8 Time-varying volatility: Autoregressive conditional heteroskedasticity and generalized autoregressive conditional heteroskedasticity models

Chapter map

If you read the previous two chapters carefully, you now know everything there is to know about time-series analysis. Well, almost everything. You learned in chapter 6 how time-series models are built up from a few fundamental building blocks, and you understand how complex dynamic behavior can be modeled by combining low-order autoregressive (AR) and moving-average (MA) components. Chapter 7 taught you to be a time-series "detective", weighing the "clues" provided by the autocorrelations and partial autocorrelations, then grilling your "suspects" by fitting candidate models, examining the fits, residuals, and forecasts to weed out the red herrings until the "culprit" is identified.

So what is left to learn? Well, what we have learned so far is the fundamental approach to time-series models. And econometricians always build upon fundamental models in three ways. First, they generalize a univariate model to handle multiple relationships simultaneously. (We will cover that in chapter 9.) Second, they extend the model to a broader class of phenomena. (That is in chapter 10.) Third, they find ways to handle violations of the classical assumptions. We will cover that topic now.

In the previous two chapters, we assumed the random contribution to our time series, ϵ_t, has constant variance, σ_ϵ^2. In practice, this assumption appears to be violated frequently, especially in time series of financial or economic data. Markets and economies seem subject to periods of "normal" volatility interrupted by periods of unusual turbulence that persist for some time. In other words, the volatility of a time series can itself follow a time-series process.

Nonconstant variance of the error term, or heteroskedasticity, is a familiar problem in cross-section econometrics with familiar solutions (for example, weighted regression when the source of the heteroskedasticity is known and measurable or heteroskedasticity-consistent estimators in more general situations). Time-varying volatility requires a different approach.

We set the stage in the next section by examining some data that exhibit time-varying volatility. The following section covers some of the first models introduced to handle these types of data. The final section reviews generalizations that followed these first models and discusses how to select among the extensive model choices available.

8.1 Examples of time-varying volatility

Financial markets, among other things, produce "price discovery", that is, asset prices that balance the willingness of relatively bullish investors to buy with the willingness of relatively bearish investors to sell. Of course, news arrives every day that shifts the balance of investor opinion about an asset's likely future performance. Some news—the unemployment and inflation rates, the balance of trade, and the like—affects investor views of the market as a whole, while other news—changes in management, updated sales figures, etc.—affects only a specific sector or company. Investors digest this information, update their willingness to buy or sell assets, and their actions change prices. The types of information that affect asset prices are many and varied, and it is impractical to model them all. As a result, their impact on prices is part of the random component of our models, and the typical magnitude of their impact helps determine the magnitude of the residual variance, σ_ϵ^2.

During uneventful periods, little news of import arrives, so investor opinions change relatively little. Moreover, long periods of relatively stable prices tend to narrow the range of investor opinions of asset values. In contrast, sudden and unexpected news can shift investor opinion significantly. More importantly, extraordinary events can sow confusion among investors. Investors may struggle for some time to settle on an updated opinion of value. In these circumstances, investors may exhibit heightened reactions to scraps of information that either increase concern or restore confidence. During these periods, the variance of the random component of prices is increased—at least, that is the way these events are reflected in our models, which, after all, are simplified representations of reality.

As an example of this phenomenon, let's revisit the TED spread that we used in chapter 3 to illustrate the use of some of Stata's smoothers. Recall that the TED spread is the difference in basis points—that is, hundredths of a percent—between the 3-month London interbank offered rate (a rate large banks charge each other) and the yield on the 3-month on-the-run Treasury bill. The TED spread measures the market's opinion of the creditworthiness of the banking sector. United States Treasury securities are regarded as having no risk of default; hence, the yields on Treasury securities provide a benchmark for all other interest rates. When investors become concerned about the strength of the banking sector, interbank interest rates increase relative to Treasury yields. In other words, the TED spread increases.

Figure 8.1 displays the TED spread from 2 January 1987 through 30 December 2011. In chapter 3, we used smoothers to track changes in the level of the TED spread, and we noted the sudden increase in the spread during the financial distress that began in 2007. Here we are more interested in changes in the volatility of the spread than in its level.

