1.06e^{-8}$	0.0055	0.9846	0.0599	0.8987	5.787	-28457
$[1.26e^{-8}]$	[0.0060]	[0.0106]	[0.0109]	[0.0375]	[0.5329]	

Notes: QML standard errors are in brackets.

Table 10: Estimates of Two Component GARCH(1,1) Models.

7.3 Integrated GARCH Model

The high persistence often observed in fitted GARCH(1,1) models suggests that volatility might be nonstationary implying that $a_1 + b_1 = 1$, in which case the GARCH(1,1) model becomes the integrated GARCH(1,1) or IGARCH(1,1) model. In the IGARCH(1,1) model the unconditional variance is not finite and so the model does not exhibit volatility mean reversion. However, it can be shown that the model is strictly stationary provided $E[\ln(a_1 z_t^2 + b_1)] < 0$. If the IGARCH(1,1) model is strictly stationary then the parameters of the model can still be consistently estimated by MLE.

[27] argued against the IGARCH specification for modeling highly persistent volatility processes for two reasons. First, they argue that the observed convergence toward normality of aggregated returns is inconsistent with the IGARCH model. Second, they argue that observed IGARCH behavior may result from misspecification of the conditional variance function. For example, a two components structure or ignored structural breaks in the unconditional variance ([58] and [70]) can result in IGARCH behavior.

7.4 Long Memory GARCH Models for Daily Returns

Table 9 gives Lo's modified R/S statistic (20) applied to r_t^2 and $|r_t|$ for Microsoft and the S&P 500. The 1% right tailed critical value for the test is 2.098 ([62] Table 5.2) and so the modified R/S statistics are significant at the 1% level for both series providing evidence for long memory behavior in volatility.

Table 10 shows estimates of the two component GARCH(1,1) with $d = 2$, using Gaussian and Student's t errors, for the daily returns on Microsoft and the S&P 500. Notice that the BIC values are smaller than the BIC values for the unconstrained GARCH(2,2) models given in Table 3, which confirms the better numerical stability of the two component model. For both series, the two components are present and satisfy $1 < (\alpha_1 + \beta_1) < (\alpha_2 + \beta_2)$. For Microsoft, the half-lives of the two components from the Gaussian (Student's t) models are 21 (26.8) days and 3.1 (3.9) days, respectively. For the S&P 500, the half-lives of the two components from the Gaussian (Student's t) models are 75 (69.9) days and 7.3 (16.4) days, respectively.

8 GARCH Model Prediction

An important task of modeling conditional volatility is to generate accurate forecasts for both the future value of a financial time series as well as its conditional volatility. Volatility forecasts are used for risk management, option pricing, portfolio allocation, trading strategies and model evaluation. Since the conditional mean of the general GARCH model (10) assumes a traditional ARMA form, forecasts of future values of the underlying time series can be obtained following the traditional approach for ARMA prediction. However, by also allowing for a time varying conditional variance, GARCH models can generate accurate forecasts of future volatility, especially over short horizons. This Section illustrates how to forecast volatility using GARCH models.

8.1 GARCH and Forecasts for the Conditional Mean

Suppose one is interested in forecasting future values of y_T in the standard GARCH model described by (2), (3) and (6). For simplicity assume that $E_T[y_{T+1}] = c$. Then the minimum mean squared error $h-$ step ahead forecast of y_{T+h} is just c, which does not depend on the GARCH parameters, and the corresponding forecast error is

$$\epsilon_{T+h} = y_{T+h} - E_T[y_{T+h}].$$

The conditional variance of this forecast error is then

$$\mathrm{var}_T(\epsilon_{T+h}) = E_T[\sigma^2_{T+h}],$$

which does depend on the GARCH parameters. Therefore, in order to produce confidence bands for the $h-$step ahead forecast the $h-$step ahead volatility forecast $E_T[\sigma^2_{T+h}]$ is needed.

8.2 Forecasts from the GARCH(1,1) Model

For simplicity, consider the basic GARCH$(1,1)$ model (7) where $\epsilon_t = z_t\sigma_t$ such that $z_t \sim$ iid $(0,1)$ and has a symmetric distribution. Assume the model is to be estimated over the time period $t = 1, 2, \cdots, T$. The optimal, in terms of mean-squared error, forecast of σ^2_{T+k} given information at time T is $E_T[\sigma^2_{T+k}]$ and can be computed using a simple recursion. For $k = 1$,

$$\begin{aligned} E_T[\sigma^2_{T+1}] &= a_0 + a_1 E_T[\epsilon^2_T] + b_1 E_T[\sigma^2_T] \\ &= a_0 + a_1\epsilon^2_T + b_1\sigma^2_T, \end{aligned} \tag{25}$$

where it assumed that ϵ^2_T and σ^2_T are known[11]. Similarly, for $k = 2$

$$\begin{aligned} E_T[\sigma^2_{T+2}] &= a_0 + a_1 E_T[\epsilon^2_{T+1}] + b_1 E_T[\sigma^2_{T+1}] \\ &= a_0 + (a_1 + b_1)E_T[\sigma^2_{T+1}]. \end{aligned}$$

since $E_T[\epsilon^2_{T+1}] = E_T[z^2_{T+1}\sigma^2_{T+1}] = E_T[\sigma^2_{T+1}]$. In general, for $k \geq 2$

$$\begin{aligned} E_T[\sigma^2_{T+k}] &= a_0 + (a_1 + b_1)E_T[\sigma^2_{T+k-1}] \\ &= a_0 \sum_{i=0}^{k-1}(a_1 + b_1)^i + (a_1 + b_1)^{k-1}(a_1\epsilon^2_T + b_1\sigma^2_T). \end{aligned} \tag{26}$$

