Statistics 435/535/711 20 April 2021
P. Shaman

Combination of Forecasts

Often one has available several different forecasts, generated by different methods, for the same setting. For example, one may have tried regression modelling and ARIMA analysis on the same data. And there may be alternative models within these three methodologies. Combination of the competing forecasts is a good option in this situation.

Let $f_{1t}, f_{2t}, \ldots, f_{kt}$ denote k alternative forecasts of y_t. Consider a weighted average,

$$f_{ct} = w_1 f_{1t} + \cdots + w_k f_{kt}, \qquad 0 \le w_i \le 1, \qquad \sum_{i=1}^{k} w_i = 1.$$

This approach is suggestive of a portfolio and the strategy of hedging. The forecast errors are

$$e_{it} = y_t - f_{it}, \; i = 1, \ldots, k, \qquad e_{ct} = y_t - f_{ct}.$$

If each of the individual forecasts is unbiased, so is f_{ct}.

Several methods for selection of the weights have been studied.

1. Equal weights. Here

$$f_{ct} = \frac{1}{k}(f_{1t} + \cdots + f_{kt}).$$

If $k = 2$,

$$e_{ct} = y_t - f_{ct} = y_t - \frac{1}{2}(f_{1t} + f_{2t}) = \frac{1}{2}(e_{1t} + e_{2t}).$$

The error variance is

$$Var(e_{ct}) = \frac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2),$$

where $\rho = \text{Corr}(e_{1t}, e_{2t})$. If $\sigma_1^2 = \sigma_2^2$ (the forecasts are of equal quality), then e_{ct} has smaller variance than e_{1t} and e_{2t}, and the risk has been lowered. If $k > 2$,

$$Var(e_{ct}) = \frac{1}{k^2}\left(\sigma_1^2 + \cdots + \sigma_k^2 + 2\sum_{i<j}\rho_{ij}\sigma_i\sigma_j\right).$$

2. Weights inversely proportional to sum of squared forecast errors. In this method higher weight is given to forecasts which have performed better in the past, and, further, the weights may be chosen to *adapt* over time. Suppose we are at time $t-1$, and we want to forecast y_t, one step ahead. Consider the n most recent one-step-ahead forecast errors (n is an arbitrary choice, and we may want to consider several different values of n),

$$e_{i,t-j} = y_{t-j} - f_{i,t-j}\,, \quad j = 1,\ldots,n, \quad i = 1,\ldots,k.$$

For method i the sum of squared forecast errors is

$$\sum_{j=1}^{n}(y_{t-j} - f_{i,t-j})^2 = \sum_{j=1}^{n} e_{i,t-j}^2\,, \qquad i = 1,\ldots,k.$$

Note this employs adaptation. This expression can be modified if we are at time $t-2$ and want to forecast two steps ahead, etc. The weight attached to the ith forecast is

$$w_i = \frac{\left(\sum_j e_{i,t-j}^2\right)^{-1}}{\left(\sum_j e_{1,t-j}^2\right)^{-1} + \cdots + \left(\sum_j e_{k,t-j}^2\right)^{-1}}\,, \qquad i = 1,\ldots,k.$$

The choice of n is often in the range from 6 to 12.

3. Regression-based weights. Suppose $k = 2$. The observation at time t is the sum of forecast and forecast error,

$$\begin{aligned} y_t &= f_{ct} + e_{ct} \\ &= w_1 f_{1t} + w_2 f_{2t} + e_{ct} \\ &= w_1 f_{1t} + (1-w_1) f_{2t} + e_{ct}\,, \end{aligned}$$

or

$$y_t - f_{2t} = w_1(f_{1t} - f_{2t}) + e_{ct}.$$

Use regression, without an intercept, to estimate w_1 from the data, employing the last n observations:

$$\text{regress } y_{t-j} - f_{2,\,t-j} \text{ on } f_{1,\,t-j} - f_{2,\,t-j}, \quad j=1,\ldots, n.$$

Note there is no guarantee that the estimate of w_1 will be in (0,1). If there are k forecasts, then

$$\text{regress } y_{t-j} - f_{k,\,t-j} \text{ on } f_{1,\,t-j} - f_{k,\,t-j}, \ldots, f_{k-1,\,t-j} - f_{k,\,t-j}, \quad j=1,\ldots, n.$$

If an estimate of w_i is negative or greater than 1, then drop the corresponding forecast and reestimate, and continue in this fashion. Some analysts argue there is no need to restrict the weights to (0,1), or to force them to add to 1. The procedure has some drawbacks: n is typically small and the regression errors are often correlated.

Empirical evidence suggests the following observations:

(i) Combined forecasts usually work well, often better than individual forecasts.

(ii) The simple average (equal weights) usually works well.

(iii) Method 2 is usually better than method 3.

(iv) If k is small and some forecasts appear to be better than others, regression-based weights are often better than those given by methods 1 and 2 (methods 1 and 2 can be quite poor under these conditions).

ARCH Models

ARCH is an acronym for autoregressive conditional heteroscedastic.

Consider the regression model

(1) $$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_r x_{rt} + u_t,$$

where

(2) $$u_t = \sigma_t \varepsilon_t,$$

with ε_t a sequence of independent and identically distributed random variables with mean 0 and variance 1, and

(3) $$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 .$$

The structure for u_t given by (2) and (3) defines an ARCH(q) model. Thus, in the regression specification (1) the error term arises from an ARCH(q) model.

The restrictions we place on the ARCH parameters are

(4) $$\alpha_0 > 0, \quad \alpha_i \geq 0, \ i = 1, \ldots, q, \quad \alpha_1 + \alpha_2 + \cdots + \alpha_q < 1.$$

The last restriction is explained below. Equations (2) and (3) imply

(5) $$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_q u_{t-q}^2 + w_t,$$

where

$$w_t = \sigma_t^2 (\varepsilon_t^2 - 1).$$

One can show that the w_t process so defined is white noise, that is,

$$Ew_t = 0,$$

$$\begin{aligned} Ew_t w_s &= 0, \qquad t \neq s, \\ &= \lambda^2, \qquad t = s, \end{aligned}$$

for some λ^2.

The ARCH structure attempts to model a time series with changing local variance and also with global stationarity in the variance pattern. In the mathematical development next we first calculate the local (conditional) mean and variance, and then we consider the global (unconditional) mean and variance.

Let Ψ_{t-1} denote the information generated by the u_t process up to time $t-1$. Then

$$E(u_t \mid \Psi_{t-1}) = E(\sigma_t \varepsilon_t \mid \Psi_{t-1}) = \sigma_t E(\varepsilon_t \mid \Psi_{t-1}) = \sigma_t E(\varepsilon_t) = \sigma_t 0 = 0.$$

This states that the local (conditional) mean is 0. Further,

$$\begin{aligned} Var(u_t \mid \Psi_{t-1}) &= E(u_t^2 \mid \Psi_{t-1}) - (E(u_t \mid \Psi_{t-1}))^2 \\ &= E(\sigma_t^2 \varepsilon_t^2 \mid \Psi_{t-1}) - 0 \\ &= \sigma_t^2 E(\varepsilon_t^2 \mid \Psi_{t-1}) \\ &= \sigma_t^2 E(\varepsilon_t^2) \\ &= \sigma_t^2. \end{aligned}$$

That is, the local (conditional) variance is σ_t^2, which depends on the q most recent values.

Next, consider the global (unconditional) mean and variance of u_t. The mean is

$$E(u_t) = E(E(u_t \mid \Psi_{t-1})) = E(0) = 0.$$

The variance is

$$\begin{aligned} Var(u_t) &= E(u_t^2) - (E(u_t))^2 \\ &= E(u_t^2) - 0 \\ &= E(E(u_t^2 \mid \Psi_{t-1})) \\ &= E(\sigma_t^2) \\ &= E(\alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_q u_{t-q}^2) \\ &= \alpha_0 + \alpha_1 E(u_{t-1}^2) + \cdots + \alpha_q E(u_{t-q}^2). \end{aligned}$$

Thus

$$\begin{aligned} Var(u_t) &= E(u_t^2) \\ &= \alpha_0 + \alpha_1 E(u_{t-1}^2) + \cdots + \alpha_q E(u_{t-q}^2). \end{aligned}$$

If we assume that all the α_i's are nonnegative and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_q < 1,$$

the u_t process is variance stationary, and we may solve the last equation to obtain

$$Var(u_t) = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_q}.$$

Finally, let us consider the covariance between u_t and u_{t+r} for r not equal to 0. This is

$$\begin{aligned} E(u_t u_{t+r}) &= E(\sigma_t \sigma_{t+r} \varepsilon_t \varepsilon_{t+r}) \\ &= E(E(\sigma_t \sigma_{t+r} \varepsilon_t \varepsilon_{t+r} \mid \Psi_{t+r-1})) \\ &= E(\sigma_t \sigma_{t+r} \varepsilon_t E(\varepsilon_{t+r} \mid \Psi_{t+r-1})) \\ &= E(\sigma_t \sigma_{t+r} \varepsilon_t E(\varepsilon_{t+r})) \\ &= E(\sigma_t \sigma_{t+r} \varepsilon_t 0) \\ &= 0. \end{aligned}$$

Therefore, the ARCH(q) process we have defined is a white noise process.

The data sets described below give simulations of the u_t values for several ARCH(q) processes, q = 1, 2, and 4. As the above discussion reveals, the process levels u_t have white noise structure, but despite this they do exhibit changing volatility. Variances are constructed by considering squared process values, and thus we consider the squared values u_t^2 to assess volatility.