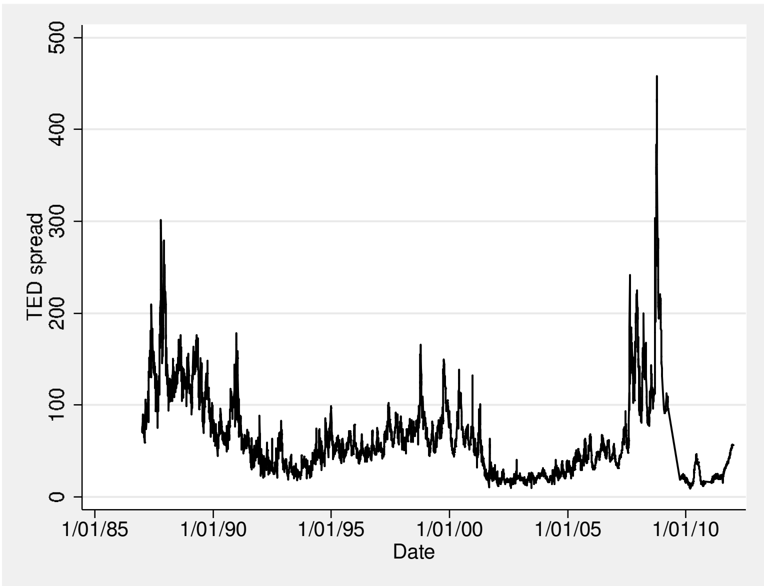

Figure 8.1. The TED spread, 1/2/1987–12/30/2012

You can think of the TED spread as the price the market places on bank risk, and, like many financial prices, the TED spread is reasonably approximated by a random walk.[1]

$$\text{TED spread}_t = \text{TED spread}_{t-1} + \epsilon_t$$

Again, as is common with financial prices, the one-day changes in the TED spread come from a fat-tailed, nonnormal distribution. Figure 8.2 displays a very-squashed box plot of the one-day changes in the TED spread.[2] Half the one-day changes lie within the shaded box in a range of ± 2 basis points. The mass of dots above and below the box and whiskers mark outside values, extreme one-day changes—one as large as 100 basis points—that reflect the fat tails of this distribution.

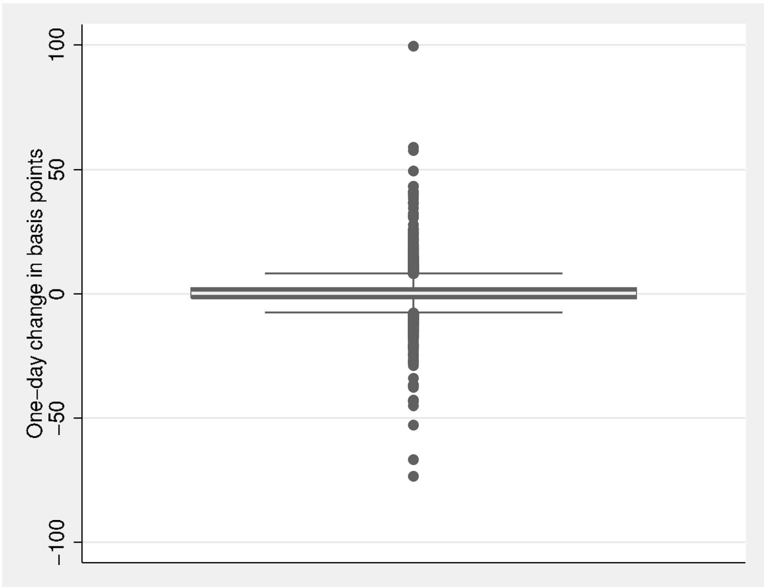

Figure 8.2. The distribution of one-day changes in the TED spread

In this chapter, we are not concerned with the fat tails per se. Rather we are interested in the stability of the dispersion of one-day changes; that is, we are interested in the behavior of σ^2_ϵ over time. Figure 8.3 displays the standard deviation of one-day

1. Actually, the TED spread appears to follow a random walk with drift—the mean of the one-day change in the spread is slightly positive (0.2 basis points) and significant. This drift term is unimportant for our purposes here.
2. Box plots provide convenient summaries of the distributions of random variables. The shaded box depicts the interquartile range; half the observations lie within the box. The horizontal line in the interior of the shaded box marks the median. The "antennae" extending above and below the box indicate the upper and lower adjacent values, which are robust measures of the dispersion in the tails. The dots that appear above and below the adjacent values are outside values, that is, unusually extreme values.

