Chapter 8

Noise: Conditional Variance Dynamics

The celebrated Wold decomposition makes clear that every covariance stationary series may be viewed as ultimately driven by underlying weak white noise innovations. Hence it is no surprise that every forecasting model discussed in this book is driven by underlying white noise. To take a simple example, if the series y_t follows an AR(1) process, then $y_t = \phi y_{t-1} + \varepsilon_t$, where ε_t is white noise. In some situations it is inconsequential whether ε_t is weak or strong white noise, that is, whether ε_t is independent, as opposed to merely serially uncorrelated. Hence, so to simplify matters we sometimes assume strong white noise, $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$. Throughout this book, we have thus far taken that approach, sometimes explicitly and sometimes implicitly.

When ε_t is independent, there is no distinction between the unconditional distribution of ε_t and the distribution of ε_t conditional upon its past, by definition of independence. Hence σ^2 is both the unconditional and conditional variance of ε_t. The Wold decomposition, however, does not require that ε_t be serially independent; rather it requires only that ε_t be serially uncorrelated. If ε_t is dependent, then its unconditional and conditional distributions will differ. We denote the unconditional innovation distribution by $\varepsilon_t \sim (0, \sigma^2)$.

We are particularly interested in conditional dynamics characterized by heteroskedasticity, or time-varying volatility. Hence we denote the conditional distribution by $\varepsilon_t \mid \Omega_{t-1} \sim (0, \sigma_t^2)$, where $\Omega_{t-1} = \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$. The conditional variance σ_t^2 will in general evolve as Ω_{t-1} evolves, which focuses attention on the possibility of time-varying innovation volatility.[1]

Allowing for time-varying volatility is crucially important in certain economic and financial contexts. The volatility of financial asset returns, for example, is often time-varying. That is, markets are sometimes tranquil and sometimes turbulent, as can readily be seen by examining the time series of stock market returns in Figure 1, to which we shall return in detail. Time-varying volatility has important implications for financial risk management, asset allocation and asset pricing, and it has therefore become central part of the emerging field of financial econometrics. Quite apart from financial applications, however, time-varying volatility also has direct implications for interval and density forecasting in a wide variety of applications: correct confidence intervals and density forecasts in the presence of volatility fluctuations require time-varying confidence interval widths and time-varying density forecast spreads. The forecasting models that we have considered thus far, however, do not allow for that possibility. In this chapter we do so.

8.1 The Basic ARCH Process

Consider the general linear process,

$$\begin{aligned} y_t &= B(L)\varepsilon_t \\ B(L) &= \sum_{i=0}^{\infty} b_i L^i \end{aligned}$$

[1] In principle, aspects of the conditional distribution other than the variance, such as conditional skewness, could also fluctuate. Conditional variance fluctuations are by far the most important in practice, however, so we assume that fluctuations in the conditional distribution of ε are due exclusively to fluctuations in σ_t^2 .

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

$$b_0 = 1$$

$\varepsilon_t \sim WN(0, \sigma^2)$.

We will work with various cases of this process.

Suppose first that ε_t is strong white noise, $\varepsilon_t \overset{iid}{\sim} WN(0, \sigma^2)$. Let us review some results already discussed for the general linear process, which will prove useful in what follows. The *unconditional* mean and variance of y are

$$E(y_t) = 0$$

and

$$E(y_t^2) = \sigma^2 \sum_{i=0}^{\infty} b_i^2,$$

which are both time-invariant, as must be the case under covariance stationarity. However, the *conditional* mean of y is time-varying:

$$E(y_t|\Omega_{t-1}) = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i},$$

where the information set is

$$\Omega_{t-1} = \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots.$$

The ability of the general linear process to capture covariance stationary conditional mean dynamics is the source of its power.

Because the volatility of many economic time series varies, one would hope that the general linear process could capture conditional variance dynamics as well, but such is not the case for the model as presently specified: the conditional variance of y is constant at

$$E\left((y_t \; - \; E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}\right) \;=\; \sigma^2.$$

This potentially unfortunate restriction manifests itself in the properties of the h-step-ahead conditional prediction error variance. The minimum mean squared error forecast is the conditional mean,

$$E(y_{t+h} \mid \Omega_t) \;=\; \sum_{i=0}^{\infty} b_{h+i}\varepsilon_{t-i} \; ,$$

and so the associated prediction error is

$$y_{t+h} \; - \; E(y_{t+h} \mid \Omega_t) \;=\; \sum_{i=0}^{h-1} b_i\varepsilon_{t+h-i} \; ,$$

which has a conditional prediction error variance of

$$E\left((y_{t+h} \; - \; E(y_{t+h} \mid \Omega_t))^2 \; \mid \Omega_t\right) \;=\; \sigma^2 \sum_{i=0}^{h-1} b_i^2 \; .$$

The conditional prediction error variance is different from the unconditional variance, but it is not time-varying: it depends only on h, not on the conditioning information Ω_t. In the process as presently specified, the conditional variance is not allowed to adapt to readily available and potentially useful conditioning information.

So much for the general linear process with iid innovations. Now we extend it by allowing ε_t to be weak rather than strong white noise, *with a particular nonlinear dependence structure*. In particular, suppose that, as before,

$$y_t \;=\; B(L)\varepsilon_t$$

$$B(L) = \sum_{i=0}^{\infty} b_i L^i$$

$$\sum_{i=0}^{\infty} b_i^2 < \infty$$

$$b_0 = 1 ,$$

but now suppose as well that

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \gamma(L)\varepsilon_t^2$$

$$\omega > 0 \quad \gamma(L) = \sum_{i=1}^{p} \gamma_i L^i \quad \gamma_i \geq 0 \text{ for all i} \quad \sum \gamma_i < 1 .$$

Note that we parameterize the innovation process in terms of its conditional density,

$$\varepsilon_t \mid \Omega_{t-1},$$

which we assume to be normal with a zero conditional mean and a conditional variance that depends linearly on p past squared innovations. ε_t is serially uncorrelated but not serially independent, because the current conditional variance σ_t^2 depends on the history of ε_t .[2] The stated regularity conditions are sufficient to ensure that the conditional and unconditional variances are positive and finite, and that y_t is covariance stationary.
The unconditional moments of ε_t are constant and are given by

$$E(\varepsilon_t) = 0$$

[2] In particular, σ_t^2 depends on the previous p values of ε_t via the distributed lag

$$\gamma(L)\varepsilon_t^2.$$

and

$$E(\varepsilon_t - E(\varepsilon_t))^2 = \frac{\omega}{1 - \sum \gamma_i} .$$

The important result is not the particular formulae for the unconditional mean and variance, but the fact that they are fixed, as required for covariance stationarity. As for the conditional moments of ε_t , its conditional variance is time-varying,

$$E\left((\varepsilon_t - E(\varepsilon_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}\right) = \omega + \gamma(L)\varepsilon_t^2 ,$$

and of course its conditional mean is zero by construction.

Assembling the results to move to the unconditional and conditional moments of y as opposed to ε_t , it is easy to see that both the unconditional mean and variance of y are constant (again, as required by covariance stationarity), but that both the conditional mean and variance are time-varying:

$$E(y_t \mid \Omega_{t-1}) = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i}$$

$$E\left((y_t - E(y_t \mid \Omega_{t-1}))^2 \mid \Omega_{t-1}\right) = \omega + \gamma(L)\varepsilon_t^2 .$$

Thus, we now treat conditional mean and variance dynamics in a symmetric fashion by allowing for movement in each, as determined by the evolving information set Ω_{t-1} . In the above development, ε_t is called an ARCH(p) process, and the full model sketched is an infinite-ordered moving average with ARCH(p) innovations, where ARCH stands for autoregressive conditional heteroskedasticity. Clearly ε_t is conditionally heteroskedastic, because its conditional variance fluctuates. There are many models of conditional heteroskedasticity, but most are designed for cross-sectional contexts, such as when the variance of a cross-sectional regression disturbance depends on one or more of the regressors.[3] However, heteroskedasticity is often present as well in the time-series contexts relevant for forecasting, particularly in financial markets. The particular conditional variance function associated with the ARCH process,

$$\sigma_t^2 = \omega + \gamma(L)\varepsilon_t^2 ,$$

is tailor-made for time-series environments, in which one often sees volatility clustering, such that large changes tend to be followed by large changes, and small by small, *of either sign*. That is, one may see persistence, or serial correlation, in volatility dynamics (conditional variance dynamics), quite apart from persistence (or lack thereof) in conditional mean dynamics. The ARCH process approximates volatility dynamics in an autoregressive fashion; hence the name *autoregressive* conditional heteroskedasticity. To understand why, note that the ARCH conditional variance function links today's conditional variance positively to earlier lagged ε_t^2 's, so that large ε_t^2 's in the recent past produce a large conditional variance today, thereby increasing the likelihood of a large ε_t^2 today. Hence ARCH processes are to conditional variance dynamics precisely as standard autoregressive processes are to conditional mean dynamics. The ARCH process may be viewed as a model for the disturbance in a broader model, as was the case when we introduced it above as a model for the innovation in a general linear process. Alternatively, if there are no conditional mean dynamics of interest, the ARCH process may be used for an observed series. It turns out that financial asset returns often have negligible conditional mean dynamics but strong conditional variance dynamics; hence in much of what follows we will view the ARCH process as a model for an observed series, which for convenience we will sometimes call a "return."