ARCH(1) processes

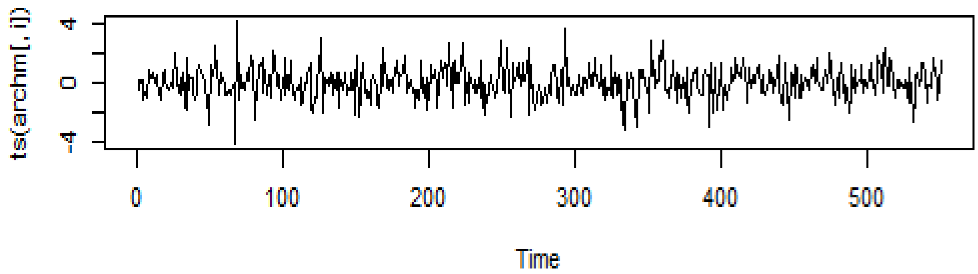
1. $\sigma_t^2 = 1 + 0.2u_{t-1}^2$

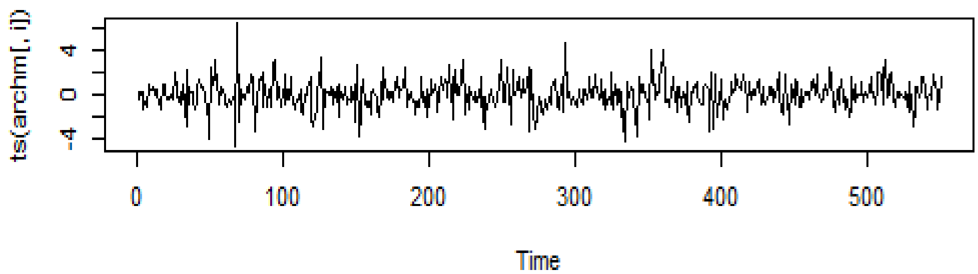
2. $\sigma_t^2 = 1 + 0.5u_{t-1}^2$

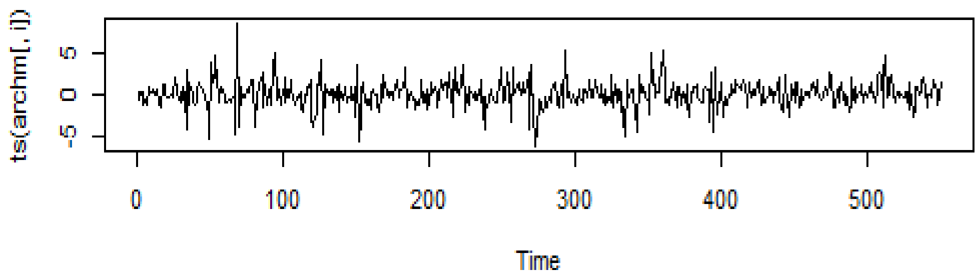
3. $\sigma_t^2 = 1 + 0.8u_{t-1}^2$

For these ARCH(1) process simulations, the following output shows time plots for the u_t series, and partial correlation plots for the u_t series and for the u_t^2 series. We expect that the partial correlation plots of the u_t series will not show significant results, and that the partial correlation plots of the u_t^2 series will exhibit some significance.

```
> arch1<-read.csv("F:/Stat71122Spring/arch1.txt")
> attach(arch1)
> head(arch1)
      arch11     arch12     arch13
1  0.1370843  0.1821179  0.2443061
2 -0.4952586 -0.4984125 -0.5059946
3  0.1552472  0.1607117  0.1663743
4  0.2259576  0.2268658  0.2278971
5 -1.1844503 -1.1935156 -1.2026817
6 -0.5671175 -0.6557695 -0.7360539

> archm<-matrix(c(arch11,arch12,arch13),ncol=3)
> par(mfrow=c(3,1))
> for(i in 1:3){
+ plot(ts(archm[,i]))
+ }
```

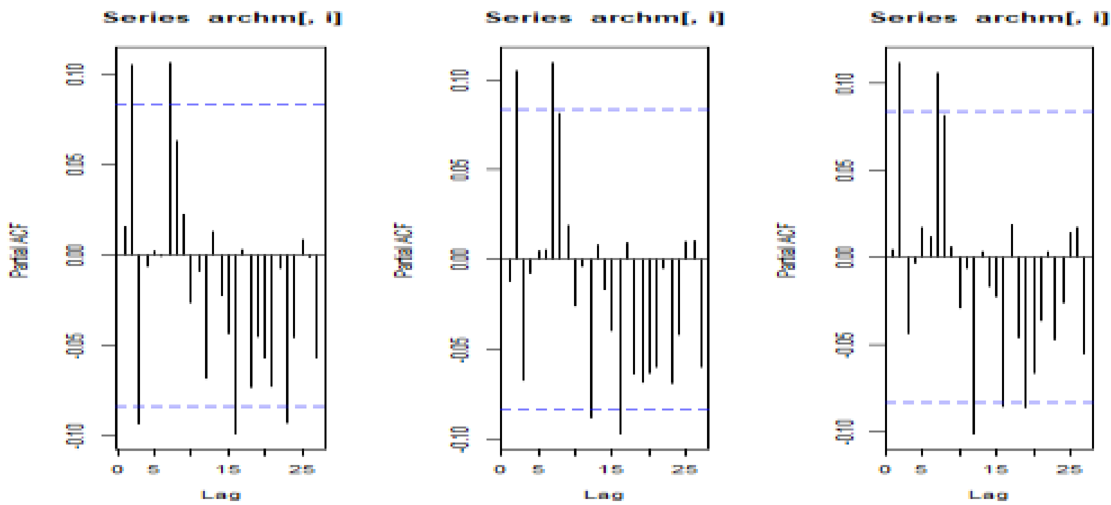






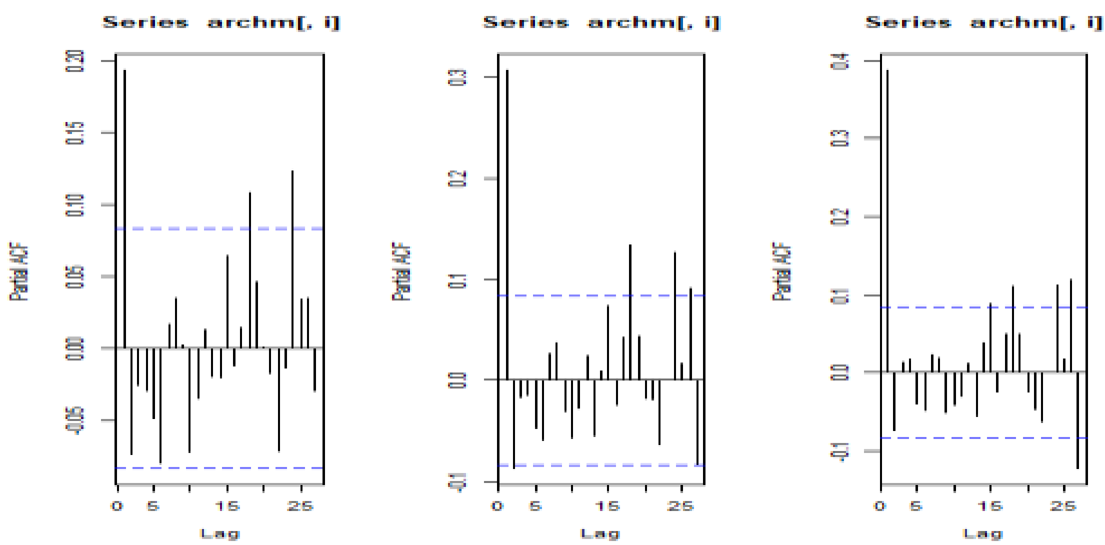
There are outliers visible in the plots. The patterns observed include short local episodes of relatively high volatility. After each such episode, the series returns to a less volatile state. Let's next examine partial correlation plots for the series.

```
> par(mfrow=c(1,3))
> for(i in 1:3){
+ pacf(archm[,i])
+ }
```



Although there are significant partial correlations for each of the simulated series, the values of the significant partial correlations are quite small and are only slightly beyond the dashed blue lines. We conclude that there is little structure, if any, beyond white noise for each of the u_t series. The partial correlations for the u_t^2 series are given next.

```
> archm<-matrix(c(arch11^2,arch12^2,arch13^2),ncol=3)
> par(mfrow=c(1,3))
> for(i in 1:3){
+ pacf(archm[,i])
+ }
```



Here the plots are different. There is clearly significant lag 1 partial correlation for each series. That is, for these ARCH(1) processes, there is little signal present in the u_t series, but notable signal in the u_t^2 series.

ARCH(2) processes

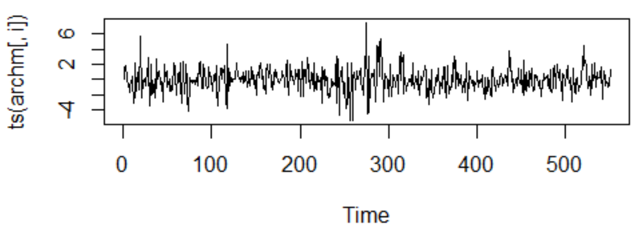
1. $\sigma_t^2 = 1 + 0.4u_{t-1}^2 + 0.3u_{t-2}^2$

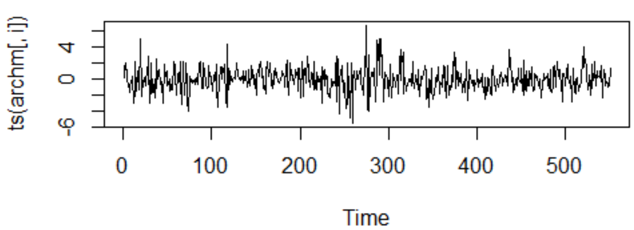
2. $\sigma_t^2 = 1 + 0.3u_{t-1}^2 + 0.4u_{t-2}^2$

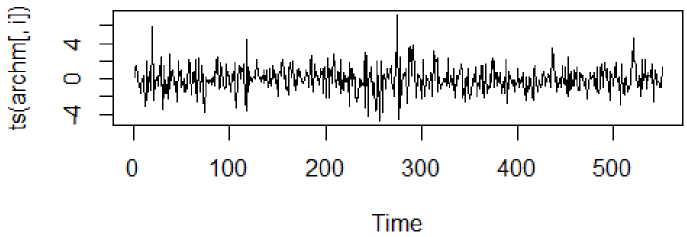
3. $\sigma_t^2 = 1 + 0.5u_{t-1}^2 + 0.1u_{t-2}^2$

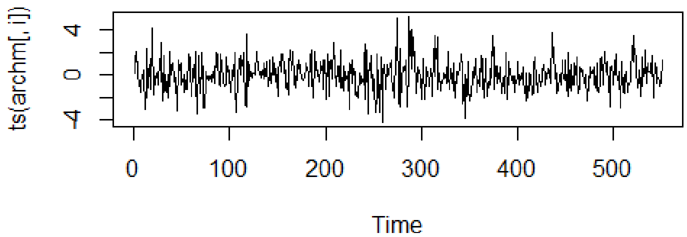
4. $\sigma_t^2 = 1 + 0.1u_{t-1}^2 + 0.5u_{t-2}^2$

```
> arch2<-read.csv("F:/Stat71122Spring/arch2.txt")
> attach(arch2)
> head(arch2)
      arch21     arch22     arch23     arch24
1  1.5996584  1.6982155  1.2785048  1.7074683
2  0.1362986  0.1308797  0.1291145  0.1089403
3  1.7878994  1.9716452  1.4526357  2.1042759
4  0.4913578  0.4792551  0.4662513  0.3913128
5 -0.2731941 -0.3086582 -0.2189005 -0.3424221
6 -0.4889579 -0.4929717 -0.4762421 -0.4858438

> archm<-matrix(c(arch21,arch22,arch23,arch24),ncol=4)
```