changes in the TED spread by year from 1987 through 2011. Two episodes stand out. The spike in 1987 reflects the uncertainty of investors as they tried to gauge the exposure of financial institutions to the stock market crash in October 1987. Volatility receded in 1988 and continued to decline secularly through 2006. The spike in the volatility of the spread in 2007 and 2008 reflects investor fear and uncertainty as the housing crisis triggered failures and bailouts at large financial firms (Washington Mutual, Countrywide, Bear Stearns, Lehman Brothers, and AIG) and unprecedented government interventions (the TARP program, which injected capital into the banking sector, and the placing of Fannie Mae and Freddie Mac into conservatorship). Even without any formal statistical test, it appears unlikely that a model with constant σ_ϵ^2 will fit these data well.

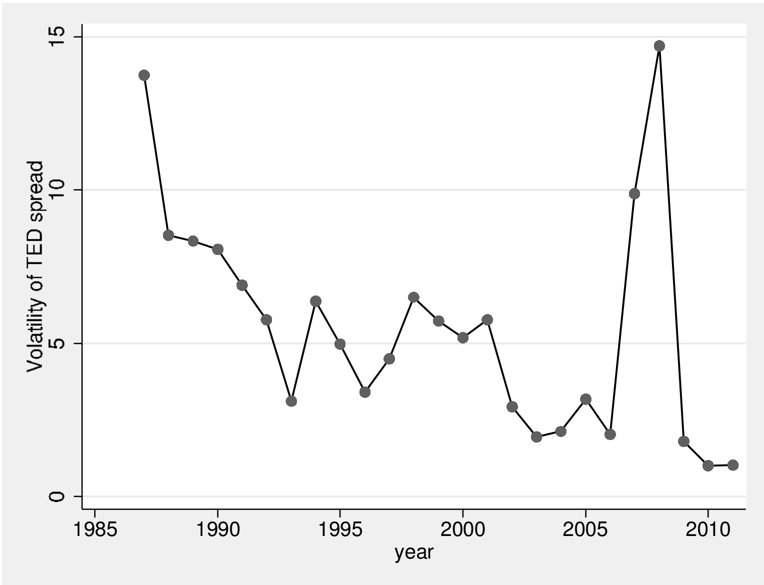

Figure 8.3. Variation in the volatility of the TED spread

While financial time series are particularly prone to time-varying volatility, other time series exhibit similar swings in variance. The real growth rate of GDP, which we analyzed in the previous chapter, provides a good example. Figure 8.4 displays the standard deviation of the quarterly real growth rate of the economy.[3] The graph has a choppy appearance because there are only four quarterly observations in each annual standard deviation. Nonetheless, three eras stand out. The early, post–World War II years (1947–1959) are particularly volatile—the 9.5% standard deviation in 1958 is the highest value in our sample. The variability of real growth drops sharply in 1961, then ratchets up steadily until the turbulent years from the late 1970s through 1982, when the Federal Reserve's so-called monetarist experiment wrung inflation from the economy

3. Annualized percentage growth rate.

at the cost of a severe recession. From 1983 through 2007, the volatility of real growth remains muted, an era that has been dubbed the "Great Moderation". This period of calm ends suddenly in 2008 and 2009 as the global financial crisis takes hold.