[3] The variance of the disturbance in a model of household expenditure, for example, may depend on income.

8.2 The GARCH Process

Thus far we have used an ARCH(p) process to model conditional variance dynamics. We now introduce the GARCH(p,q) process (GARCH stands for generalized ARCH), which we shall subsequently use almost exclusively. As we shall see, GARCH is to ARCH (for conditional variance dynamics) as ARMA is to AR (for conditional mean dynamics).

The pure GARCH(p,q) process is given by[4]

$$y_t = \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2$$

$$\alpha(L) = \sum_{i=1}^{p} \alpha_i L^i, \quad \beta(L) = \sum_{i=1}^{q} \beta_i L^i$$

$$\omega > 0, \ \alpha_i \geq 0, \ \beta_i \geq 0, \ \sum \alpha_i + \sum \beta_i < 1 .$$

The stated conditions ensure that the conditional variance is positive and that y_t is covariance stationary.

Back substitution on σ_t^2 reveals that the GARCH(p,q) process can be represented as a restricted infinite-ordered ARCH process,

$$\sigma_t^2 = \frac{\omega}{1 - \sum \beta_i} + \frac{\alpha(L)}{1 - \beta(L)}\varepsilon_t^2 = \frac{\omega}{1 - \sum \beta_i} + \sum_{i=1}^{\infty} \delta_i \varepsilon_{t-i}^2 ,$$

which precisely parallels writing an ARMA process as a restricted infinite-ordered AR. Hence the GARCH(p,q) process is a parsimonious approximation to what may truly be infinite-ordered ARCH volatility dynamics.

[4] By "pure" we mean that we have allowed only for conditional variance dynamics, by setting $y_t = \varepsilon_t$. We could of course also introduce conditional mean dynamics, but doing so would only clutter the discussion while adding nothing new.

It is important to note a number of special cases of the GARCH(p,q) process. First, of course, the ARCH(p) process emerges when

$$\beta(\mathrm{L}) \;=\; 0.$$

Second, if *both* $\alpha(\mathrm{L})$ and $\beta(\mathrm{L})$ are zero, then the process is simply iid Gaussian noise with variance ω. Hence, although ARCH and GARCH processes may at first appear unfamiliar and potentially ad hoc, they are in fact much more general than standard iid white noise, which emerges as a potentially highly-restrictive special case.

Here we highlight some important properties of GARCH processes. All of the discussion of course applies as well to ARCH processes, which are special cases of GARCH processes. First, consider the second-order moment structure of GARCH processes. The first two unconditional moments of the pure GARCH process are constant and given by

$$\mathrm{E}(\varepsilon_{\mathrm{t}}) \;=\; 0$$

and

$$\mathrm{E}(\varepsilon_{\mathrm{t}} \;-\; \mathrm{E}(\varepsilon_{\mathrm{t}}))^2 \;=\; \frac{\omega}{1 \;-\; \sum \alpha_{\mathrm{i}} \;-\; \sum \beta_{\mathrm{i}}}\;,$$

while the conditional moments are

$$\mathrm{E}(\varepsilon_{\mathrm{t}} \mid \Omega_{\mathrm{t-1}}) \;=\; 0$$

and of course

$$\mathrm{E}\left((\varepsilon_{\mathrm{t}} \;-\; \mathrm{E}(\varepsilon_{\mathrm{t}}|\Omega_{\mathrm{t-1}}))^2|\Omega_{\mathrm{t-1}}\right) \;=\; \omega \;+\; \alpha(\mathrm{L})\varepsilon_{\mathrm{t}}^2 \;+\; \beta(\mathrm{L})\sigma_{\mathrm{t}}^2\;.$$

In particular, the unconditional variance is fixed, as must be the case under covariance stationarity, while the conditional variance is time-varying. It is no *surprise* that the conditional variance is time-varying – the GARCH process was of course *designed* to allow for a time-varying conditional variance – but it is certainly worth emphasizing: the conditional variance is itself a serially correlated time series process.

Second, consider the unconditional higher-order (third and fourth) moment structure of GARCH processes. Real-world financial asset returns, which are often modeled as GARCH processes, are typically unconditionally symmetric but leptokurtic (that is, more peaked in the center and with fatter tails than a normal distribution). It turns out that the implied unconditional distribution of the conditionally Gaussian GARCH process introduced above is also symmetric and leptokurtic. The unconditional leptokurtosis of GARCH processes follows from the persistence in conditional variance, which produces clusters of "low volatility" and "high volatility" episodes associated with observations in the center and in the tails of the unconditional distribution, respectively. Both the unconditional symmetry and unconditional leptokurtosis agree nicely with a variety of financial market data.

Third, consider the conditional prediction error variance of a GARCH process, and its dependence on the conditioning information set. Because the conditional variance of a GARCH process is a serially correlated random variable, it is of interest to examine the optimal h-step-ahead prediction, prediction error, and conditional prediction error variance. Immediately, the h-step-ahead prediction is

$$E(\varepsilon_{t+h} \mid \Omega_t) = 0,$$

and the corresponding prediction error is

$$\varepsilon_{t+h} - E(\varepsilon_{t+h} \mid \Omega_t) = \varepsilon_{t+h} .$$

This implies that the conditional variance of the prediction error,

$$E\left((\varepsilon_{t+h} - E(\varepsilon_{t+h} \mid \Omega_t))^2 \mid \Omega_t\right) = E(\varepsilon_{t+h}^2 \mid \Omega_t) ,$$

depends on both h *and*

$$\Omega_t,$$

because of the dynamics in the conditional variance. Simple calculations reveal that the expression for the GARCH(p, q) process is given by

$$E(\varepsilon_{t+h}^2 \mid \Omega_t) = \omega \left(\sum_{i=0}^{h-2} (\alpha(1) + \beta(1))^i \right) + (\alpha(1) + \beta(1))^{h-1} \sigma_{t+1}^2 .$$

In the limit, this conditional variance reduces to the unconditional variance of the process,

$$\lim_{h \to \infty} E(\varepsilon_{t+h}^2 \mid \Omega_t) = \frac{\omega}{1 - \alpha(1) - \beta(1)} .$$

For finite h, the dependence of the prediction error variance on the current information set Ω_t can be exploited to improve interval and density forecasts.