An alternative representation of the forecasting equation (26) starts with the mean-adjusted form

$$\sigma^2_{T+1} - \bar{\sigma}^2 = a_1(\epsilon^2_T - \bar{\sigma}^2) + b_1(\sigma^2_T - \bar{\sigma}^2),$$

where $\bar{\sigma}^2 = a_0/(1 - a_1 - b_1)$ is the unconditional variance. Then by recursive substitution

$$E_T[\sigma^2_{T+k}] - \bar{\sigma}^2 = (a_1 + b_1)^{k-1}(E[\sigma^2_{T+1}] - \bar{\sigma}^2). \tag{27}$$

Notice that as $k \to \infty$, the volatility forecast in (26) approaches $\bar{\sigma}^2$ if the GARCH process is covariance stationary and the speed at which the forecasts approaches $\bar{\sigma}^2$ is captured by $a_1 + b_1$.

The forecasting algorithm (26) produces forecasts for the conditional variance σ^2_{T+k}. The forecast for the conditional volatility, σ_{T+k}, is usually defined as the square root of the forecast for σ^2_{T+k}.

The GARCH(1,1) forecasting algorithm (25) is closely related to an exponentially weighted moving average (EWMA) of past values of ϵ^2_t. This type of forecast is commonly used by RiskMetrics ([54]). The EWMA forecast of σ^2_{T+1} has the form

$$\sigma^2_{T+1,EWMA} = (1 - \lambda)\sum_{s=0}^{\infty}\lambda^s\epsilon^2_{t-s} \tag{28}$$

for $\lambda \in (0, 1)$. In (28), the weights sum to one, the first weight is $1-\lambda$, and the remaining weights decline exponentially. To relate the EWMA forecast to the GARCH(1,1) formula (25), (28) may be re-expressed as

$$\sigma^2_{T+1,EWMA} = (1 - \lambda)\epsilon^2_T + \lambda\sigma^2_{T,EWMA} = \epsilon^2_T + \lambda(\sigma^2_{T,EWMA} - \epsilon^2_T),$$

[11] In practice, a_0, a_1, b_1, ϵ_T and σ^2_T are the fitted values computed from the estimated GARCH(1,1) model instead of the unobserved "true" values.

which is of the form (25) with $a_0 = 0$, $a_1 = 1 - \lambda$ and $b_1 = \lambda$. Therefore, the EWMA forecast is equivalent to the forecast from a restricted IGARCH(1,1) model. It follows that for any $h > 0$, $\sigma^2_{T+h,EWMA} = \sigma^2_{T,EWMA}$. As a result, unlike the GARCH(1,1) forecast, the EWMA forecast does not exhibit mean reversion to a long-run unconditional variance.

8.3 Forecasts from Asymmetric GARCH(1,1) Models

To illustrate the asymmetric effects of leverage on forecasting, consider the TGARCH(1,1) model (17) at time T

$$\sigma_T^2 = a_0 + a_1\epsilon_{T-1}^2 + \gamma_1 S_{T-1}\epsilon_{T-1}^2 + b_1\sigma_{T-1}^2.$$

Assume that ϵ_t has a symmetric distribution about zero. The forecast for $T+1$ based on information at time T is

$$E_T[\sigma_{T+1}^2] = a_0 + a_1\epsilon_T^2 + \gamma_1 S_T\epsilon_T^2 + b_1\sigma_T^2,$$

where it assumed that ϵ_T^2, S_T and σ_T^2 are known. Hence, the TGARCH(1,1) forecast for $T+1$ will be different than the GARCH(1,1) forecast if $S_T = 1$ ($\epsilon_T < 0$). The forecast at $T+2$ is

$$\begin{aligned} E_T[\sigma_{T+2}^2] &= a_0 + a_1 E_T[\epsilon_{T+1}^2] + \gamma_1 E_T[S_{T+1}\epsilon_{T+1}^2] + b_1 E_T[\sigma_{T+1}^2] \\ &= a_0 + \left(\frac{\gamma_1}{2} + a_1 + b_1\right) E_T[\sigma_{T+1}^2], \end{aligned}$$

which follows since $E_T[S_{T+1}\epsilon_{T+1}^2] = E_T[S_{T+1}]E_T[\epsilon_{T+1}^2] = \frac{1}{2}E_T[\sigma_{T+1}^2]$. Notice that the asymmetric impact of leverage is present even if $S_T = 0$. By recursive substitution for the forecast at $T+h$ is

$$E_T[\sigma_{T+h}^2] = a_0 + \left(\frac{\gamma_1}{2} + a_1 + b_1\right)^{h-1} E_T[\sigma_{T+1}^2], \qquad (29)$$

which is similar to the GARCH(1,1) forecast (26). The mean reverting form (29) is

$$E_T[\sigma_{T+h}^2] - \bar{\sigma}^2 = \left(\frac{\gamma_1}{2} + a_1 + b_1\right)^{h-1} \left(E_T[\sigma_{T+h}^2] - \bar{\sigma}^2\right)$$

where $\bar{\sigma}^2 = a_0/(1 - \frac{\gamma_1}{2} - a_1 - b_1)$ is the long run variance.