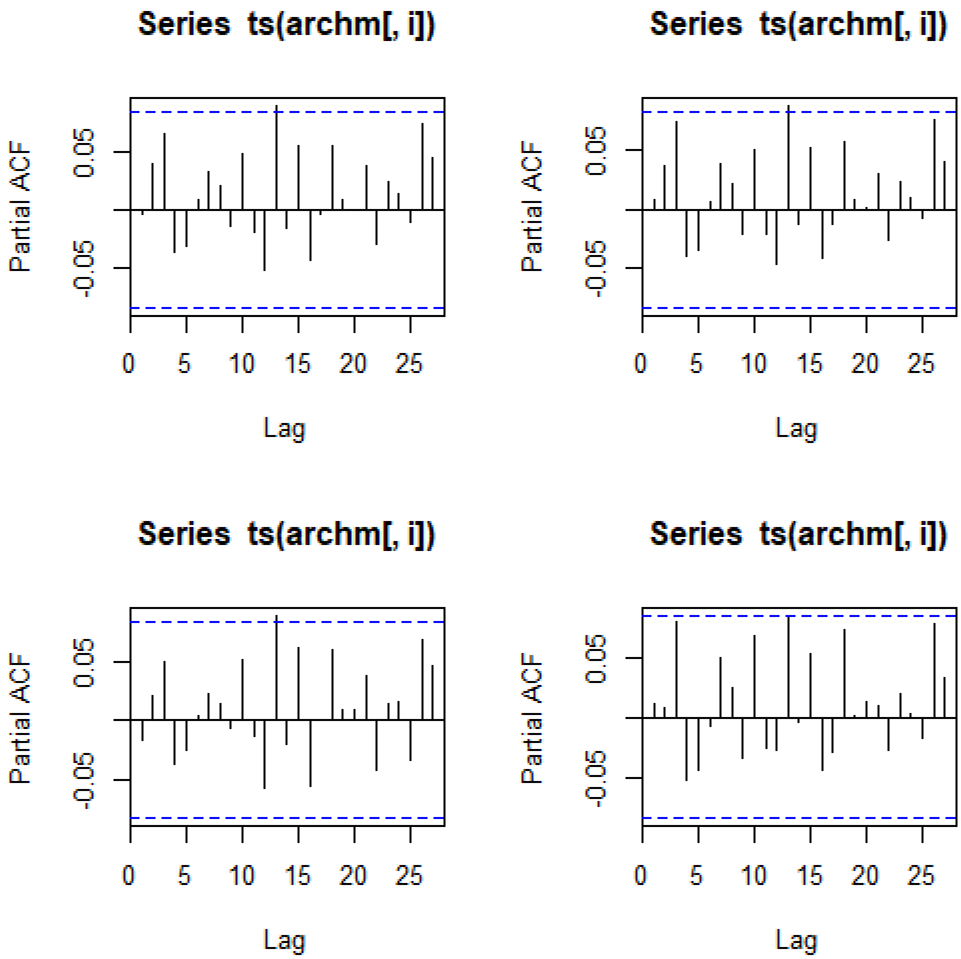




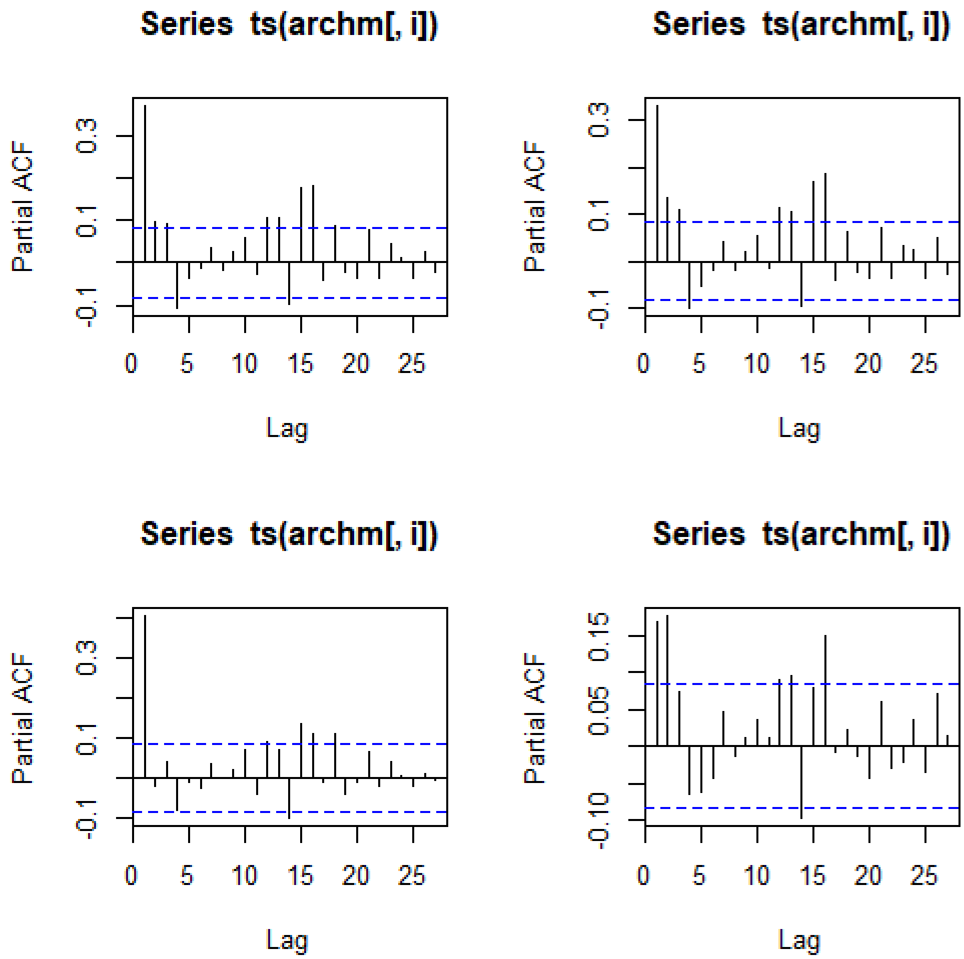




The time plots show changing volatility dominated by short bursts of relatively high variation.



The partial correlation plots for the u_t series are essentially consistent with white noise structure.

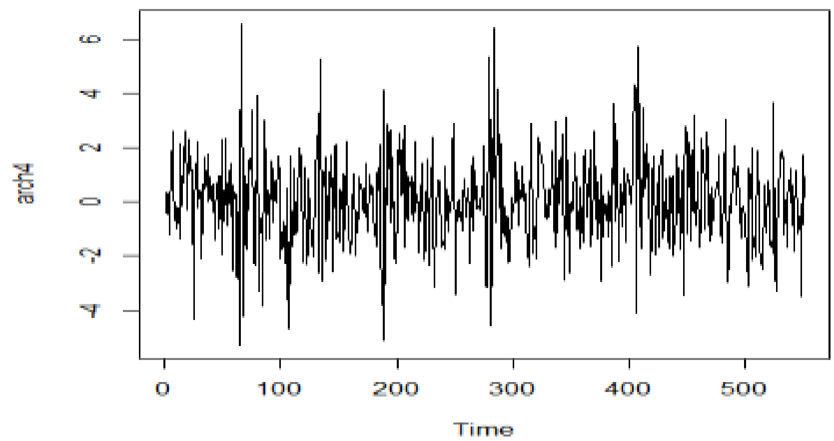


The partial correlation plots for the u_t^2 series show strong significance at lag 1, and at lag 2 for the last series.

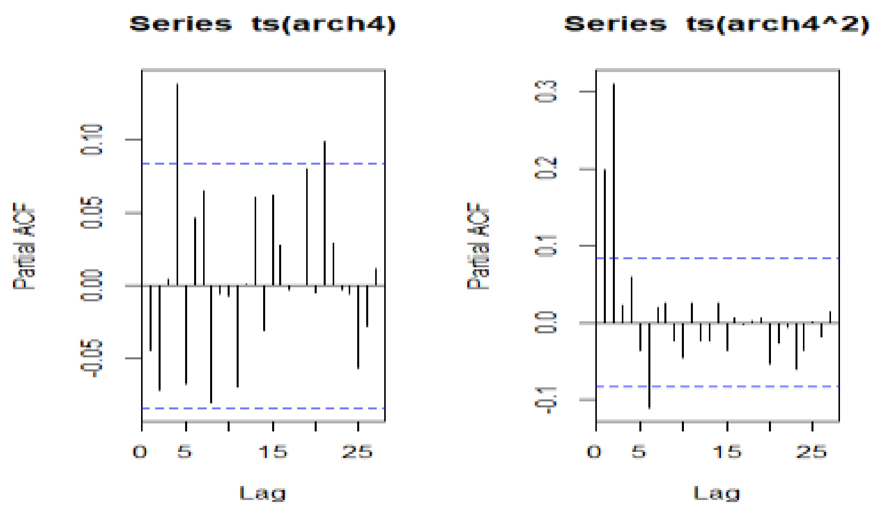
ARCH(4) process

$$\sigma_t^2 = 1 + 0.3u_{t-1}^2 + 0.2u_{t-2}^2 + 0.1u_{t-2}^2 + 0.1u_{t-4}^2$$

```
> arch4<-read.csv("F:/Stat71122Spring/arch4.txt")
> attach(arch4)
> head(arch4)
       arch4
1 -0.4235666
2  0.3650616
3  0.3152147
4 -1.2055684
5 -0.3351709
6  1.1518614
```



The partial correlation plots for the u_t and the u_t^2 series are next shown side-by-side.



How are the parameters of an ARCH model estimated? Maximum likelihood estimation is used. Often it is assumed that the errors u_t are Gaussian (normally distributed). Then y_t in (1) is conditionally Gaussian, given past time, with

$$E(y_t \mid x_{1t}, \ldots, x_{rt}, y_{t-1}, y_{t-2}, \ldots) = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_r x_{rt},$$

$$Var(y_t \mid x_{1t}, \ldots, x_{rt}, y_{t-1}, y_{t-2}, \ldots) = \sigma_t^2.$$

Then the conditional density of y_t has the form

$$\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left(-\frac{1}{2\sigma_t^2}\left(y_t-\beta_0-\beta_1x_{1t}-\cdots-\beta_rx_{rt}\right)^2\right),$$

and the likelihood function to be maximized is the product over t of these expressions.