Fourth, consider the relationship between ε_t^2 and σ_t^2 . The relationship is important: GARCH dynamics in σ_t^2 turn out to introduce ARMA dynamics in ε_t^2 .[5] More precisely, if ε_t is a GARCH(p,q) process, then

$$\varepsilon_t^2$$

has the ARMA representation

$$\varepsilon_t^2 = \omega + (\alpha(L) + \beta(L))\varepsilon_t^2 - \beta(L)\nu_t + \nu_t ,$$

where

$$\nu_t = \varepsilon_t^2 - \sigma_t^2$$

[5] Put differently, the GARCH process approximates conditional variance dynamics in the same way that an ARMA process approximates conditional mean dynamics.

is the difference between the squared innovation and the conditional variance at time t. To see this, note that if ε_t is GARCH(p,q), then

$$\sigma_t^2 \;=\; \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2.$$

Adding and subtracting

$$\beta(L)\varepsilon_t^2$$

from the right side gives

$$\begin{aligned}\sigma_t^2 \;&=\; \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\varepsilon_t^2 - \beta(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \\ &=\; \omega \;+\; (\alpha(L) \;+\; \beta(L))\varepsilon_t^2 \;-\; \beta(L)(\varepsilon_t^2 \;-\; \sigma_t^2)\;.\end{aligned}$$

Adding

$$\varepsilon_t^2$$

to each side then gives

$$\sigma_t^2 \;+\; \varepsilon_t^2 \;=\; \omega \;+\; (\alpha(L) \;+\; \beta(L))\varepsilon_t^2 \;-\; \beta(L)(\varepsilon_t^2 \;-\; \sigma_t^2) \;+\; \varepsilon_t^2\;,$$

so that

$$\begin{aligned}\varepsilon_t^2 \;&=\; \omega \;+\; (\alpha(L) \;+\; \beta(L))\varepsilon_t^2 \;-\; \beta(L)(\varepsilon_t^2 \;-\; \sigma_t^2) \;+\; (\varepsilon_t^2 \;-\; \sigma_t^2)\;, \\ &=\; \omega \;+\; (\alpha(L) \;+\; \beta(L))\varepsilon_t^2 \;-\; \beta(L)\nu_t \;+\; \nu_t\;.\end{aligned}$$

Thus,

$$\varepsilon_t^2$$

is an ARMA((max(p,q)), p) process with innovation ν_t , where

$$\nu_t \;\in\; [-\sigma_t^2,\; \infty).$$

ε_t^2 is covariance stationary if the roots of $\alpha(L)+\beta(L)=1$ are outside the unit circle.

Fifth, consider in greater depth the similarities and differences between σ_{t}^2 and

$$\varepsilon_{\mathrm{t}}^2.$$

It is worth studying closely the key expression,

$$\nu_{\mathrm{t}} \; = \; \varepsilon_{\mathrm{t}}^2 - \sigma_{\mathrm{t}}^2,$$

which makes clear that

$$\varepsilon_{\mathrm{t}}^2$$

is effectively a "proxy" for σ_{t}^2 , behaving similarly but not identically, with ν_{t} being the difference, or error. In particular, ε_t^2 is a *noisy* proxy: ε_t^2 is an unbiased estimator of σ_t^2 , but it is more volatile. It seems reasonable, then, that reconciling the noisy proxy ε_t^2 and the true underlying σ_t^2 should involve some sort of smoothing of ε_t^2. Indeed, in the GARCH(1,1) case σ_t^2 is precisely obtained by exponentially smoothing ε_t^2. To see why, consider the exponential smoothing recursion, which gives the current smoothed value as a convex combination of the current unsmoothed value and the lagged smoothed value,

$$\bar{\varepsilon}_{\mathrm{t}}^2 \; = \; \gamma\varepsilon_{\mathrm{t}}^2 \; + \; (1-\gamma)\bar{\varepsilon}_{\mathrm{t-1}}^2 \; .$$

Back substitution yields an expression for the current smoothed value as an exponentially weighted moving average of past actual values:

$$\bar{\varepsilon}_{\mathrm{t}}^2 \; = \; \sum \mathrm{w_j}\varepsilon_{\mathrm{t-j}}^2 \; ,$$

where

$$\mathrm{w_j} \; = \; \gamma(1-\gamma)^{\mathrm{j}} \; .$$

Now compare this result to the GARCH(1,1) model, which gives the current volatility as a linear combination of lagged volatility and the lagged squared return, $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$.

Back substitution yields $\sigma_t^2 = \frac{\omega}{1-\beta} + \alpha \sum \beta^{j-1}\varepsilon_{t-j}^2$, so that the GARCH(1,1) process gives current volatility as an exponentially weighted moving average of past squared returns.

Sixth, consider the temporal aggregation of GARCH processes. By temporal aggregation we mean aggregation over time, as for example when we convert a series of daily returns to weekly returns, and then to monthly returns, then quarterly, and so on. It turns out that convergence toward normality under temporal aggregation is a feature of real-world financial asset returns. That is, although high-frequency (e.g., daily) returns tend to be fat-tailed relative to the normal, the fat tails tend to get thinner under temporal aggregation, and normality is approached. Convergence to normality under temporal aggregation is also a property of covariance stationary GARCH processes. The key insight is that a low-frequency change is simply the sum of the corresponding high-frequency changes; for example, an annual change is the sum of the internal quarterly changes, each of which is the sum of its internal monthly changes, and so on. Thus, if a Gaussian central limit theorem can be invoked for sums of GARCH processes, convergence to normality under temporal aggregation is assured. Such theorems can be invoked if the process is covariance stationary.

In closing this section, it is worth noting that the symmetry and leptokurtosis of the unconditional distribution of the GARCH process, as well as the disappearance of the leptokurtosis under temporal aggregation, provide nice independent confirmation of the accuracy of GARCH approximations to asset return volatility dynamics, insofar as GARCH was certainly not invented with the intent of explaining those features of financial asset return data. On the contrary, the unconditional distributional results emerged as unanticipated byproducts of allowing for conditional variance dynamics, thereby providing a unified explanation of phenomena that were previously believed unrelated.

8.3 Extensions of ARCH and GARCH Models

There are numerous extensions of the basic GARCH model. In this section, we highlight several of the most important. One important class of extensions allows for asymmetric response; that is, it allows for last period's squared return to have different effects on today's volatility, depending on its sign.[6] Asymmetric response is often present, for example, in stock returns.

8.3.1 Asymmetric Response

The simplest GARCH model allowing for asymmetric response is the threshold GARCH, or TGARCH, model.[7] We replace the standard GARCH conditional variance function, $\sigma_{\mathrm{t}}^2 = \omega + \alpha\varepsilon_{\mathrm{t-1}}^2 + \beta\sigma_{\mathrm{t-1}}^2$, with $\sigma_{\mathrm{t}}^2 = \omega + \alpha\varepsilon_{\mathrm{t-1}}^2 + \gamma\varepsilon_{\mathrm{t-1}}^2\mathrm{D}$

where $\mathrm{D_t} = \begin{array}{l} 1,\ if\ \varepsilon_t < 0 \\ 0\ otherwise\ . \end{array}$.

The dummy variable D keeps track of whether the lagged return is positive or negative. When the lagged return is positive (good news yesterday), D=0, so the effect of the lagged squared return on the current conditional variance is simply α. In contrast, when the lagged return is negative (bad news yesterday), D=1, so the effect of the lagged squared return on the current conditional variance is $\alpha+\gamma$. If $\gamma = 0$, the response is symmetric and we have a standard GARCH model, but if $\gamma \neq 0$ we have asymmetric response of volatility to news. Allowance for asymmetric response has proved useful for modeling "leverage effects" in stock returns, which occur when γ <0.[8]

[6] In the GARCH model studied thus far, only the *square* of last period's return affects the current conditional variance; hence its sign is irrelevant.

[7] For expositional convenience, we will introduce all GARCH extensions in the context of GARCH(1,1), which is by far the most important case for practical applications. Extensions to the GARCH(p,q) case are immediate but notationally cumbersome.

[8] Negative shocks appear to contribute more to stock market volatility than do positive shocks. This is called the leverage effect, because a negative shock to the market value of equity increases the aggregate debt/equity ratio (other things the same), thereby increasing leverage.

Asymmetric response may also be introduced via the exponential GARCH (EGARCH) model,

$$\ln(\sigma_t^2) = \omega + \alpha \left| \varepsilon_{\frac{t-1}{\sigma_{t-1}}} \right| + \gamma \varepsilon_{\frac{t-1}{\sigma_{t-1}}} + \beta \ln(\sigma_{t-1}^2) .$$

Note that volatility is driven by both size and sign of shocks; hence the model allows for an asymmetric response depending on the sign of news.[9] The log specification also ensures that the conditional variance is automatically positive, because σ_t^2 is obtained by exponentiating $\ln(\sigma_t^2)$; hence the name "exponential GARCH."

8.3.2 Exogenous Variables in the Volatility Function

Just as ARMA models of conditional mean dynamics can be augmented to include the effects of exogenous variables, so too can GARCH models of conditional variance dynamics.