Forecasting algorithms for σ_{T+h}^d in the PGARCH$(1, d, 1)$ and for $\ln \sigma_{T+h}^2$ in the EGARCH(1,1) follow in a similar manner and the reader is referred to [31], and [71] for further details.

8.4 Simulation-Based Forecasts

The forecasted volatility can be used together with forecasted series values to generate confidence intervals of the forecasted series values. In many cases, the forecasted volatility is of central interest, and confidence intervals for the forecasted volatility can be obtained as well. However, analytic formulas for confidence intervals of forecasted volatility are only known for some special cases (see [6]). In models for which analytic formulas for confidence intervals are not known, a simulation-based method can be used to obtain confidence intervals for forecasted volatility from any GARCH that can be simulated. To obtain volatility forecasts from a fitted GARCH model, simply simulate σ^2_{T+k} from the last observation of the fitted model. This process can be repeated many times to obtain an "ensemble" of volatility forecasts. The point forecast of σ^2_{T+k} may then be computed by averaging over the simulations, and a 95% confidence interval may be computed using the 2.5% and 97.5% quantiles of the simulation distribution, respectively.

8.5 Forecasting the Volatility of Multiperiod Returns

In many situations, a GARCH model is fit to daily continuously compounded returns $r_t = \ln(P_t) - \ln(P_{t-1})$, where P_t denotes the closing price on day t. The resulting GARCH forecasts are for daily volatility at different horizons. For risk management and option pricing with stochastic volatility, volatility forecasts are needed for multi-period returns. With continuously compounded returns, the $h-$day return between days T and $T+h$ is simply the sum of h single day returns

$$r_{T+h}(h) = \sum_{j=1}^{h} r_{T+j}.$$

Assuming returns are uncorrelated, the conditional variance of the $h-$period return is then

$$\mathrm{var}_T(r_{T+h}(h)) = \sigma^2_T(h) = \sum_{j=1}^{h} \mathrm{var}_T(r_{T+j}) = E_T[\sigma^2_{T+1}] + \cdots + E_T[\sigma^2_{T+h}]. \qquad (30)$$

If returns have constant variance $\bar{\sigma}^2$, then $\sigma^2_T(h) = h\bar{\sigma}^2$ and $\sigma_T(h) = \sqrt{h}\bar{\sigma}$. This is known as the "square root of time" rule as the $h-$day volatility scales with $\sqrt{h}$. In this case, the $h-$day variance per day, $\sigma^2_T(h)/h$, is constant. If returns are described by a GARCH model then the square root of time rule does not necessarily apply. To see this, suppose returns follow a GARCH(1,1) model. Plugging the GARCH(1,1) model forecasts (27) for $E_T[\sigma^2_{T+1}], \ldots, E_T[\sigma^2_{T+h}]$ into (30) gives

$$\sigma^2_T(h) = h\bar{\sigma}^2 + (E[\sigma^2_{T+1}] - \bar{\sigma}^2)\left[\frac{1-(a_1+b_1)^h}{1-(a_1+b_1)}\right]$$

For the GARCH(1,1) process the square root of time rule only holds if $E[\sigma^2_{T+1}] = \bar{\sigma}^2$. Whether $\sigma^2_T(h)$ is larger or smaller than $h\bar{\sigma}^2$ depends on whether $E[\sigma^2_{T+1}]$ is larger or smaller than $\bar{\sigma}^2$.

8.6 Evaluating Volatility Predictions

GARCH models are often judged by their out-of-sample forecasting ability, see [22] for an overview. This forecasting ability can be measured using traditional forecast error metrics as well as with specific economic considerations such as value-at-risk violations, option pricing accuracy, or portfolio performance. Out-of-sample forecasts for use in model comparison are typically computed using one of two methods. The first method produces recursive forecasts. An initial sample using data from $t = 1, \ldots, T$ is used to estimate the models, and $h-$step ahead out-of-sample forecasts are produced starting at time T. Then the sample is increased by one, the models are re-estimated, and $h-$step ahead forecasts are produced starting at $T+1$. This process is repeated until no more $h-$step ahead forecasts can be computed. The second method produces rolling forecasts. An initial sample using data from $t = 1, \ldots, T$ is used to determine a window width T, to estimate the models, and to form $h-$step ahead out-of-sample forecasts starting at time T. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2, \ldots, T+1$, and $h-$step ahead out-of-sample forecasts are produced starting at time $T+1$. This process is repeated until no more $h-$step ahead forecasts can be computed.