However, often the error term in models for financial returns has heavier tails than those of the Gaussian distribution. In such cases it is more appropriate to use a heavy-tailed conditional distribution in the maximum likelihood estimation of ARCH parameters. Choices which have been used include *t* distributions with low degrees of freedom and the Laplace (double exponential) distribution. Skew versions of these distributions are also used in applications.

We can term the model given by (1)–(5) a regression–ARCH(*q*) model. The regression function addresses the mean level of the process y_t, and the ARCH(*q*) formulation describes the volatility.

Instead of a regression function for the mean level, we can use an ARMA(*m*, *n*) model, that is,

$$(1-\varphi_1B-\cdots-\varphi_mB^m)y_t=\varphi_0+(1+\theta_1B+\cdots+\theta_nB^n)u_t, \tag{6}$$

where u_t is defined by (2)–(5). We call this an ARMA(*m*, *n*)–ARCH(*q*) model.

A GARCH(*q*, *p*) model for u_t generalizes the ARCH(*q*) structure similarly to the way that an ARMA(*p*, *q*) model generalizes an AR(*p*) model. The GARCH(*q*, *p*) model can be defined by (2) and

$$\sigma_t^2=\alpha_0+\alpha_1u_{t-1}^2+\alpha_2u_{t-2}^2+\cdots+\alpha_qu_{t-q}^2+\delta_1\sigma_{t-1}^2+\delta_2\sigma_{t-2}^2+\cdots+\delta_p\sigma_{t-p}^2. \tag{7}$$

The restrictions on the GARCH(*q*, *p*) parameters are

$$\begin{gathered}\alpha_0>0,\quad \alpha_i\ge 0,\ \ i=1,\ldots,q,\quad \delta_j\ge 0,\ \ j=1,\ldots,p,\\ \alpha_1+\alpha_2+\cdots+\alpha_q+\delta_1+\delta_2+\cdots+\delta_p<1.\end{gathered} \tag{8}$$

Define $\eta_t=u_t^2-\sigma_t^2$. Then equation (7) implies

$$\begin{aligned}u_t^2&=\alpha_0+\alpha_1u_{t-1}^2+\cdots+\alpha_qu_{t-q}^2+\delta_1u_{t-1}^2+\cdots+\delta_pu_{t-p}^2+\eta_t-\delta_1\eta_{t-1}-\cdots-\delta_p\eta_{t-p}\\&=\alpha_0+\sum_{i=1}^{\max(p,q)}(\alpha_i+\delta_i)u_{t-i}^2+\eta_t-\sum_{j=1}^{p}\delta_j\eta_{t-j},\end{aligned} \tag{9}$$

where α_i is 0 if $i > q$ and δ_j is 0 if $j > p$. Moreover, the η_t sequence has mean 0, and η_t and η_{t-j} are uncorrelated if $j \geq 1$. The representation (9) shows that u_t^2 has the structure of an ARMA(max(*p*, *q*), *p*) model.

Recall that for an ARCH(*q*) process the variance of u_t is

(10)
$$\mathrm{Var}(u_t) = \mathrm{E}(u_t^2) = \frac{\alpha_0}{1-\alpha_1-\cdots-\alpha_q}.$$

For the GARCH(*q*,*p*) process defined by (2) and (7)–(9), the variance of u_t is

(11)
$$\mathrm{Var}(u_t) = \mathrm{E}(u_t^2) = \frac{\alpha_0}{1-\sum_{i=1}^{\max(p,q)}(\alpha_i+\delta_i)}.$$

The kurtosis of a random variable is defined to be the fourth central moment divided by the square of the variance. Kurtosis can be used to compare tail behavior to that of a normal random variable. The kurtosis value for a normal random variable is 3. For the u_t process the kurtosis is (the process has mean 0)

$$\frac{\mathrm{E}(u_t^4)}{\left[\mathrm{E}(u_t^2)\right]^2},$$

and the excess kurtosis is this expression minus 3. If u_t is ARCH(1) and is fourth-order stationary, the kurtosis is

(12)
$$\frac{\mathrm{E}(u_t^4)}{\left[\mathrm{E}(u_t^2)\right]^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3.$$

Thus, the ARCH(1) process has excess kurtosis relative to that of a normal random variable—it tends to produce more outliers than the normal variable. The same result holds for ARCH(*q*) processes with $q > 1$. Expression (12) shows that for fourth-order stationarity we require, beyond $0 \leq \alpha_1 < 1$, that $1-3\alpha_1^2 > 0$, hence that $0 \leq \alpha_1 < 1/\sqrt{3}$, further restricting the parameter space.

The GARCH(1, 1) process is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1\sigma_{t-1}^2, \quad \text{or} \quad u_t^2 = \alpha_0 + (\alpha_1+\delta_1)u_{t-1}^2 + \eta_t - \delta_1\eta_{t-1}.$$

If $1-(\alpha_1+\delta_1)^2 - 2\alpha_1^2 > 0$, the kurtosis is

(13) $$\frac{\mathrm{E}(u_t^4)}{\left[\mathrm{E}(u_t^2)\right]^2} = \frac{3\left[1-(\alpha_1+\delta_1)^2\right]}{1-(\alpha_1+\delta_1)^2-2\alpha_1^2} > 3.$$

That is, as with the ARCH(1) process, the GARCH(1, 1) model has tails that are heavier than those of a normal distribution. Also, note that this kurtosis calculation, which assumes fourth-order stationarity, requires further restriction on the parameter space, namely, that $1-(\alpha_1+\delta_1)^2-2\alpha_1^2 > 0$.

In most financial applications, low order GARCH models are fit to the data. These include the GARCH(1, 1), GARCH(1, 2) and GARCH(2, 1) models. In fact, the GARCH(1, 1) model is most commonly fit.

Typically, appropriate software is used to fit GARCH models. For an ARMA–GARCH model, one estimates the ARMA and GARCH parameters simultaneously. However, the representation (9) for a GARCH model suggests that one may employ a two-step procedure, as follows. See page 140 in Tsay.

1. Estimate the mean level of the time series with an ARMA(m, n) (or a regression model, if appropriate) and calculate the residuals, denoted by u_t.
2. Square the residuals, and estimate the parameters in the representation (9). That is, for a GARCH(q, p) model, fit an ARMA(max(p, q), p) model to the squared residuals. In this ARMA fit the autoregressive parameters are estimates of the parameters $\alpha_i + \delta_i$, and the moving average parameters give estimates of the δ_i.

This procedure produces results which approximate the correct maximum likelihood methodology. Tsay notes that "limited experience shows that this simple approach often provides good approximations, especially when the sample size is moderate or large" (page 140). Of course, this two-step procedure employs maximum likelihood estimation with a normal distribution, rather than with a t distribution, or a Laplace distribution, or even a skew t distribution, for the GARCH fitting.

It is common to fit a GARCH(1, 1) model to the returns of a financial time series. The shorter the time period over which the return is calculated (e.g., the return might be calculated for a year, a month, a week, a day, an hourly, etc.), the greater the kurtosis tends to be, and thus the more spiky and more volatile the return series tends to be. This is because of the central limit theorem (I'll discuss this in class).

In the example below we'll employ both the above two-step procedure and maximum likelihood methodology.

For use of the two-step procedure, we begin by examining the time plot of the return series and the correlations and partial correlations of the series. We then fit an ARMA model to the returns, to model the mean level. The residuals from this ARMA fit are the

u_t values. Next we fit a GARCH model to these residuals. To do so, we plot the squared residuals and fit an ARMA model to them, as in step 2 above. From this ARMA we determine the α_i and δ_i values. Further, we calculate the estimated σ_t values from the fitted GARCH model [using (7)] and form the standardized residuals as

$$\tilde{u}_t = \frac{u_t}{\sigma_t}. \tag{14}$$

If the model has been properly fit to describe the data, the standardized residuals should resemble a white noise sequence. We examine the time plot, correlations, and partial correlations of the standardized residuals. In addition, we look at their spectral density.

Tsay notes some problems associated with ARCH and GARCH modelling (page 119).

(i) The models use the squares of residuals and thus assume that positive and negative shocks have the same effects on volatility. This is somewhat questionable for financial time series. Extensions of the GARCH model have been developed. One is the exponential GARCH model, written EGARCH. It allows for asymmetry in positive and negative asset returns, for example. One can also employ a skew t distribution in the maximum likelihood estimation.

(ii) The ARCH and GARCH models are quite restrictive, given the limits on the parameter values, such as those shown in (4) and (8), and, in addition, those required for fourth-order stationarity. For GARCH(q, p) with values of p and q greater than 1, the restrictions on the parameter space are complicated and hard to describe; they are rather strong. Recall that I have discussed in class the fact that there are also limitations on the parameter space for ARMA processes, although they are not as severe as those for the ARCH and GARCH models.

(iii) GARCH models cannot anticipate volatility changes. They can only react to past shocks. As such, they often tend to overpredict volatility after a burst has dissipated, or to be slow to catch a burst after it has begun.

(iv) The GARCH model is not structural. That is, it does not relate to variables which might explain sources of variation for the time series.