We simply modify the standard GARCH volatility function in the obvious way, writing

$$\sigma_t^2 = \omega + \alpha \, \varepsilon_{t-1}^2 + \beta \, \sigma_{t-1}^2 + \gamma x_t ,$$

where γ is a parameter and x is a positive exogenous variable.[10] Allowance for exogenous variables in the conditional variance function is sometimes useful. Financial market volume, for example, often helps to explain market volatility.

[9] The absolute "size" of news is captured by $|r_{t-1}/\sigma_{t-1}|$, and the sign is captured by r_{t-1}/σ_{t-1} .

[10] Extension to allow multiple exogenous variables is straightforward.

8.3.3 Regression with GARCH disturbances and GARCH-M

Just as ARMA models may be viewed as models for disturbances in regressions, so too may GARCH models. We write

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. Consider now a regression model with GARCH disturbances of the usual sort, with one additional twist: the conditional variance enters as a regressor, thereby affecting the conditional mean. We write

$$y_t = \beta_0 + \beta_1 x_t + \gamma\sigma_t^2 + \varepsilon_t$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$. This model, which is a special case of the general regression model with GARCH disturbances, is called GARCH-in-Mean (GARCH-M). It is sometimes useful in modeling the relationship between risks and returns on financial assets when risk, as measured by the conditional variance, varies.[11]

8.3.4 Component GARCH

Note that the standard GARCH(1,1) process may be written as $(\sigma_t^2 - \bar{\omega}) = \alpha(\varepsilon_{t-1}^2 - \bar{\omega})$ where $\bar{\omega} = \frac{\omega}{1-\alpha-\beta}$ is the unconditional variance.[12] This is precisely the GARCH(1,1) model introduced earlier, rewritten it in a slightly different but equivalent form. In this model, short-run volatility dynamics are governed by the parameters α and β, and there are no long-run volatility dynamics, because $\bar{\omega}$ is constant. Sometimes we might want to allow for both long-run and

[11] One may also allow the conditional standard deviation, rather than the conditional variance, to enter the regression.

[12] $\bar{\omega}$ is sometimes called the "long-run" variance, referring to the fact that the unconditional variance is the long-run average of the conditional variance.

short-run, or persistent and transient, volatility dynamics in addition to the short-run volatility dynamics already incorporated. To do this, we replace $\bar{\omega}$ with a time-varying process, yielding $(\sigma_t^2 - q_t) \; = \; \alpha(\varepsilon_{t-1}^2 - q_{t-1}) \; + \; \beta(\sigma_{t-1}^2 - q_{t-1})$, where the time-varying long-run volatility, q_t , is given by $q_t \; = \; \omega \; + \; \rho(q_{t-1} - \omega) \; + \; \phi(\varepsilon_t$ This "component GARCH" model effectively lets us decompose volatility dynamics into long-run (persistent) and short-run (transitory) components, which sometimes yields useful insights. The persistent dynamics are governed by ρ , and the transitory dynamics are governed by α and β .[13]

8.3.5 Mixing and Matching

In closing this section, we note that the different variations and extensions of the GARCH process may of course be mixed. As an example, consider the following conditional variance function: $(\sigma_t^2 - q_t) \; = \; \alpha(\varepsilon_{t-1}^2 - q_{t-1}) \; + \; \gamma(\varepsilon_{t-1}^2 - q_{t-1})D_{t-1}$ This is a component GARCH specification, generalized to allow for asymmetric response of volatility to news via the sign dummy D, as well as effects from the exogenous variable x.

8.4 Estimating, Forecasting and Diagnosing GARCH Models

Recall that the likelihood function is the joint density function of the data, viewed as a function of the model parameters, and that maximum likelihood estimation finds the parameter values that maximize the likelihood function. This makes good sense: we choose those parameter values that maximize the likelihood of obtaining the data that were actually obtained. It turns

[13] It turns out, moreover, that under suitable conditions the component GARCH model introduced here is covariance stationary, and equivalent to a GARCH(2,2) process subject to certain nonlinear restrictions on its parameters.

out that construction and evaluation of the likelihood function is easily done for GARCH models, and maximum likelihood has emerged as the estimation method of choice.[14] No closed-form expression exists for the GARCH maximum likelihood estimator, so we must maximize the likelihood numerically.[15] Construction of optimal forecasts of GARCH processes is simple. In fact, we derived the key formula earlier but did not comment extensively on it. Recall, in particular, that

$$\sigma^2_{t+h,t} \;=\; E\left[\varepsilon^2_{t+h} \mid \Omega_t\right] \;=\; \omega\left(\sum_{i=1}^{h-1}\left[\alpha(1) \;+\; \beta(1)\right]^i\right) \;+\; \left[\alpha(1) \;+\; \beta(1)\right]^{h-1}\sigma^2_{t+1}\;.$$

In words, the optimal h-step-ahead forecast is proportional to the optimal 1-step-ahead forecast. The optimal 1-step-ahead forecast, moreover, is easily calculated: all of the determinants of σ^2_{t+1} are lagged by at least one period, so that there is no problem of forecasting the right-hand side variables. In practice, of course, the underlying GARCH parameters α and β are unknown and so must be estimated, resulting in the feasible forecast $\hat{\sigma}^2_{t+h,t}$ formed in the obvious way. In financial applications, volatility forecasts are often of direct interest, and the GARCH model delivers the optimal h-step-ahead point forecast, $\sigma^2_{t+h,t}$. Alternatively, and more generally, we might not be intrinsically interested in volatility; rather, we may simply want to use GARCH volatility forecasts to improve h-step-ahead interval or density forecasts of ε_t , which are crucially dependent on the h-step-ahead prediction error variance, $\sigma^2_{t+h,t}$. Consider, for example, the case of interval forecasting. In the case of constant volatility, we earlier worked with Gaussian ninety-five percent

[14] The precise form of the likelihood is complicated, and we will not give an explicit expression here, but it may be found in various of the surveys mentioned in the Bibliographical and Computational Notes at the end of the chapter.

[15] Routines for maximizing the GARCH likelihood are available in a number of modern software packages such as Eviews. As with any numerical optimization, care must be taken with startup values and convergence criteria to help insure convergence to a global, as opposed to merely local, maximum.

interval forecasts of the form

$$y_{t+h,t} \pm 1.96\sigma_h ,$$

where σ_h denotes the unconditional h-step-ahead standard deviation (which also equals the conditional h-step-ahead standard deviation in the absence of volatility dynamics). Now, however, in the presence of volatility dynamics we use

$$y_{t+h,t} \pm 1.96\sigma_{t+h,t} .$$

The ability of the conditional prediction interval to adapt to changes in volatility is natural and desirable: when volatility is low, the intervals are naturally tighter, and conversely. In the presence of volatility dynamics, the unconditional interval forecast is correct on average but likely incorrect at any given time, whereas the conditional interval forecast is correct at all times. The issue arises as to how to detect GARCH effects in observed returns, and related, how to assess the adequacy of a fitted GARCH model. A key and simple device is the correlogram of squared returns, ε_t^2 . As discussed earlier, ε_t^2 is a proxy for the latent conditional variance; if the conditional variance displays persistence, so too will ε_t^2 .[16] Once can of course also fit a GARCH model, and assess significance of the GARCH coefficients in the usual way.

Note that we can write the GARCH process for returns as $\varepsilon_t = \sigma_t v_t$, where $v_t \overset{iid}{\sim} N(0, 1)$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$.

Equivalently, the *standardized* return, v, is iid,

[16] Note well, however, that the converse is not true. That is, if ε_t^2 displays persistence, it does not necessarily follow that the conditional variance displays persistence. In particular, neglected serial correlation associated with conditional mean dynamics may cause serial correlation in ε_t and hence also in ε_t^2 . Thus, before proceeding to examine and interpret the correlogram of ε_t^2 as a check for volatility dynamics, it is important that any conditional mean effects be appropriately modeled, in which case ε_t should be interpreted as the disturbance in an appropriate conditional mean model.