8.6.1 Traditional Forecast Evaluation Statistics

Let $E_{i,T}[\sigma^2_{T+h}]$ denote the $h-$step ahead forecast of σ^2_{T+h} at time T from GARCH model i using either recursive or rolling methods. Define the corresponding forecast error as $e_{i,T+h|T} = E_{i,T}[\sigma^2_{T+h}] - \sigma^2_{T+h}$. Common forecast evaluation statistics based on N out-of-sample forecasts from $T = T+1, \ldots, T+N$ are

$$\mathrm{MSE}_i = \frac{1}{N}\sum_{j=T+1}^{T+N} e^2_{i,j+h|j},$$

$$\mathrm{MAE}_i = \frac{1}{N}\sum_{j=T+1}^{T+N} \left| e_{i,j+h|j} \right|,$$

$$\mathrm{MAPE}_i = \frac{1}{N}\sum_{j=T+1}^{T+N} \frac{\left| e_{i,j+h|j} \right|}{\sigma_{j+h}}.$$

The model which produces the smallest values of the forecast evaluation statistics is judged to be the best model. Of course, the forecast evaluation statistics are random variables and a formal statistical procedure should be used to determine if one model exhibits superior predictive performance.

[28] proposed a simple procedure to test the null hypothesis that one model has superior predictive performance over another model based on traditional forecast evaluation statistics. Let $\{e_{1,j+h|j}\}_{T+1}^{T+N}$, and $\{e_{2,j+h|j}\}_{T+1}^{T+N}$ denote forecast errors from two different GARCH models. The accuracy of each forecast is measured by a particular loss function $L(e_{i,T+h|T})$, $i = 1, 2$. Common choices are the squared error loss function $L(e_{i,T+h|T}) = \left(e_{i,T+h|T}\right)^2$ and the absolute error loss function $L(e_{i,T+h|T}) = \left|e_{i,T+h|T}\right|$. The Diebold-Mariano (DM) test is based on the loss differential

$$d_{T+h} = L(e_{1,T+h|T}) - L(e_{2,T+h|T}).$$

The null of equal predictive accuracy is $H_0 : E[d_{T+h}] = 0$.The DM test statistic is

$$\text{S} = \frac{\bar{d}}{\left(\widehat{\text{avar}}(\bar{d})\right)^{1/2}}, \tag{31}$$

where $\bar{d} = N^{-1} \sum_{j=T+1}^{T+N} d_{j+h}$, and $\widehat{\text{avar}}(\bar{d})$ is a consistent estimate of the asymptotic variance of $\sqrt{N}\bar{d}$. [28] recommend using the Newey-West estimate for $\widehat{\text{avar}}(\bar{d})$ because the sample of loss differentials $\{d_{j+h}\}_{T+1}^{T+N}$ are serially correlated for $h > 1$. Under the null of equal predictive accuracy, S has an asymptotic standard normal distribution. Hence, the DM statistic can be used to test if a given forecast evaluation statistic (e.g. MSE_1) for one model is statistically different from the forecast evaluation statistic for another model (e.g. MSE_2).

Forecasts are also often judged using the forecasting regression

$$\sigma_{T+h}^2 = \alpha + \beta E_{i,T}[\sigma_{T+h}^2] + e_{i,T+h}. \tag{32}$$

Unbiased forecasts have $\alpha = 0$ and $\beta = 1$, and accurate forecasts have high regression R^2 values. In practice, the forecasting regression suffers from an errors-in-variables problem when estimated GARCH parameters are used to form $E_{i,T}[\sigma_{T+h}^2]$ and this creates a downward bias in the estimate of β. As a result, attention is more often focused on the R^2 from (32).

An important practical problem with applying forecast evaluations to volatility models is that the $h-$step ahead volatility σ_{T+h}^2 is not directly observable. Typically, ϵ_{T+h}^2 (or just the squared return) is used to proxy σ_{T+h}^2 since $E_T[\epsilon_{T+h}^2] = E_T[z_{T+h}^2 \sigma_{T+h}^2] = E_T[\sigma_{T+h}^2]$. However, ϵ_{T+h}^2 is a very noisy proxy for σ_{T+h}^2 since $\text{var}(\epsilon_{T+h}^2) = E[\sigma_{T+h}^4](\kappa - 1)$, where κ is the fourth moment of z_t, and this causes problems for the interpretation of the forecast evaluation metrics.

Many empirical papers have evaluated the forecasting accuracy of competing GARCH models using ϵ_{T+h}^2 as a proxy for σ_{T+h}^2. [77] gave a comprehensive survey. The typical findings are that the forecasting evaluation statistics tend to be large, the forecasting regressions tend to be slightly biased, and the regression R^2 values tend to be very low (typically below 0.1). In general, asymmetric GARCH models tend to have the lowest forecast evaluation statistics. The overall conclusion, however, is that GARCH models do not forecast very well.