The above discussion shows that the parameter spaces for ARCH and GARCH models are severely restricted. See (4), (8), and (12), e.g. As an aside, let's emphasize that restrictions also occur for ARMA(p, q) models, and these can be severe if p and q are large. Consider the AR(p) model $(1-\phi_1 \mathrm{B} - \cdots - \phi_p \mathrm{B}^p)\left(y_y - \mu\right) = \varepsilon_t$. The required parameter restriction is that the zeros of the polynomial

$$1 - \phi_1 z - \cdots - \phi_p z^p$$

be strictly greater than 1 in magnitude. For p = 1, we require $-1 < \phi_1 < 1$, and, for p = 2, (ϕ_1, ϕ_2) has to be inside the triangle bounded by the lines

$$\phi_2 = 1 - \phi_1, \quad \phi_2 = 1 + \phi_1, \quad \phi_2 = -1.$$

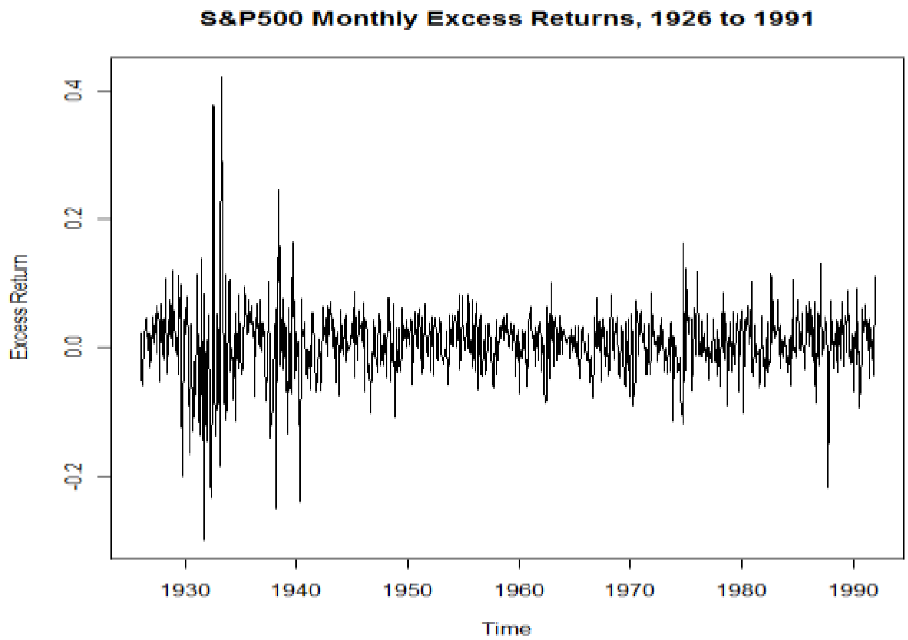
Piccolo (1982, *Journal of Time Series Analysis*, Vol. 3, pp. 245–247) has calculated the hypervolumes of the parameter regions for autoregressive models. These calculations illustrate how severe the parameter constraints become as p increases. Some numerical values follow.

p	Volume
1	2
2	4
3	5.33333
4	7.11111
5	7.58519
6	8.09086
7	7.39736
8	6.76330
14	1.13407
20	0.05312
30	0.00005

I'll discuss these numbers in class.

Example 1. Let's consider monthly excess returns for the S&P 500 for the period 1926 to 1991. This is the series Tsay considers in his Example 3.3. The excess returns are the S&P 500 returns minus the return for a risk-free investment such as that for the last 30 days of a T bill.

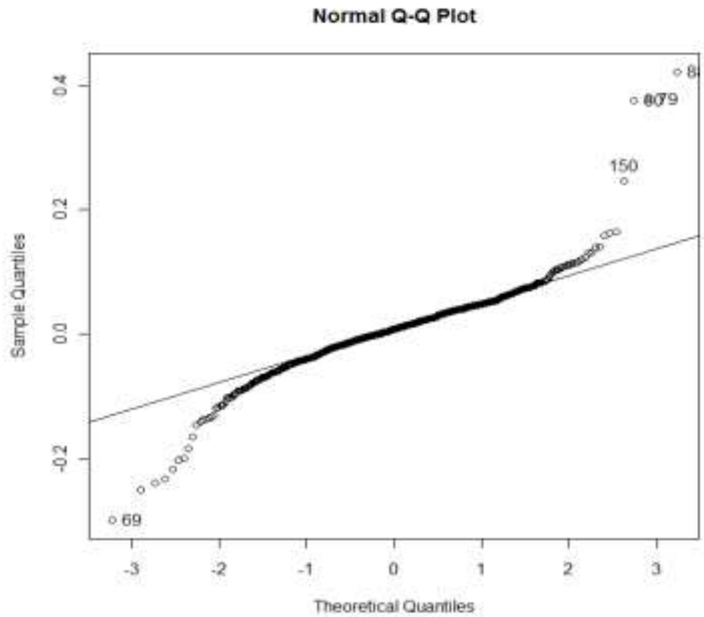
```
> sp500monthly<-read.csv("F:/Stat71122Spring/sp500.txt")
> attach(sp500monthly)
> head(sp500monthly)

  Year Month S.P500ExcessReturn
1 1926     1             0.0225
2 1926     2            -0.0440
3 1926     3            -0.0591
4 1926     4             0.0227
5 1926     5             0.0077
6 1926     6             0.0432

> sp500.ts<-ts(S.P500ExcessReturn,start=c(1926,1),freq=12)
> plot(sp500.ts,xlab="Time",ylab="Excess Return",main="S&P500 Monthly
Excess Returns, 1926 to 1991")
```



There was especially high volatility during the 1930s, the depression years, and this volatility didn't abate until World War II began. The outlier for October 1987 is also visible.

Before fitting a model, let's look at the distribution of the data.

```
> qq<-qqnorm(S.P500ExcessReturn)
> qqline(S.P500ExcessReturn)
> identify(qq)
```



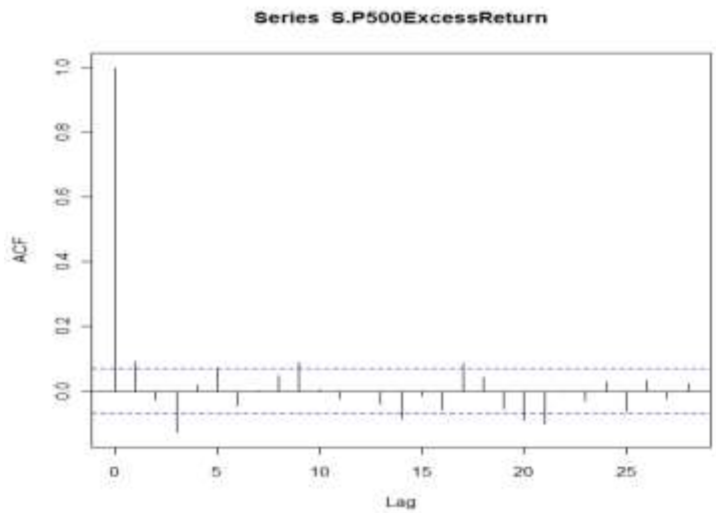
The outlier data points marked are almost all during the Great Depression years. To obtain skewness and kurtosis measures, we install the **moments** package. Skewness is a measure of asymmetry for a distribution. If a distribution is symmetric, its skewness is 0. Skewness is estimated with the third central moment, normalized by division by the third power of the standard deviation. Recall that kurtosis is estimated with the fourth central moment, normalized by division by the fourth power of the standard deviation.

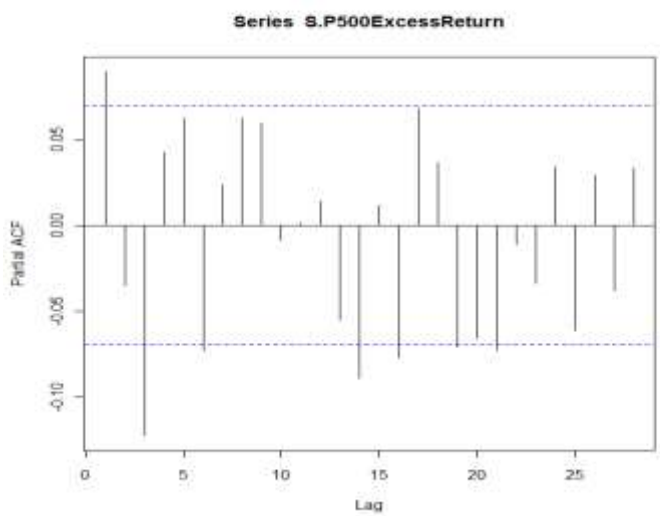
```
> library("moments")

> skewness(S.P500ExcessReturn)
[1] 0.4113397
> kurtosis(S.P500ExcessReturn)
[1] 12.30025
```

Relative to normality, the distribution is heavy-tailed, as the kurtosis calculation shows. Moreover, there is slight skewness toward larger values.

We proceed to model both the level and the volatility of this time series. We'll provide an ARMA model for the level, followed by a GARCH model for the volatility.

Tsay fits an ARMA–GARCH model, with simultaneous estimation of the ARMA and GARCH structures. He also refers to use of the two-stage estimation procedure described above (see his page 140). I'll start with the two-stage procedure. This, of course, employs the normal distribution to do the fitting. Here are the correlations and partial correlations of the excess returns.



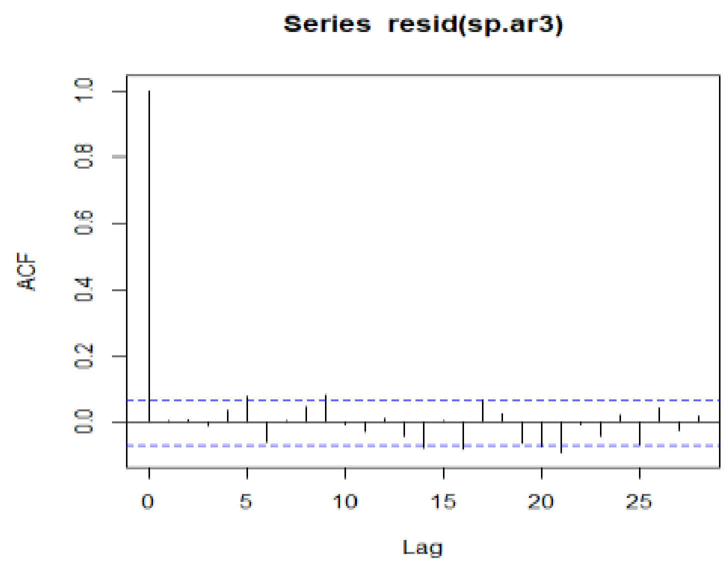


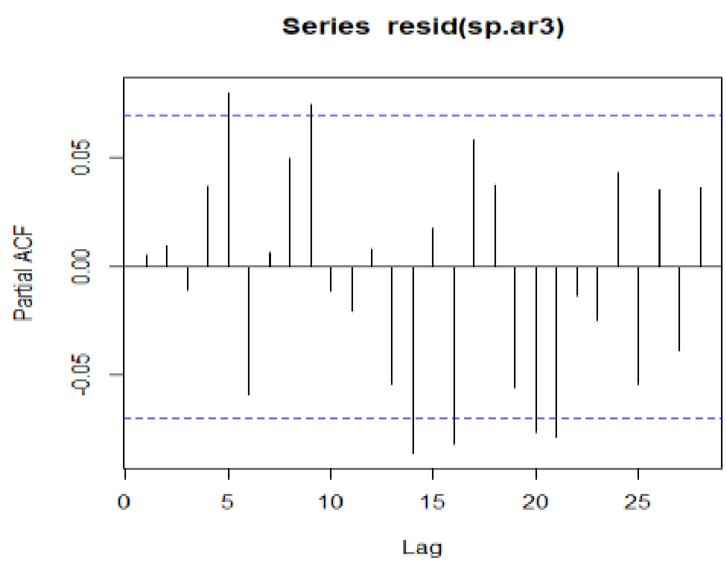
The signal is not strong. There are quite a few slightly significant autocorrelations and partial correlations. Let's try an AR(3) fit.

```
> sp.ar3<-arima(sp500.ts,order=c(3,0,0))
> sp.ar3

Call:
arima(x = sp500.ts, order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3  intercept
      0.0890  -0.0238  -0.1229     0.0062
s.e.  0.0353   0.0355   0.0353     0.0019

sigma^2 estimated as 0.00333:  log likelihood = 1135.25,  aic = -2260.5

> library("lmtest")
```