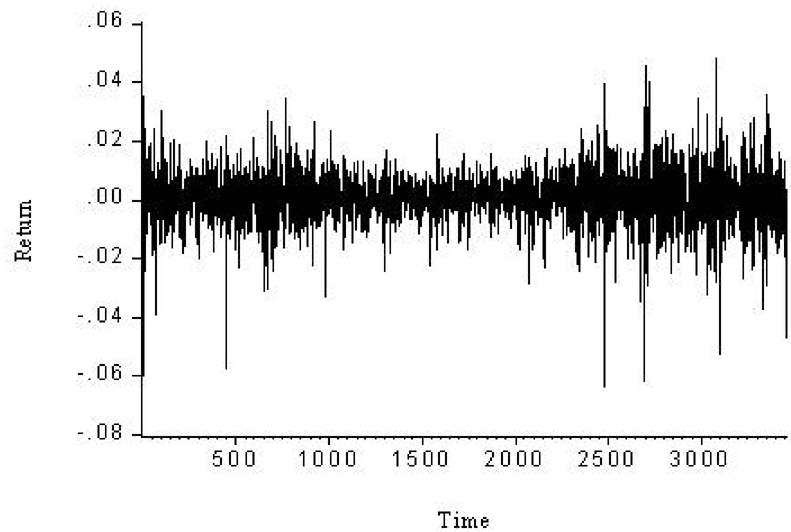

Figure 8.1: NYSE Returns

$$\frac{\varepsilon_t}{\sigma_t} = v_t \overset{iid}{\sim} N(0,\ 1)\ .$$

This observation suggests a way to evaluate the adequacy of a fitted GARCH model: standardize returns by the conditional standard deviation from the fitted GARCH model, $\hat{\sigma}$, and then check for volatility dynamics missed by the fitted model by examining the correlogram of the squared *standardized* return, $(\varepsilon_t/\hat{\sigma}_t)^2$. This is routinely done in practice.

8.5 Application: Stock Market Volatility

We model and forecast the volatility of daily returns on the New York Stock Exchange (NYSE) from January 1, 1988 through December 31, 2001, excluding holidays, for a total of 3531 observations. We estimate using observations 1-3461, and then we forecast observations 3462-3531.

In Figure 8.1 we plot the daily returns, r_t . There is no visual evidence of serial correlation in the returns, but there *is* evidence of serial correlation in the *amplitude* of the returns. That is, volatility appears to cluster: large

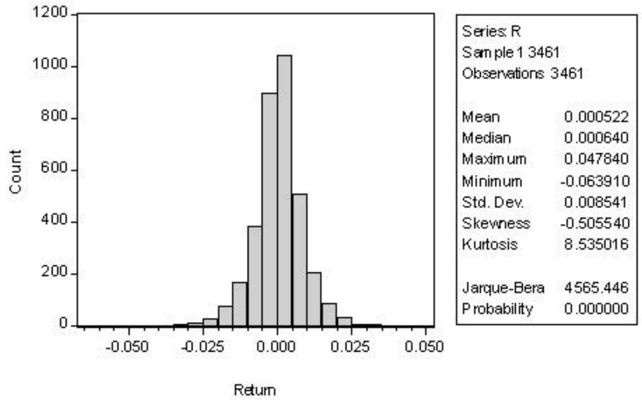

Figure 8.2: Histogram of NYSE Returns

changes tend to be followed by large changes, and small by small, *of either sign*. In Figure 8.2 we show the histogram and related statistics for r_t . The mean daily return is slightly positive. Moreover, the returns are approximately symmetric (only slightly left skewed) but highly leptokurtic. The Jarque-Bera statistic indicates decisive rejection of normality.

In Figure 8.3 we show the correlogram for r_t . The sample autocorrelations are tiny and usually insignificant relative to the Bartlett standard errors, yet the autocorrelation function shows some evidence of a systematic cyclical pattern, and the Q statistics (not shown), which cumulate the information across all displacements, reject the null of weak white noise. Despite the weak serial correlation evidently present in the returns, we will proceed for now as if returns were weak white noise, which is approximately, if not exactly, the case.[17]

In Figure 8.4 we plot r_t^2 . The volatility clustering is even more evident than it was in the time series plot of returns. Perhaps the strongest evidence of all comes from the correlogram of r_t^2 , which we show in Figure 8.5: all sample autocorrelations of r_t^2 are positive, overwhelmingly larger than those of the returns themselves, and statistically significant. As a crude first pass at modeling the stock market volatility, we fit an AR(5) model directly to r_t^2

[17] In the Exercises, Problems and Complements at the end of this chapter we model the conditional mean, as well as the conditional variance, of returns.

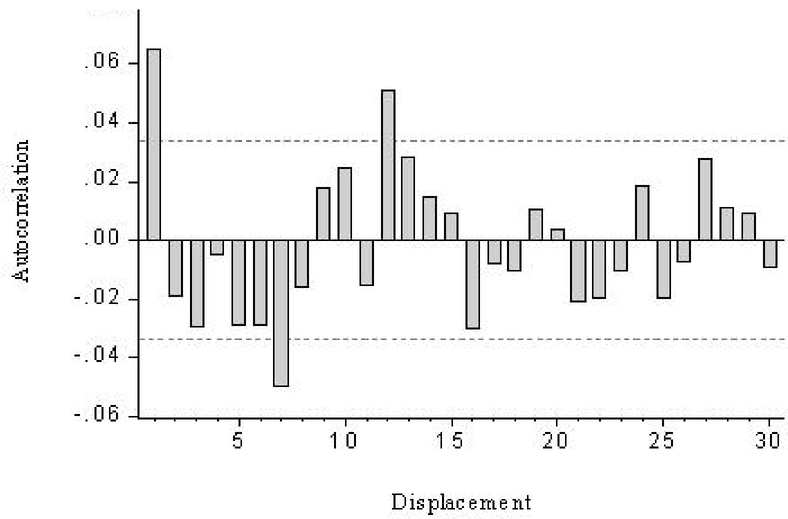

Figure 8.3: Correlogram of NYSE Returns

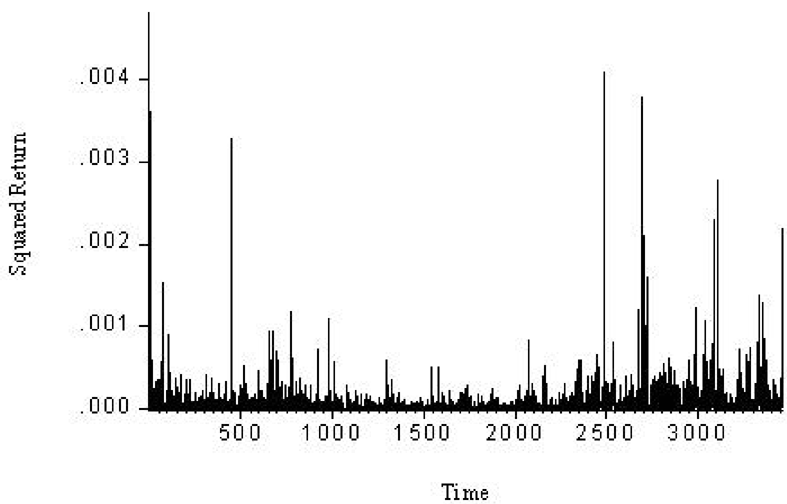

Figure 8.4: Squared NYSE Returns

; the results appear in Table 8.6. It is interesting to note that the t-statistics on the lagged squared returns are often significant, even at long lags, yet the R^2 of the regression is low, reflecting the fact that r_t^2 is a very noisy volatility proxy. As a more sophisticated second pass at modeling NYSE volatility, we fit an ARCH(5) model to r_t ; the results appear in Table 8.7. The lagged squared returns appear significant even at long lags. The correlogram of squared standardized residuals shown in Figure 8.8, however, displays some remaining systematic behavior, indicating that the ARCH(5) model fails to

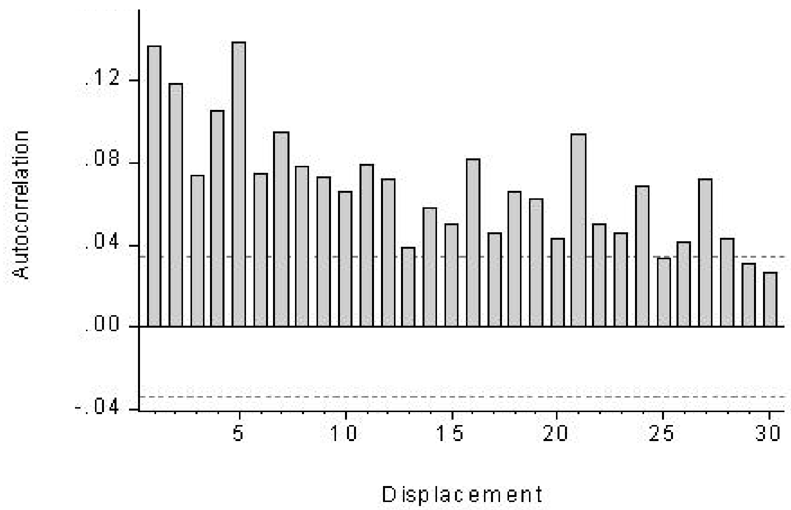

Figure 8.5: Squared NYSE Returns Correlogram

capture all of the volatility dynamics, potentially because even longer lags are needed.[18]

[18] In the Exercises, Problems and Complements at the end of this chapter we also examine ARCH(p) models with p>5.