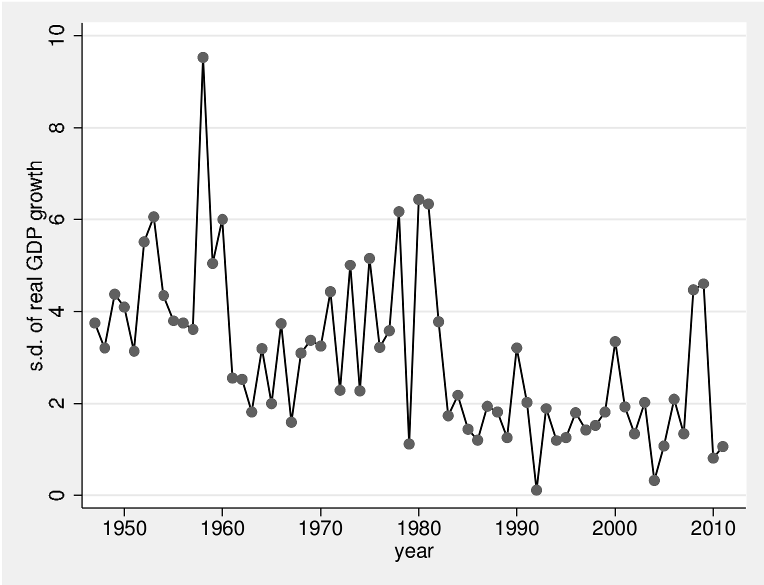

Figure 8.4. Variation in the volatility of real GDP growth

Figure 8.5 provides a clearer picture of those patterns by calculating the standard deviation of real growth over nonoverlapping five-year periods.[4]

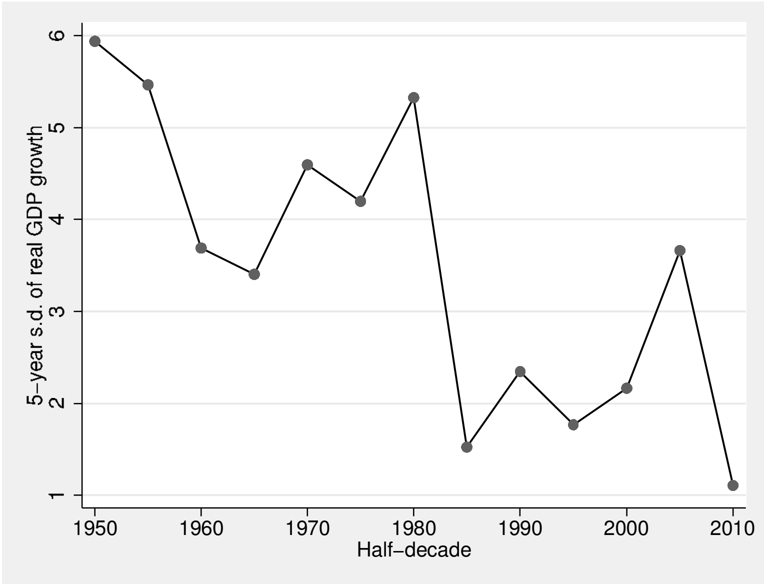

Figure 8.5. Five-year standard deviations of real GDP growth

8.2 ARCH: A model of time-varying volatility

How can we model the empirical observation that many time series exhibit episodes of higher and lower variability? This challenge is a little different, a little less concrete, than what we have been dealing with so far. Up to now, we have built models of observable series, our y_t. To select the appropriate dynamic characterization of y_t, we have relied on statistical summaries—autocorrelations, partial autocorrelations, and the like—but we have been modeling something we can observe, something we can compare directly with model fits and forecasts. Now, however, we are trying to model the evolution over time of an unobservable parameter, σ_y^2. We are at one more degree of separation from observable phenomena.

Engle (1982) introduced the ARCH model to capture time-varying volatility within the structure of standard time-series models. Engle's innovation sparked an explosion of variations of the ARCH approach, and we will look at some of the leading examples a little later. For now, let's walk through Engle's original insight.

4. The magnitudes of the estimated five-year and annual standard deviations are not directly comparable, because they cover varying amounts of time.

The autoregressive integrated moving-average (ARIMA) model we have discussed so far is

$$\phi(L)y_t = \mu + \theta(L)\epsilon_t$$

where ϵ_t is white noise with variance equal to σ_ϵ^2. The conditional mean of y_t depends on prior values of y_t (through the coefficients in $\phi[L]$). However, the unconditional mean of y_t is constant:

$$Ey_t = \frac{\mu}{(1 - \phi_1 - \cdots - \phi_p)} \equiv \mu^*$$

Likewise, the unconditional variance of y_t is also constant. In the previous chapter, we showed that this variance—the initial autocovariance, γ_0—is given by

$$\gamma_0 = \phi_1\gamma_1 + \cdots + \phi_p\gamma p - \gamma_{y\epsilon}(0) - \theta_1\gamma\gamma_{y\epsilon}(1) - \cdots - -\theta_q\gamma\gamma_{y\epsilon}(q)$$

where $\gamma_{y\epsilon}(j)$ is the cross-covariance between y_t and ϵ_{t-j}. The expression is messy, but it is composed entirely of constant parameters.