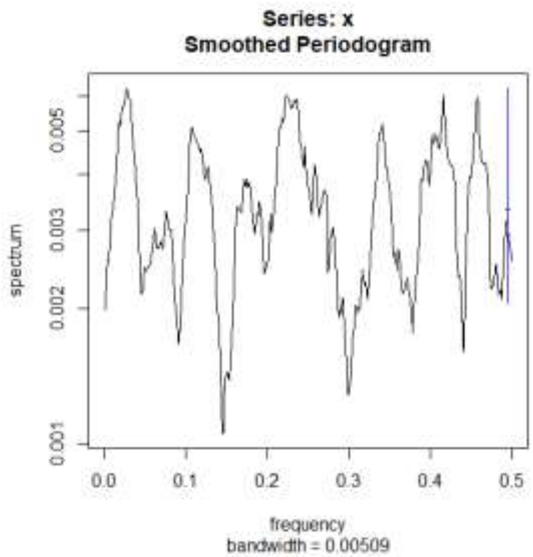
```
> coeftest(sp.ar3)

z test of coefficients:

            Estimate  Std. Error z value  Pr(>|z|)
ar1        0.0890483  0.0353216  2.5211 0.0116999 *
ar2       -0.0238404  0.0354719 -0.6721 0.5015251
ar3       -0.1228632  0.0353403 -3.4766 0.0005079 ***
intercept  0.0061599  0.0019399  3.1754 0.0014962 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tsay gives the same fit at the bottom of page 135. The residual correlations and partial correlations for this AR(3) fit follow.



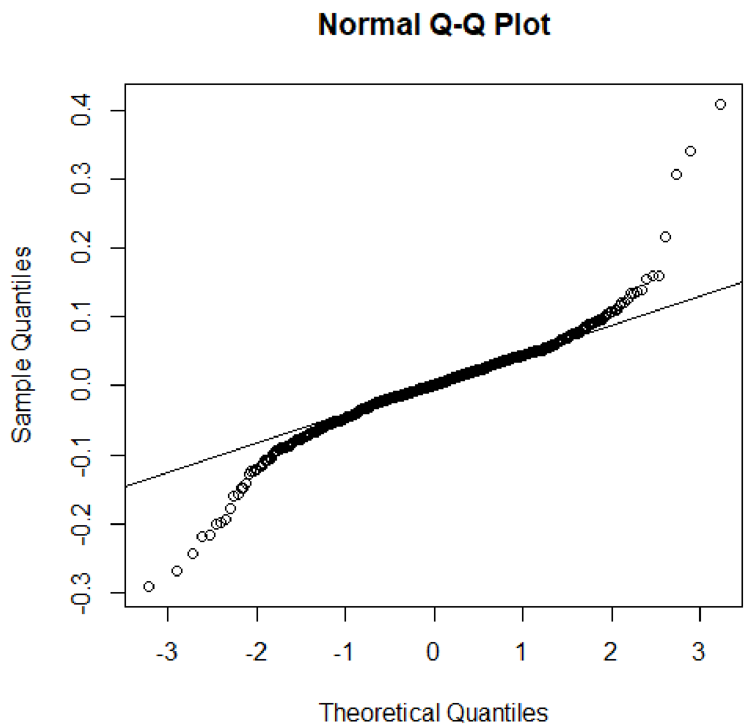


Some of the residual correlations and partial correlations are statistically significant, but only modestly so. The residual spectral density and the Bartlett statistic follow.




```
> library("hwwntest")

> bartlettB.test(ts(resid(sp.ar3)))

        Bartlett B Test for white noise

data:
= 0.61763, p-value = 0.8401
```

The Bartlett test does not reject the hypothesis of reduction to white noise by the model. Before continuing to model the volatility, let's look at the normal quantile plot of the AR(3) residuals and also calculate their skewness and kurtosis.




```
> skewness(resid(sp.ar3))
[1] 0.2448624
> kurtosis(resid(sp.ar3))
[1] 10.64358
```

Thus, the residuals from the AR(3) fit have a distribution with tails larger than for the normal distribution. They have large kurtosis and some positive skewness, as prior to the fit. However, the skewness and kurtosis values have abated somewhat as a result of fitting the AR(3) model.

Next, continuing with the two-step procedure, we square the residuals and estimate GARCH structure via the representations (7) and (9). Let's use the GARCH(1, 1) model shown above (13),

(15) $$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 \sigma_{t-1}^2.$$

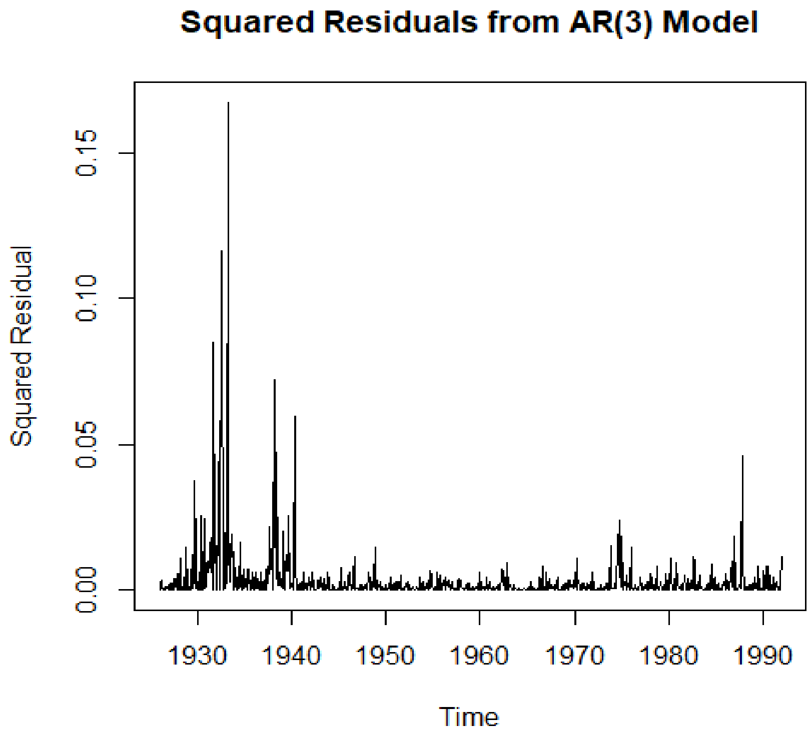
We rewrite this as in (9),

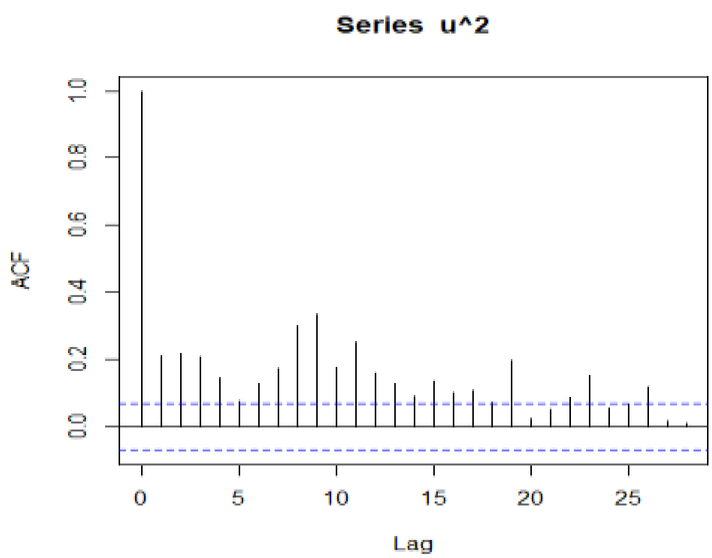
(16) $$u_t^2 = \alpha_0 + (\alpha_1 + \delta_1)\, u_{t-1}^2 + \eta_t - \delta_1 \eta_{t-1}.$$

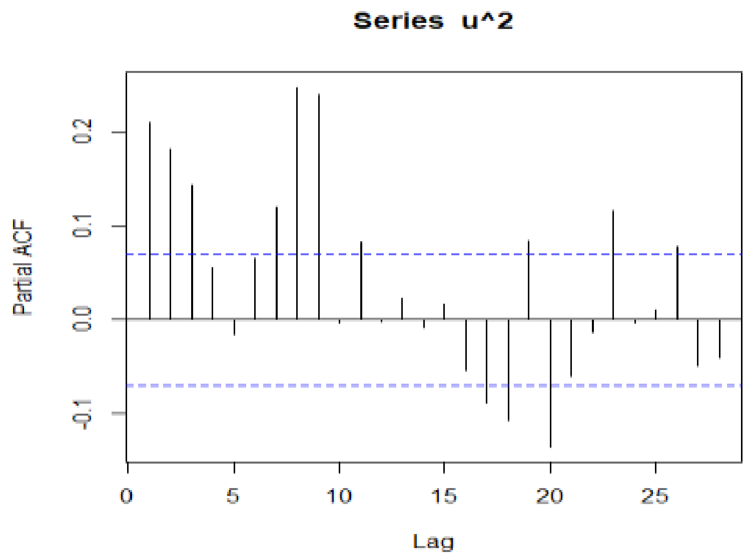
To estimate this model, we fit an ARMA(1, 1) to the squared residuals.

First here are the plots of the squared residuals, and of their correlations and partial correlations.

```
> u<-resid(sp.ar3)
```







There is clearly a good deal of structure in the squared residuals. The second stage of the two-step procedure follows.