Dependent Variable: R2
Method: Least Squares

Sample(adjusted): 6 3461
Included observations: 3456 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.40E-05	3.78E-06	11.62473	0.0000
R2(-1)	0.107900	0.016137	6.686547	0.0000
R2(-2)	0.091840	0.016186	5.674167	0.0000
R2(-3)	0.028981	0.016250	1.783389	0.0746
R2(-4)	0.039312	0.016481	2.385241	0.0171
R2(-5)	0.116436	0.016338	7.126828	0.0000

R-squared	0.052268	Mean dependent var	7.19E-05
Adjusted R-squared	0.050894	S.D. dependent var	0.000189
S.E. of regression	0.000184	Akaike info criterion	-14.36434
Sum squared resid	0.000116	Schwarz criterion	-14.35366
Log likelihood	24827.58	F-statistic	38.05372
Durbin-Watson stat	1.975672	Prob(F-statistic)	0.000000

Figure 8.6: Squared NYSE Returns, AR(5) Model

Dependent Variable: R
Method: ML - ARCH (Marquardt)

Sample: 1 3461
Included observations: 3461
Convergence achieved after 13 iterations
Variance backcast: ON

Coefficient	Std. Error	z-Statistic	Prob.	
C	0.000689	0.000127	5.437097	0.0000
Variance Equation				
C	3.16E-05	1.08E-06	29.28536	0.0000
ARCH(1)	0.128948	0.013847	9.312344	0.0000
ARCH(2)	0.166852	0.015055	11.08281	0.0000
ARCH(3)	0.072551	0.014345	5.057526	0.0000
ARCH(4)	0.143778	0.015363	9.358870	0.0000
ARCH(5)	0.089254	0.018480	4.829789	0.0000

R-squared	-0.000381	Mean dependent var	0.000522
Adjusted R-squared	-0.002118	S.D. dependent var	0.008541
S.E. of regression	0.008550	Akaike info criterion	-6.821461
Sum squared resid	0.252519	Schwarz criterion	-6.809024
Log likelihood	11811.54	Durbin-Watson stat	1.861036

Figure 8.7: Squared NYSE Returns, ARCH(5) Model

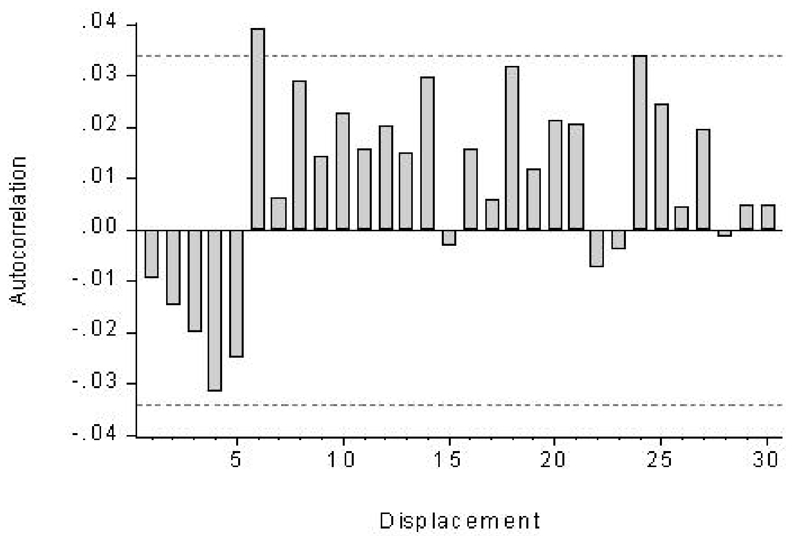

Figure 8.8: NYSE Returns, Correlogram of Squared Standardized Residuals from ARCH(5)

In Table 8.9 we show the results of fitting a GARCH(1,1) model. All of the parameter estimates are highly statistically significant, and the "ARCH coefficient" (α) and "GARCH coefficient" (β) sum to a value near unity (.987), with β substantially larger then α, as is commonly found for financial asset returns. We show the correlogram of squared standardized GARCH(1,1) residuals in Figure 8.10. All sample autocorrelations are tiny and inside the Bartlett bands, and they display noticeably less evidence of any systematic pattern than for the squared standardized ARCH(5) residuals. In Figure 8.11 we show the time series of estimated conditional standard deviations implied by the estimated GARCH(1,1) model. Clearly, volatility fluctuates a great deal and is highly persistent. For comparison we show in Figure 8.12 the series of exponentially smoothed r_t^2 , computed using a standard smoothing parameter of .05.[19] Clearly the GARCH and exponential smoothing volatility estimates behave similarly, although not at all identically. The difference reflects the fact that the GARCH smoothing parameter is effectively estimated by the method of maximum likelihood, whereas the exponential smoothing parameter is set rather arbitrarily. Now, using the model estimated using observations 1-3461, we generate a forecast of the conditional standard deviation for the out-of-sample observations 3462-3531. We show the results in Figure 8.13. The forecast period begins just following a volatility burst, so it is not surprising that the forecast calls for gradual volatility reduction. For greater understanding, in Figure 8.14 we show both a longer history and a longer forecast. Clearly the forecast conditional standard deviation is reverting exponentially to the unconditional standard deviation (.009), per the formula discussed earlier.

[19] For comparability with the earlier-computed GARCH estimated conditional standard deviation, we actually show the square root of exponentially smoothed r_t^2 .

Dependent Variable: R
Method: ML - ARCH (Marquardt)

Sample: 1 3461
Included observations: 3461
Convergence achieved after 19 iterations
Variance backcast: ON

Coefficient	Std. Error	z-Statistic	Prob.	
C	0.000640	0.000127	5.036942	0.0000
Variance Equation				
C	1.06E-06	1.49E-07	7.136840	0.0000
ARCH(1)	0.067410	0.004955	13.60315	0.0000
GARCH(1)	0.919714	0.006122	150.2195	0.0000
R-squared	-0.000191		Mean dependent var	0.000522
Adjusted R-squared	-0.001059		S.D. dependent var	0.008541
S.E. of regression	0.008546		Akaike info criterion	-6.868008
Sum squared resid	0.252471		Schwarz criterion	-6.860901
Log likelihood	11889.09		Durbin-Watson stat	1.861389

Figure 8.9: Squared NYSE Returns, GARCH(1,1)

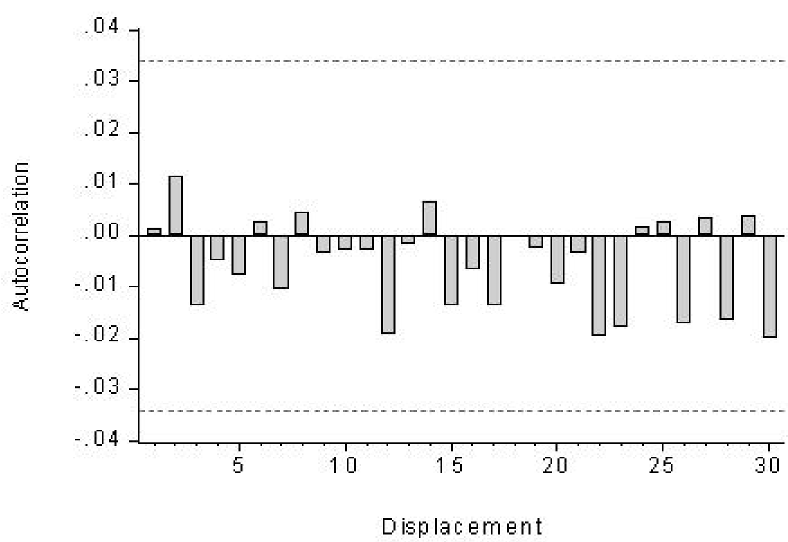

Figure 8.10: NYSE Returns, Correlogram of Squared Standardized Residuals from GARCH(1,1)