[2] provided an explanation for the apparent poor forecasting performance of GARCH models when ϵ_{T+h}^2 is used as a proxy for σ_{T+h}^2 in (32). For the GARCH(1,1) model in which z_t has finite kurtosis κ, they showed that the population R^2 value in (32) with $h = 1$ is equal to

$$R^2 = \frac{a_1^2}{1 - b_1^2 - 2a_1 b_1},$$

and is bounded from above by $1/\kappa$. Assuming $z_t \sim N(0, 1)$, this upper bound is 1/3. With a fat-tailed distribution for z_t the upper bound is smaller. Hence, very low R^2 values are to be expected even if the true model is a GARCH(1,1). Moreover, [49] found that the substitution of ϵ_{T+h}^2 for σ_{T+h}^2 in the evaluation of GARCH models using the DM statistic (31) can result in inferior models being chosen as the best

	Error pdf	GARCH	TGARCH	PGARCH
MSFT	Gaussian	0.0253	0.0257	0.0256
	Student's t	0.0247	0.0253	0.0250
S&P 500	Gaussian	0.0138	0.0122	0.0108
	Student's t	0.0138	0.0128	0.0111

Table 11: Unconditional Volatilities from Estimated GARCH(1,1) Models.

with probability one. These results indicate that extreme care must be used when interpreting forecast evaluation statistics and tests based on ϵ^2_{T+h}.

If high frequency intraday data are available, then instead of using ϵ^2_{T+h} to proxy σ^2_{T+h} [2] suggested using the so-called realized variance

$$RV^m_{t+h} = \sum_{j=1}^{m} r^2_{t+h,j},$$

where $\{r_{T+h,1}, \ldots, r_{T+h,m}\}$ denote the squared intraday returns at sampling frequency $1/m$ for day $T+h$. For example, if prices are sampled every 5 minutes and trading takes place 24 hours per day then there are $m = 288$ 5-minute intervals per trading day. Under certain conditions (see [4]), RV^m_{t+h} is a consistent estimate of σ^2_{T+h} as $m \to \infty$. As a result, RV^m_{t+h} is a much less noisy estimate of σ^2_{T+h} than ϵ^2_{T+h} and so forecast evaluations based on RV^m_{t+h} are expected to be much more accurate than those based on ϵ^2_{T+h}. For example, in evaluating GARCH(1,1) forecasts for the Deutschemark-US daily exchange rate, [2] reported R^2 values from (32) of $0.047, 0.331$ and 0.479 using ϵ^2_{T+1}, RV^{24}_{T+1} and RV^{288}_{T+1}, respectively.

8.7 Forecasting the Volatility of Microsoft and the S&P 500

Figure 8 shows $h-$day ahead volatility predictions ($h = 1, \ldots, 250$) from the fitted GARCH(1,1) models with normal errors for the daily returns on Microsoft and the S&P 500. The horizontal line in the figures represents the estimated unconditional standard deviation from the fitted models. At the beginning of the forecast period, $\hat{\sigma}_T < \widehat{\bar{\sigma}}$ for both series and so the forecasts revert upward toward the unconditional volatility. The speed of volatility mean reversion is clearly shown by the forecast profiles. The forecasts for Microsoft revert to the unconditional level after about four months, whereas the forecasts for the S&P 500 take over one year.

Figure 8 shows the volatility forecasts from the asymmetric and long memory GARCH(1,1) models, and Table 11 gives the unconditional volatility from the estimated models. For Microsoft, the forecasts and unconditional volatilities from the different models are similar. For the S&P 500, in contrast, the forecasts and unconditional volatilities differ considerably across the models.

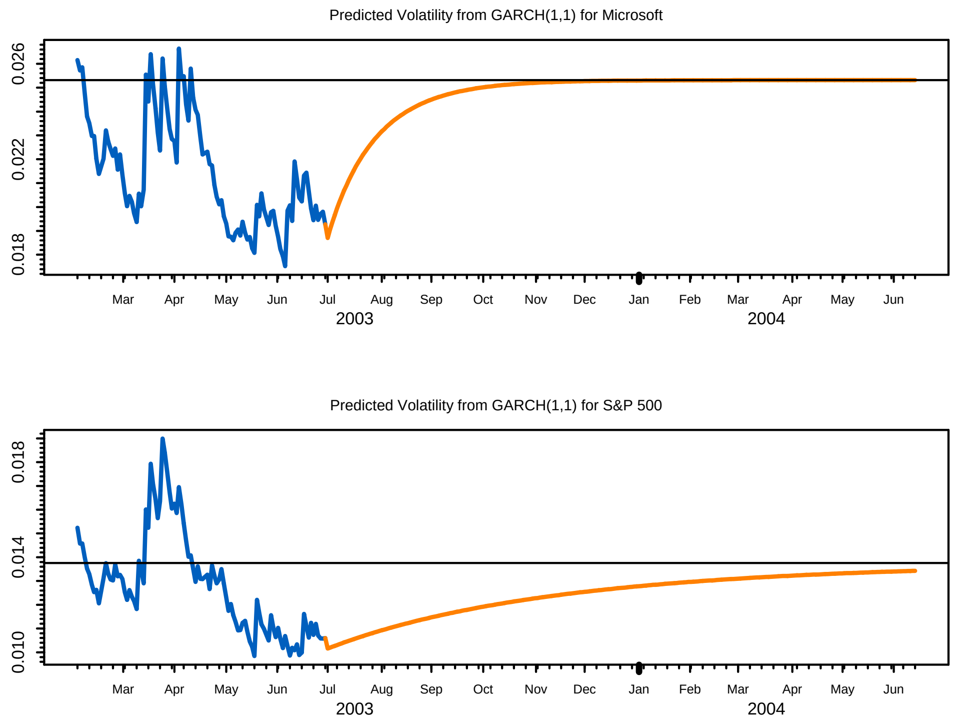

Figure 8: Predicted Volatility from GARCH(1,1) Models

9 Final Remarks

This paper surveyed some of the practical issues associated with estimating univariate GARCH models and forecasting volatility. Some practical issues associated with the estimation of multivariate GARCH models and forecasting of conditional covariances are given in [80].