```
> spu2arma<-arima(ts(u^2),order=c(1,0,1))
> spu2arma

Call:
arima(x = ts(u^2), order = c(1, 0, 1))

Coefficients:
         ar1      ma1  intercept
      0.9662  -0.8680     0.0033
s.e.  0.0129   0.0233     0.0013

sigma^2 estimated as 9.293e-05:  log likelihood = 2552.3,  aic = -
5096.61

> coeftest(spu2arma)

z test of coefficients:

            Estimate Std. Error  z value Pr(>|z|)
ar1        0.9662123  0.0129488  74.6180  < 2e-16 ***
ma1       -0.8680031  0.0233130 -37.2325  < 2e-16 ***
intercept  0.0033128  0.0013030   2.5424  0.01101 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

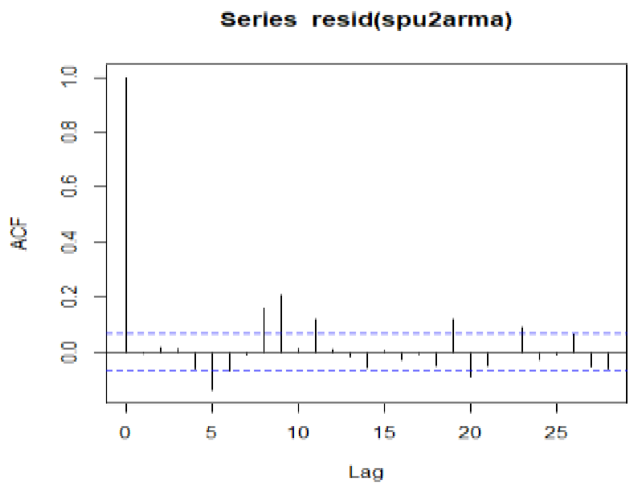
Note that the *t*-ratios are very large. This fitted model is, from (16) (we need to reverse the algebraic sign of the moving average coefficient estimate),

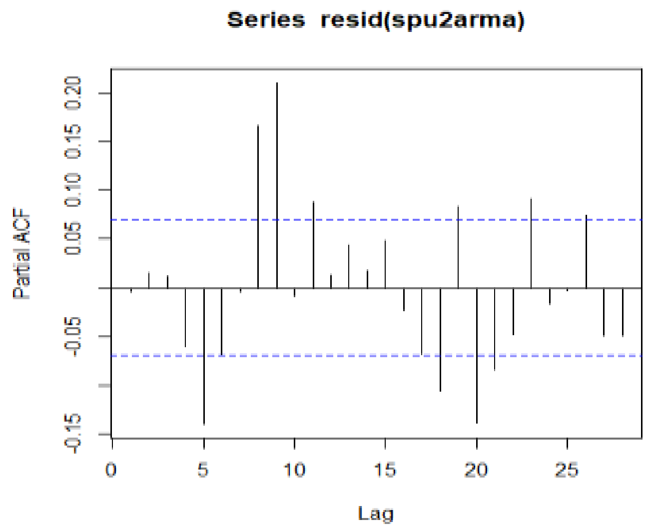
$$
\begin{aligned}
u_t^2 &= \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\delta}_1)\, u_{t-1}^2 + \eta_t - \hat{\delta}_1 \eta_{t-1} \\
&= 0.00011 + 0.9662 u_{t-1}^2 + \eta_t - 0.8680 \eta_{t-1}.
\end{aligned}
$$

The correct value for the constant here is tricky, given that we are using the two-step procedure. The value is determined from part B on page 2 of 22 March notes. Writing this fitted model in the form of (15), we have

(17)
$$
\begin{aligned}
\sigma_t^2 &= \hat{\alpha}_0 + \hat{\alpha}_1 u_{t-1}^2 + \hat{\delta}_1 \sigma_{t-1}^2 \\
&= 0.00011 + 0.0982 u_{t-1}^2 + 0.8680\, \sigma_{t-1}^2.
\end{aligned}
$$

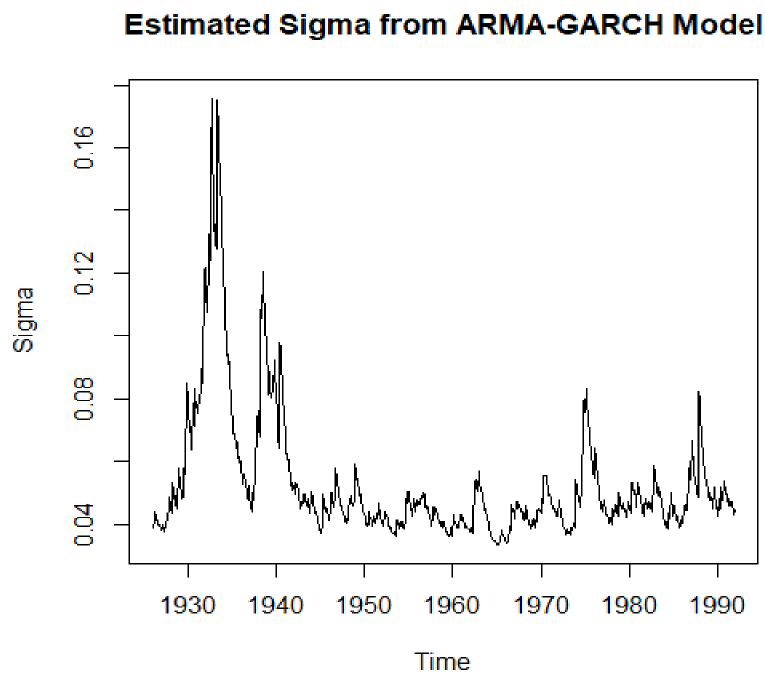
Tsay reports similar values (see page 140). Here are the residual correlations and partial correlations from this fit.





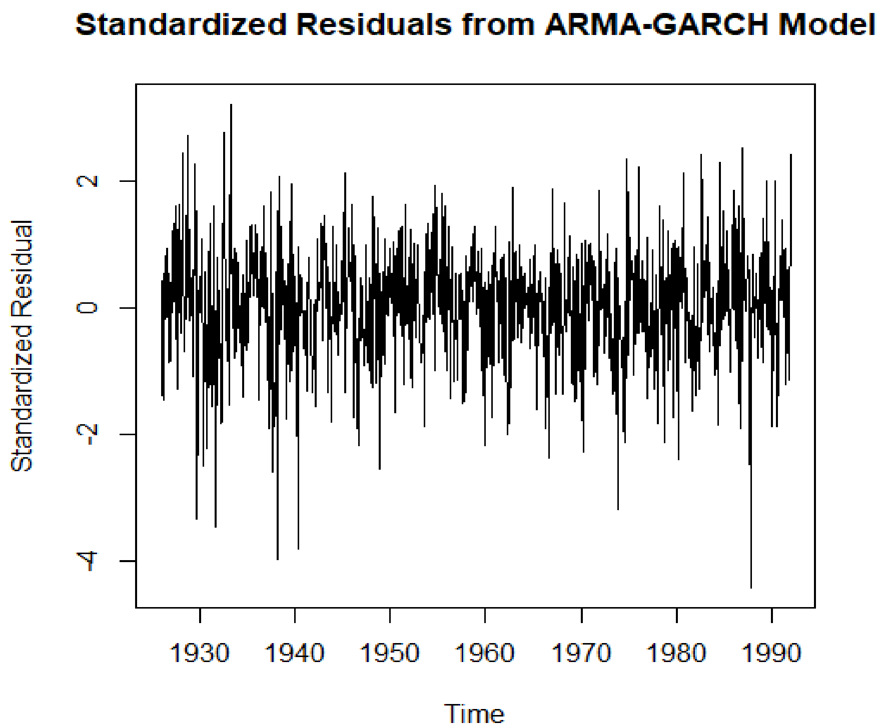
There is remaining structure in these residuals. Next let's plot the estimate of σ_t, as calculated from (17).

```
> cf<-coef(spu2arma)
> cf
         ar1          ma1    intercept
 0.966212320 -0.868003054  0.003312807

> T<-length(u)
> u2<-u^2
> cf<-coef(spu2arma)
> sigma2<-rep(0,T)
> cc<-cf[3]*(1-cf[1])
> cf1<-cf[1]+cf[2];sigma2[1]<-0.0016
> for(j in 2:T){
+ j1<-j-1
+ sigma2[j]<-cc+cf1*u2[j1]-cf[2]*sigma2[j1]
+ }
>
plot(ts(sqrt(sigma2),start=c(1926,1),freq=12),xlab="Time",ylab="Sigma",
main="Estimated Sigma from ARMA-GARCH Model")
```



Asymmetry in the behavior of the local volatility is evident. Let's form the standardized residuals, as at (14), and see if the volatility has been properly estimated.

```
> stdresid<-u/sqrt(sigma2)
>
plot(ts(stdresid,start=c(1926,1),freq=12),xlab="Time",ylab="Standardize
d Residual",main="Standardized Residuals from ARMA-GARCH Model")
```



There is a considerable amount of attenuation of the varying volatility. However, some variation does remain. Attempts to fit GARCH(2, 1) and GARCH(1, 2) models do not offer improvement.

Finally, let's look at the skewness and kurtosis of the standardized residuals.

```
> library("moments")
> skewness(stdresid)
[1] -0.3638164
> kurtosis(stdresid)
[1] 4.243771
```

The skewness is modest. The kurtosis is now only 1.24 in excess of 3, the value for the normal distribution. Recall that the kurtosis of S&P 500 series was calculated as 12.30, and the kurtosis for the residuals from the AR(3) model fit to the series was found to be 10.64. Thus, the present value, 4.24, indicates that the GARCH fit has substantially captured the volatility structure. Also, the results perhaps suggest that the GARCH fit should employ maximum likelihood with a distribution other than the normal, such as a *t* distribution or a Laplace distribution. Also, it may be advisable to employ an asymmetric *t* distribution.

Next we turn to options for GARCH estimation which are available in R. We start with the **tseries** package.

```
> install.packages("tseries")
> library(tseries)
```

The default for the **garch** command is to fit a GARCH(1,1) model. **trace=FALSE** suppresses output showing the iteration steps in fitting, and **grad="numerical"** provides numerical estimation of the gradient and is more robust in providing algorithmic convergence.