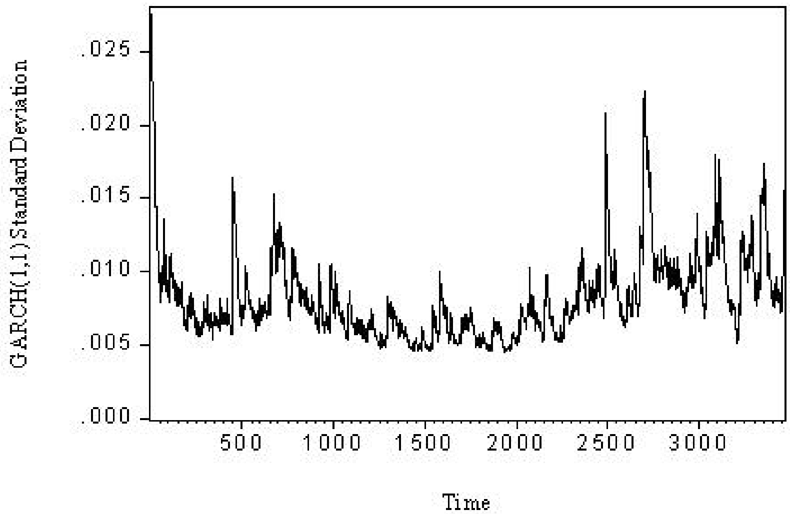

Figure 8.11: Estimated Conditional Standard Deviations from GARCH(1,1)

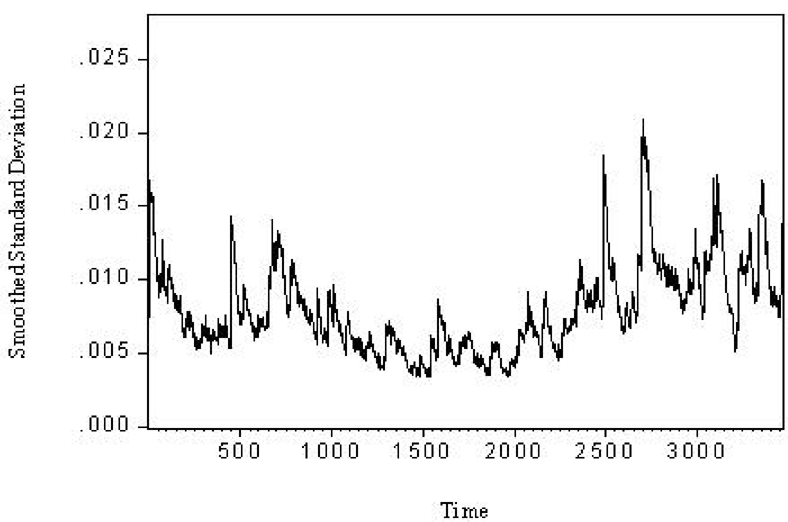

Figure 8.12: Estimated Conditional Standard Deviations - Exponential Smoothing

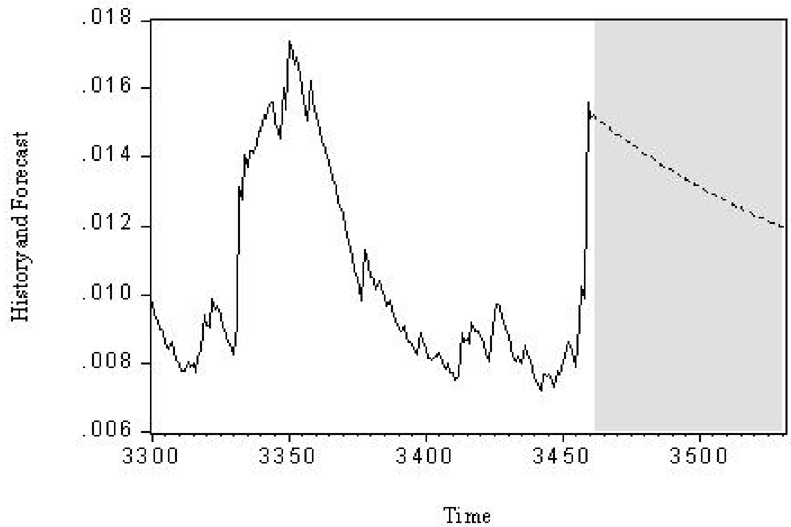

Figure 8.13: Conditional Standard Deviations, History and Forecast from GARCH(1,1)

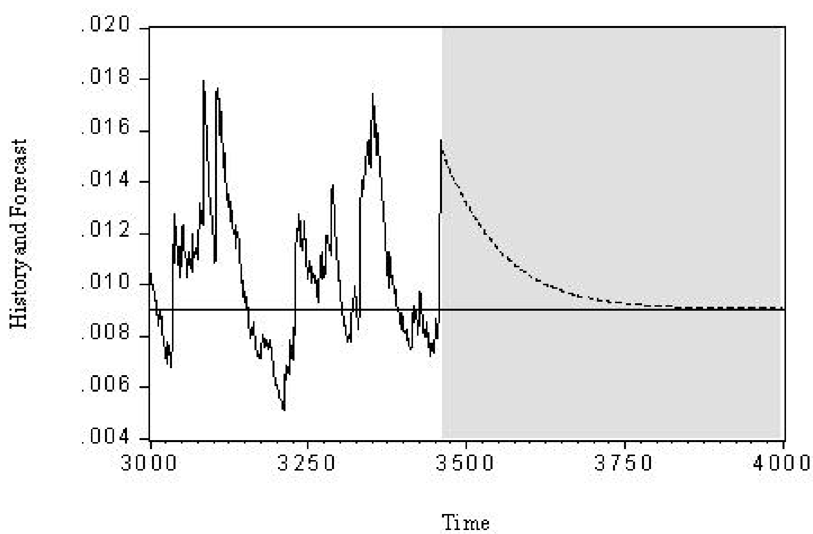

Figure 8.14: Conditional Standard Deviation, History and Extended Forecast from GARCH(1,1)

8.6 Exercises, Problems and Complements

1. Volatility dynamics: correlograms of squares.

   In the Chapter 3 EPC, we suggested that a time series plot of a squared residual, e_t^2, can reveal serial correlation in squared residuals, which corresponds to non-constant volatility, or heteroskedasticity, in the levels of the residuals. Financial asset returns often display little systematic variation, so instead of examining residuals from a model of returns, we often examine returns directly. In what follows, we will continue to use the notation e_t , but you should interpret e_t it as an observed asset return.

   a. Find a high frequency (e.g., daily) financial asset return series, $\mathrm{e_t}$, plot it, and discuss your results.

   b. Perform a correlogram analysis of e_t, and discuss your results.

   c. Plot e_t^2, and discuss your results.

   d. In addition to plotting e_t^2, examining the correlogram of e_t^2 often proves informative for assessing volatility persistence. Why might that be so? Perform a correlogram analysis of e_t^2 and discuss your results.

2. Removing conditional mean dynamics before modeling volatility dynamics. In the application in the text we noted that NYSE stock returns appeared to have some weak conditional mean dynamics, yet we ignored them and proceeded directly to model volatility.

   a. Instead, first fit autoregressive models using the SIC to guide order selection, and then fit GARCH models to the residuals. Redo the entire empirical analysis reported in the text in this way, and discuss any important differences in the results.

b. Consider instead the simultaneous estimation of all parameters of AR(p)-GARCH models. That is, estimate regression models where the regressors are lagged dependent variables and the disturbances display GARCH. Redo the entire empirical analysis reported in the text in this way, and discuss any important differences in the results relative to those in the text and those obtained in part a above.

3. Variations on the basic ARCH and GARCH models.

   Using the stock return data, consider richer models than the pure ARCH and GARCH models discussed in the text.

   a. Estimate, diagnose and discuss a threshold GARCH(1,1) model.
   b. Estimate, diagnose and discuss an EGARCH(1,1) model.
   c. Estimate, diagnose and discuss a component GARCH(1,1) model.
   d. Estimate, diagnose and discuss a GARCH-M model.