References

[1] Alexander, C. (2001). *Market Models: A Guide to Financial Data Analysis*. John Wiley & Sons, Chichester, UK.

[2] Andersen, T., and Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39(4), 885-905.

[3] Andersen, T., and Bollerslev, T., Christoffersen, P.K., and Diebold, F.X. (2006). Volatility Forecasting. In G. Elliott, C.W.J.Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, Amsterdam, North Holland.

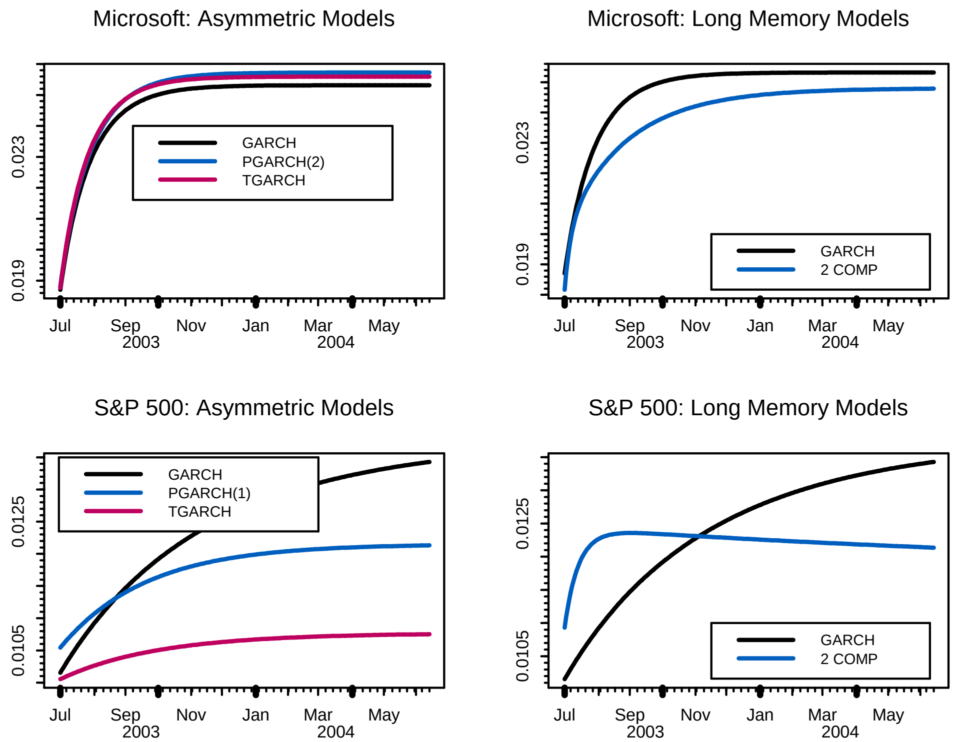

Figure 9: Predicted Volatility from Competing GARCH Models.

[4] Andersen, T, Bollerslev, T., Diebold, F.X., and Labys, P. (2003). The Distribution of Exchange Rate Volatility. *Journal of the American Statistical Association*, 96, 42-55.

[5] Andreou, E. and Ghysels, E. (2002). Rolling Volatility Estimators: Some new Theoretical, Simulation and Empirical Results. *Journal of Business and Economic Statistics*, 20(3), 363-376.

[6] Baillie, R.T., and Bollerslev, T. (1992). Prediction in Dynamic Models with Time Dependent Conditional Variances. *Journal of Econometrics*, 52, 91-113.

[7] Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 74, 3-30.

[8] Beran, J. (1994). *Statistics for Long Memory Processes*. Chapman and Hall, New York.

[9] Bera, A.K. and Higgins, M.L. (1995). On ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, 7, 305-362.

[10] Black, F. (1976). Studies in Stock Price Volatility Changes. *Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section*, American Statistical Association, 177-181.

[11] Blair, B., Poon, S.-H., and Taylor, S.J. (2001). Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns. *Journal of Econometrics*, 105, 5-26.

[12] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

[13] Bollerslev, T. (1987). A Conditional Heteroskedastic Time Series Model for Speculative Prices Rates of Return. *Review of Economics and Statistics*, 69, 542-547.

[14] Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). ARCH Models. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics*, Vol. 4, Elsevier Science B. V., Amsterdam.

[15] Bollerslev, T., and Ghysels, E. (1996). Periodic Autoregressive Conditional Heteroscedasticity. *Journal of Business and Economic Statistics*, 14, 139-151

[16] Bollerslev, T., and Wooldridge, T. (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances. *Econometric Reviews*, 11, 143-172.

[17] Bomfim, A.N. (2003). Pre-Announcement Effects, News Effects, and Volatility: Monetary Policy and the Stock Market. *Journal of Banking and Finance*, 27, 133-151.