We fit a GARCH(1,1) model to the residuals from the AR(3) fit given above for the level of the time series.

```
> sp500garch11<-garch(u,grad="numerical",trace=FALSE)
> summary(sp500garch11)

Call:
garch(x = u, grad = "numerical", trace = FALSE)

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max 
-4.54499 -0.54453  0.01363  0.62118  3.07507 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 7.779e-05   2.434e-05    3.196   0.0014 ** 
a1 1.155e-01   2.076e-02    5.564 2.64e-08 ***
b1 8.613e-01   2.067e-02   41.665  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals 
X-squared = 58.383, df = 2, p-value = 2.101e-13


        Box-Ljung test

data:  Squared.Residuals 
X-squared = 1.3546, df = 1, p-value = 0.2445


> confint(sp500garch11)
          2.5 %      97.5 %
a0 3.007584e-05 0.000125496
a1 7.480143e-02 0.156166395
b1 8.207473e-01 0.901776406
```

The parameter estimates are only slightly different from those obtained above with the approximate two-step procedure.

Both analyses above used the normal distribution for maximum likelihood estimation. The function **garchFit** in the package **fGarch** allows one to choose among the following distributions: normal, skew normal, generalized error (Laplace), skew generalized error, standardized *t* distribution, skew standardized *t* distribution.

Let's use the **fGarch** package.

```
> library(fGarch)
Loading required package: timeDate
Loading required package: timeSeries
Loading required package: fBasics
Loading required package: MASS
```

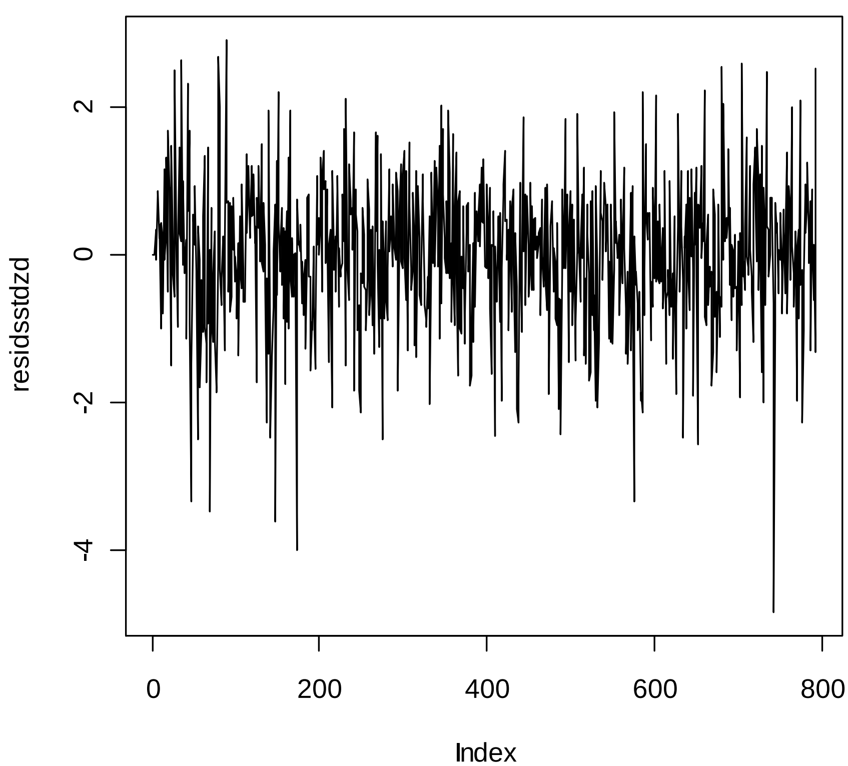
Now we obtain the AR(3) and GARCH(1,1) parameters in one estimation. To use this procedure, one needs to first determine an appropriate ARMA model for the level of the time series.

```
> model1<-garchFit(~arma(3,0)+garch(1,1),data=sp500.ts,trace=FALSE)
> summary(model1)

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(3, 0) + garch(1, 1), data = sp500.ts,
    trace = FALSE)

Mean and Variance Equation:
 data ~ arma(3, 0) + garch(1, 1)
<environment: 0x02441184>
 [data = sp500.ts]

Conditional Distribution:
 norm

Coefficient(s):
        mu         ar1         ar2         ar3       omega      alpha1
 7.7077e-03   3.1968e-02  -3.0261e-02  -1.0649e-02   7.9746e-05   1.2425e-01
      beta1
 8.5302e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      7.708e-03   1.607e-03    4.798 1.61e-06 ***
ar1     3.197e-02   3.837e-02    0.833  0.40473
ar2    -3.026e-02   3.841e-02   -0.788  0.43076
ar3    -1.065e-02   3.756e-02   -0.284  0.77677
omega   7.975e-05   2.810e-05    2.838  0.00454 **
alpha1  1.242e-01   2.247e-02    5.529 3.22e-08 ***
beta1   8.530e-01   2.183e-02   39.075  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Log Likelihood:
 1272.179    normalized:  1.606287

Standardised Residuals Tests:
                                Statistic p-Value
 Ljung-Box Test     R    Q(10)  11.56744  0.315048
 Ljung-Box Test     R    Q(15)  17.78747  0.2740039
 Ljung-Box Test     R    Q(20)  24.11916  0.2372256
 Ljung-Box Test     R^2  Q(10)  10.31614  0.4132089
 Ljung-Box Test     R^2  Q(15)  14.22819  0.5082978
 Ljung-Box Test     R^2  Q(20)  16.79404  0.6663038
 LM Arch Test       R    TR^2   13.34305  0.3446075

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.194897 -3.153581 -3.195051 -3.179018
```

The standardized residuals are easily obtained.

```
> residsstdzd<-residuals(model1,standardize=TRUE)
> plot(residsstdzd,type='l')
```




```
> kurtosis(residsstdzd)
[1] 1.256575
attr(,"method")
[1] "excess"
```

Note the output gives excess kurtosis. Recall that the excess kurtosis of the standardized residuals obtained with the two-step procedure was 4.24 – 3 = 1.24 (page 31).

The following is an AR(3)-GARCH(1,1) fit with a symmetric *t*-distribution.

```
> model2<-
garchFit(~arma(3,0)+garch(1,1),data=sp500.ts,trace=FALSE,cond.dist="std
")
> summary(model2)

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(3, 0) + garch(1, 1), data = sp500.ts,
    cond.dist = "std", trace = FALSE)

Mean and Variance Equation:
 data ~ arma(3, 0) + garch(1, 1)
<environment: 0x024ac4c0>
 [data = sp500.ts]

Conditional Distribution:
 std

Coefficient(s):
        mu          ar1          ar2          ar3        omega       alpha1
 0.00856064   0.01637895  -0.00877946  -0.00034328   0.00012656   0.11647067
      beta1        shape
 0.83942500   6.83281956

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      8.561e-03   1.613e-03    5.309 1.11e-07 ***
ar1     1.638e-02   3.699e-02    0.443  0.65790
ar2    -8.779e-03   3.660e-02   -0.240  0.81040
ar3    -3.433e-04   3.675e-02   -0.009  0.99255
omega   1.266e-04   4.598e-05    2.753  0.00591 **
alpha1  1.165e-01   2.781e-02    4.188 2.81e-05 ***
beta1   8.394e-01   3.244e-02   25.876  < 2e-16 ***
shape   6.833e+00   1.644e+00    4.157 3.22e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 1285.979    normalized:  1.623711
```

The number of degrees of freedom chosen by the estimation is 6.83.

```
Standardised Residuals Tests:
                                Statistic p-Value
 Ljung-Box Test     R    Q(10)  11.25961  0.3376532
 Ljung-Box Test     R    Q(15)  17.76819  0.2750483
 Ljung-Box Test     R    Q(20)  24.51884  0.2204625
 Ljung-Box Test     R^2  Q(10)  10.67343  0.3835203
 Ljung-Box Test     R^2  Q(15)  16.05917  0.3781302
 Ljung-Box Test     R^2  Q(20)  18.80188  0.5347357
 LM Arch Test       R    TR^2   14.91177  0.246297

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.227220 -3.180002 -3.227421 -3.209072
```

One more example, use of a skew *t*-distribution.

```
> model3<-
garchFit(~arma(3,0)+garch(1,1),data=sp500.ts,trace=FALSE,cond.dist="sst
d")

> summary(model3)

Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(3, 0) + garch(1, 1), data = sp500.ts,
    cond.dist = "sstd", trace = FALSE)

Mean and Variance Equation:
 data ~ arma(3, 0) + garch(1, 1)
<environment: 0x01edb5cc>
 [data = sp500.ts]

Conditional Distribution:
 sstd

Coefficient(s):
         mu          ar1          ar2          ar3        omega       alpha1
 0.00780992  -0.00031329  -0.01142827  -0.00645324   0.00012187   0.11423480
      beta1         skew        shape
 0.84189659   0.89892089   7.18120161

Std. Errors:
 based on Hessian
```

```
Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      7.810e-03   1.634e-03    4.780 1.76e-06 ***
ar1    -3.133e-04   3.749e-02   -0.008  0.99333
ar2    -1.143e-02   3.643e-02   -0.314  0.75373
ar3    -6.453e-03   3.679e-02   -0.175  0.86078
omega   1.219e-04   4.498e-05    2.709  0.00674 **
alpha1  1.142e-01   2.719e-02    4.202 2.65e-05 ***
beta1   8.419e-01   3.212e-02   26.215  < 2e-16 ***
skew    8.989e-01   4.695e-02   19.147  < 2e-16 ***
shape   7.181e+00   1.825e+00    3.936 8.30e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 1288.088    normalized:  1.626373



Standardised Residuals Tests:
                                Statistic p-Value
 Ljung-Box Test     R    Q(10)  11.85546  0.2948467
 Ljung-Box Test     R    Q(15)  18.36293  0.2440661
 Ljung-Box Test     R    Q(20)  25.30622  0.1899661
 Ljung-Box Test     R^2  Q(10)  10.95451  0.3610729
 Ljung-Box Test     R^2  Q(15)  16.47984  0.3508974
 Ljung-Box Test     R^2  Q(20)  19.38844  0.4967212
 LM Arch Test       R    TR^2   15.26779  0.2271175

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-3.230019 -3.176899 -3.230274 -3.209603

> residsstdzd2<-residuals(model3,standardize=TRUE)
> kurtosis(residsstdzd2)
[1] 1.49143
attr(,"method")
[1] "excess"
```

Some final comments.

GARCH models can be used to forecast volatility. However, one should remember that the models do not anticipate large shocks. Rather, they react to the shocks, and thus their forecasts are problematic. In fact, as the forecast horizon increases, the forecasts converge to the estimate of the (global) variance, that is, $E(u_t^2)$.

Multivariate GARCH models have been developed. They are discussed in Chapter 10 of Tsay. An interesting application of a multivariate GARCH model is estimation of the beta for the excess return of a stock. I'll discuss this briefly in class.