4. Empirical performance of pure ARCH models as approximations to volatility dynamics.

   Here we will fit pure ARCH(p) models to the stock return data, including values of p larger than p=5 as done in the text, and contrast the results with those from fitting GARCH(p,q) models.

   a. When fitting pure ARCH(p) models, what value of p seems adequate?
   b. When fitting GARCH(p,q) models, what values of p and q seem adequate?
   c. Which approach appears more parsimonious?

5. Direct modeling of volatility proxies.

   In the text we fit an AR(5) directly to a subset of the squared NYSE stock returns. In this exercise, use the *entire* NYSE dataset.

a. Construct, display and discuss the fitted volatility series from the AR(5) model.

b. Construct, display and discuss an alternative fitted volatility series obtained by exponential smoothing, using a smoothing parameter of .10, corresponding to a large amount of smoothing, but less than done in the text.

c. Construct, display and discuss the volatility series obtained by fitting an appropriate GARCH model.

d. Contrast the results of parts a, b and c above.

e. Why is fitting of a GARCH model preferable in principle to the AR(5) or exponential smoothing approaches?

6. GARCH volatility forecasting.

You work for Xanadu, a luxury resort in the tropics. The daily temperature in the region is beautiful year-round, with a mean around 76 (Fahrenheit!) and no conditional mean dynamics. Occasional pressure systems, however, can cause bursts of temperature volatility. Such volatility bursts generally don't last long enough to drive away guests, but the resort still loses revenue from fees on activities that are less popular when the weather isn't perfect. In the middle of such a period of high temperature volatility, your boss gets worried and asks you to make a forecast of volatility over the next ten days. After some experimentation, you find that daily temperature $\mathrm{y_t}$ follows 11111 $\mathrm{y_t}|\Omega_{\mathrm{t}-1} \sim \mathrm{N}(\mu,\ \sigma_{\mathrm{t}}^2)$, where σ_{t}^2 follows a GARCH(1,1) process, $\sigma_{\mathrm{t}}^2 = \omega + \alpha\varepsilon_{\mathrm{t}-1}^2 + \beta\sigma_{\mathrm{t}-1}^2$.

a. Estimation of your model using historical daily temperature data yields $\hat{\mu} = 76$, $\hat{\omega} = 3$, $\hat{\alpha} = .6$, and $\hat{\beta} = 0$. If yesterday's temperature was 92 degrees, generate point forecasts for each of the next ten days conditional variance.

b. According to your volatility forecasts, how many days will it take until volatility drops enough such that there is at least a 90% probability that the temperature will be within 4 degrees of 76?

c. Your boss is impressed by your knowledge of forecasting, and asks you if your model can predict the next spell of bad weather. How would you answer him?

7. Assessing volatility dynamics in observed returns and in standardized returns.

   In the text we sketched the use of correlograms of squared observed returns for the detection of GARCH, and squared standardized returns for diagnosing the adequacy of a fitted GARCH model. Examination of Ljung-Box statistics is an important part of a correlogram analysis. McLeod and Li (1983) show that the Ljung-Box statistics may be legitimately used on squared observed returns, in which case it will have the usual

$$\chi^2_{\mathrm{m}}$$

   distribution under the null hypothesis of independence. Bollerslev and Mikkelson (1996) argue that one may also use the Ljung-Box statistic on the squared standardized returns, but that a better distributional approximation is obtained in that case by using a

$$\chi^2_{\mathrm{m-k}}$$

   distribution, where k is the number of estimated GARCH parameters, to account for degrees of freedom used in model fitting.

8. Allowing for leptokurtic conditional densities.

   Thus far we have worked exclusively with conditionally Gaussian GARCH models, which correspond to $\varepsilon_t = \sigma_t v_t \; v_t \overset{iid}{\sim} N(0, 1)$, or equivalently, to normality of the standardized return, ε_t / σ_t.

a. The conditional normality assumption may sometimes be violated. However, Bollerslev and Wooldridge (1992) show that GARCH parameters are consistently estimated by Gaussian maximum likelihood even when the normality assumption is incorrect. Sketch some intuition for this result.

b. Fit an appropriate conditionally Gaussian GARCH model to the stock return data. How might you use the histogram of the standardized returns to assess the validity of the conditional normality assumption? Do so and discuss your results.

c. Sometimes the conditionally Gaussian GARCH model does indeed fail to explain all of the leptokurtosis in returns; that is, especially with very high-frequency data, we sometimes find that the conditional density is leptokurtic. Fortunately, leptokurtic conditional densities are easily incorporated into the GARCH model. For example, in Bollerslev's (1987) conditionally Student's-t GARCH model, the conditional density is assumed to be Student's t, with the degrees-of-freedom d treated as another parameter to be estimated. More precisely, we write

$$v_t \overset{iid}{\sim} \frac{t_d}{std(t_d)}.$$

$$\varepsilon_t = \sigma_t v_t$$

What is the reason for dividing the Student's t variable, t_d, by its standard deviation, $\mathrm{std}(\mathrm{t_d})$? How might such a model be estimated?

9. Optimal prediction under asymmetric loss.

   In the text we stressed GARCH modeling for improved interval and density forecasting, implicitly working under a symmetric loss function. Less obvious but equally true is the fact that, under *asymmetric* loss, volatility dynamics can be exploited to produce improved *point* forecasts, as shown by Christoffersen and Diebold (1996, 1997). The optimal predictor under asymmetric loss is not the conditional mean, but rather the conditional mean shifted by a time-varying adjustment that depends on the conditional variance. The intuition for the bias in the optimal predictor is simple – when errors of one sign are more costly than errors of the other sign, it is desirable to bias the forecasts in such a way as to reduce the chance of making an error of the more damaging type. The optimal amount of bias depends on the conditional prediction error variance of the process because, as the conditional variance grows, so too does the optimal amount of bias needed to avoid large prediction errors of the more damaging type. .

10. Multivariate GARCH models.

    In the multivariate case, such as when modeling a *set* of returns rather than a single return, we need to model not only conditional variances, but also conditional *covariances*.

    a. Is the GARCH conditional variance specification introduced earlier, say for the $\mathrm{i-th}$ return, $\sigma^2_{\mathrm{it}} = \omega + \alpha\varepsilon^2_{\mathrm{i,t-1}} + \beta\sigma^2_{\mathrm{i,t-1}}$, still appealing in the multivariate case? Why or why not?

    b. Consider the following specification for the conditional covariance between $\mathrm{i-th}$ and j-th returns: $\sigma_{\mathrm{ij,t}} = \omega + \alpha\varepsilon_{\mathrm{i,t-1}}\varepsilon_{\mathrm{j,t-1}} + \beta\sigma_{\mathrm{ij,t-1}}$. Is it appealing? Why or why not?

c. Consider a fully general multivariate volatility model, in which every conditional variance and covariance may depend on lags of every conditional variance and covariance, as well as lags of every squared return and cross product of returns. What are the strengths and weaknesses of such a model? Would it be useful for modeling, say, a set of five hundred returns? If not, how might you proceed?

8.7 Notes

This chapter draws upon the survey by Diebold and Lopez (1995), which may be consulted for additional details. Other broad surveys include Bollerslev, Chou and Kroner

(1992), Bollerslev, Engle and Nelson (1994), Taylor (2005) and Andersen et al. (2007). Engle (1982) is the original development of the ARCH model. Bollerslev (1986) provides the important GARCH extension, and Engle (1995) contains many others. Diebold (1988) shows convergence to normality under temporal aggregation. TGARCH traces to Glosten, Jagannathan and Runkle (1993), and EGARCH to Nelson (1991). Engle, Lilien and Robins (1987) introduce the GARCH-M model, and Engle and Lee (1999) introduce component GARCH. Recently, methods of volatility measurement, modeling and forecasting have been developed that exploit the increasing availability of high-frequency financial asset return data. For a fine overview, see Dacorogna et al. (2001), and for more recent developments see Andersen, Bollerslev, Diebold and Labys (2003) and Andersen, Bollerslev and Diebold (2006). For insights into the emerging field of financial econometrics, see Diebold (2001) and many of the other papers in the same collection.