[18] Breidt, F.J., Crato, N. and de Lima, P. (1998). The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics*, 73, 325-348.

[19] Brooks, C. (1997). GARCH Modeling in Finance: A Review of the Software Options. *Economic Journal*, 107(443), 1271-1276.

[20] Brooks, C., Burke, S., and Pesand, G. (2001). Benchmarks and the Accuracy of GARCH Model Estimation. *International Journal of Forecasting*, 17, 45-56.

[21] Chen, F., Yu, W.-C., and Zivot, E. (2008). Predicting Stock Volatility Using After-Hours Information. Unpublished manuscript, Department of Economics, University of Washington.

[22] Clements, M.P. (2005). *Evaluating Econometric Forecasts of Economic and Financial Variables.* Palgrave Texts in Econometrics, Palgrave Macmillan, Houndmills, UK.

[23] Conrad, C., and Haag, B.R. (2006). Inequality Constraints in the Fractionally Integrated GARCH Model. *Journal of Financial Econometrics*, 4(3), 413-449.

[24] Davis, R., and T. Mikosch (2008). Extreme Value Theory for GARCH Processes. In T.G. Andersen, R.A. Davis, J-P Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*, New York, Springer.

[25] de Lima, P.J.F. (1996). Nuisance Parameter Free Properties of Correlation Integral Based Statistics. *Econometric Reviews*, 15, 237-259.

[26] Diebold, F.X. (1988). *Empirical Modeling of Exchange Rate Behavior*. Springer-Verlag, New York.

[27] Diebold, F.X. and Lopez, J.A. (1996). Modeling Volatility Dynamics. In K. Hoover, editor, *Macroeconomics: Developments, Tensions and Prospects*. Kluwer, Boston.

[28] Diebold, F.X. and R.S. Mariano (1995). Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*. 13, 253-263.

[29] Diebold, F.X., and Schuermann, T. (1993). Exact Maximum Likelihood Estimation of ARCH Models. Unpublished manuscript, Department of Economics, University of Pennsylvania.

[30] Ding, Z., and Granger, C. W. J. (1996). Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics*, 73, 185-215.

[31] Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1, 83-106.

[32] Drost, F.C. and Nijman, T.E. (1993). Temporal Aggregation of GARCH Processes. *Econometrica*, 61, 909-927.

[33] Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.

[34] Engle, R.F. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Economics. *Journal of Economic Perspectives*, 15, 157-168.

[35] Engle, R.F. and González-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics*, 9, 345-60.

[36] Engle, R.F. and Kroner, K. (1995). Multivariate Simultaneous Generalised ARCH. *Econometric Theory* 11, 122-150.

[37] Engle, R. F., and Lee, G. J. (1999). A Long-Run and Short-Run Component Model of Stock Return Volatility. In R. F. Engle and H. White, editors, *Cointegration, Causality, and Forecasting*. Oxford University Press, Oxford.

[38] Engle, R.F., and Mezrich, J. (1995). Grappling with GARCH. *RISK*, 8(9), 112-117.

[39] Engle, R.F., and Ng, V. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, 48, 1749-78.

[40] Engle, R.F., and Patton, A. (2001). What Good is a Volatility Model?, *Quantitative Finance*, 1, 237-245.

[41] Fiorentini, G., Calzolari, G., and Panattoni, L. (1996) Analytic Derivatives and the Computation of GARCH Estimates. *Journal of Applied Econometrics* 11, 399-417.

[42] Fernandez, C., and Steel, M. (1998). On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93, 359-371.

[43] Flannery, M. and Protopapadakis, A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *The Review of Financial Studies*, 15, 751-782.

[44] Franses, P.H. and D. van Dijk (2000). *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press, Cambridge.

[45] Gallant, A.R. and Tauchen, G. (2001). SNP: A Program for Nonparametric Time Series Analysis, Version 8.8, User's Guide. Unpublished manuscrip, University of North Carolina at Chapel Hill.

[46] Gallo, G.M., and Pacini, B. (1997). Early News is Good News: The Effect of Market Opening on Market Volatility. *Studies in Nonlinear Dynamics and Econometrics*, 2(4), 115-131.

[47] Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48(5), 1779-1801.

[48] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton, NJ.

[49] Hansen, P., and Lunde, A. (2004). A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1) Model? *Journal of Applied Econometrics*, 20, 873-889.

[50] Hansen, P., and Lunde, A. (2006). Consistent Ranking of Volatility Models. *Journal of Econometrics*, 131, 97-121.

[51] He, C., Teräsvirta, T. (1999a). Properties of Moments of a Family of GARCH Processes. *Journal of Econometrics*, 92, 173-192.

[52] He, C., Teräsvirta, T. (1999b). Fourth Moment Structure of the GARCH(p, q) Process. *Econometric Theory*, 15, 824-846.

[53] Hurst, H. E. (1951). Long Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, 116, 770-799.

[54] J.P. Morgan (1997). *RiskMetrics, Technical Documents, 4th Edition*. New York.

[55] Jensen, T., and Rahbek, A. (2004). Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case. *Econometrica*, 72(2), 641-646.

[56] Kristensen, D., and Rahbek, A. (2005). Asymptotics of the QMLE for a Class of ARCH(q) Models. *Econometric Theory*, 21, 946-961.

[57] Laurent, S., and Peters, J.-P. (2002). G@RCH 2.0: An Ox Package for Estimating and Forecasting Various ARCH Models. *Journal of Economic Surveys* 16, pp.447-485.

[58] Lamoureux, C.G., and Lastrapes, W.D. (1990a). Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. *The Journal of Finance*, 45, 221-229.

[59] Lamoureux, C.G., and Lastrapes, W.D. (1990b). Persistence in Variance, Structural Change and the GARCH Model. *Journal of Business and Economic Statistics*, 8, 225-234.

[60] Lee, and Hansen, B.E. (1993). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory*, 10, 29-52.

[61] Lee, J.H.H., and King, M.L. (1993). A Locally Most Powerful Based Score Test for ARCH and GARCH Regression Disturbances. *Journal of Business and Economic Statistics*, 7, 259-279.

[62] Lo, A. W. (1991). Long Term Memory in Stock Market Prices. *Econometrica*, 59, 1279-1313.

[63] Lumsdaine, R. L, (1992). Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica*, 64, 575-596.

[64] Lumsdaine, R.L., and Ng, S. (1999). Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean. *Journal of Econometrics*, 93, 257-279.

[65] Lundbergh, S., and Teräsvirta, T. (2002). Evaluating GARCH Models. *Journal of Econometrics*, Vol. 110, 417-435

[66] Ma, J., Nelson, C.R., and Startz, R. (2006). Spurious Inference in the GARCH(1,1) Model When It Is Weakly Identified. *Studies in Nonlinear Dynamics and Econometrics*, 11(1), Article 1.

[67] Mandelbrot, B. B. (1975). Limit Theorems on the Self-Normalized Range for Weakly and Strongly Dependent Processes. *Zeitschrift für Wahr- scheinlichkeitstheorie und verwandte Gebiete*, 31, 271-285.

[68] Martens, M. (2002). Measuring and Forecasting S&P 500 Index-Futures Volatility Using High Frequency Data. *The Journal of Futures Markets*, 22, 497-518.

[69] McCullough, B.D., and Renfro, C.G. (1999). Benchmarks and Software Standards: A Case Study of GARCH Procedures. *Journal of Economic and Social Measurement*, 25, 59-71.

[70] Mikosch, T., and Starica, C. (2004). Non-stationarities in Financial Time Series, the Long-Range Dependence and the IGARCH Effects. *Review of Economics and Statistics*, 86, 378-384.

[71] Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: a New Approach. *Econometrica*, 59(2), 347-370.

[72] Nelson, D. B., and Cao, C. Q. (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business and Economic Statistics*, 10(2), 229-235.

[73] Newey, W.K. and West, K.D. (1987). A Simple Positive Semidefinite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703-708.

[74] Pagan, A. (1996). The Econometrics of Financial Markets. *Journal of Empirical Finance*, 3, 15-102.

[75] Pagan, A., and Schwert G.W. (1990). Alternative Models for Conditional Volatility. *Journal of Econometrics*, 45, 267-290.

[76] Palm, F.C. (1996). GARCH models of Volatility. In G.S. Maddala and C.R. Rao, editors, *Handbook of Statistics* (Vol 14), *Statistical Methods in Finance*, pp. 209-240. North Holland.

[77] Poon, S.-H. (2005). *A Practical Guide to Forecasting Financial Market Volatility.* John Wiley & Sons, New York.

[78] Poon, S.-H., and Granger, C. (2003). Forecasting Financial Market Volatility: A Review. *Journal of Economic Literature*, 41(2), 478-539.

[79] Poon, S.-H., and Granger, C. (2005). Practical Issues in Forecasting Volatility. *Financial Analysts Journal*, 61(1), 45-56.

[80] Silvennoinen, A., and Teräsvirta, T. (2008). Multivariate GARCH. In T.G. Andersen, R.A. Davis, J-P Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*, New York, Springer.

[81] Straumann, D. (2005). *Estimation in Conditionally Heteroskedastic Time Series Models.* Lecture Notes in Statistics 181, Springer, Berlin.

[82] Taylor, S.J., and Xu, X. (1997). The Incremental Volatilty Information in One Million Foreign Exchange Quotations. *Journal of Empirical Finance*, 4, 317-340.

[83] Teräsvirta, T. (2008). An Introduction to Univariate GARCH Models. In T.G. Andersen, R.A. Davis, J-P Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*, New York, Springer.

[84] Tsay, R.S. (2001). *Analysis of Financial Time Series*. John Wiley & Sons, New York.

[85] Venter, J.H., and de Jongh, P.J. (2002). Risk Estimation Using the Normal Inverse Gausian Distribution. *Journal of Risk*, 4, 1-23.

[86] Weiss, A.A. (1986). Asymptotic Theory for ARCH Models: Estimation and Testing. *Econometric Theory*, 2, 107-131.

[87] Zakoian, M. (1994). Threshold Heteroscedastic Models. *Journal of Economic Dynamics and Control*, 18, 931-955.

[88] Zivot, E., and Wang, J. (2005). *Modeling Financial Time Series with S-PLUS, Second Edition*. Springer-Verlag, New York.