So where does this leave us? How can we incorporate a concept of time-varying volatility? Engle's solution centered on the conditional variance of ϵ_t. Even though the unconditional variance of ϵ_t is the constant σ_ϵ^2, Engle suggested that the conditional variance of ϵ_t could follow an AR process of its own. Specifically, he proposed modeling the square of ϵ_t as an AR(m) process

$$\epsilon_t^2 = \zeta + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_m\epsilon_{t-m}^2 + \nu_t$$

or

$$A(L)\epsilon_t^2 = \zeta + \nu_t$$

where ν_t is a new white-noise process. The conditional expectation of ϵ_t^2 is

$$E(\epsilon_t^2|\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots) = \zeta + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_m\epsilon_{t-m}^2$$

A white-noise series, ϵ_t, satisfying these conditions is described as an ARCH(m) process.

ARCH models require a few more restrictions than ordinary autoregressive moving-average (ARMA) models. ϵ_t^2 can never be negative. This restriction can be guaranteed if $\nu_t \geq -\zeta, \forall t$ and $\alpha_j \geq 0, \forall j$. As with all ARMA models, the roots of

$$A(z) = 0$$

must all lie outside the unit circle. With all $\alpha_j \geq 0$, this condition is equivalent to the condition

$$\alpha_1 + \cdots + \alpha_m < 1$$

When all of these conditions hold, it can be shown that the unconditional variance of ϵ_t is

$$\sigma_\epsilon^2 = E(\epsilon_t^2) = \frac{\zeta}{1 - \alpha_1 - \cdots - \alpha_m}$$

The ARCH model extends the standard ARMA model to incorporate time-varying volatility but at the cost of the additional restrictions required to guarantee that $\epsilon_t^2 \geq 0$. In practice, it can be difficult to impose the additional stationarity and nonnegativity restrictions. Nonetheless, ARCH and its many descendants have proved useful in refining our ability to model many time series.

Let's see how Stata handles ARCH models. We begin with a model for the growth rate of real GDP in the United States. Recall that in chapter 7, we fit an AR(1) model for the growth of real GDP in the United States.[5]

```
. arima growth, ar(1) nolog
ARIMA regression
Sample:  1947:2 - 2012:1                        Number of obs      =       260
                                                Wald chi2(1)       =     58.51
Log likelihood =  -707.238                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |                 OPG
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
growth       |
       _cons |   3.115064   .3723441     8.37   0.000     2.385283    3.844845
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .3762591   .0491894     7.65   0.000     .2798497    .4726685
-------------+----------------------------------------------------------------
      /sigma |   3.672893   .1175688    31.24   0.000     3.442463    3.903324
------------------------------------------------------------------------------
```

As we showed in the previous chapter, this specification provides a reasonable fit to the data. However, our casual inspection above of the variability of real growth suggested the presence of an ARCH component. How can we test this hypothesis?

Stata provides a test for ARCH effects that can be used after a `regress` command. We will demonstrate it by using the AR(1) specification for real growth:

```
. regress L(0/1).growth
      Source |       SS       df       MS              Number of obs =     259
-------------+------------------------------           F(  1,   257) =   42.59
       Model |  579.230003     1  579.230003           Prob > F      =  0.0000
    Residual |  3494.96519   257  13.5990863           R-squared     =  0.1422
-------------+------------------------------           Adj R-squared =  0.1388
       Total |  4074.19519   258  15.7914542           Root MSE      =  3.6877

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      growth |
         L1. |   .3764496   .0576814     6.53   0.000     .2628611    .4900381
             |
       _cons |   1.961878   .2916873     6.73   0.000     1.387477     2.53628
------------------------------------------------------------------------------
```

5. You may have noticed the difference in the estimate of the constant compared with the estimate reported in chapter 7. We have annualized the quarterly growth rates here to make the units easier to read.

The `regress` estimates are very similar to the `arima` estimates. The root mean squared error in the regression is close to the estimated σ_ϵ calculated by `arima`; the ϕ_1 coefficient estimates are similar; and the estimated constants are similar after the translation to the ARIMA specification is made:

```
. display = _b[_cons]/(1 - _b[L1.growth])
3.1463024
```

The `estat` command, which we introduced in chapter 1, provides a Lagrange multiplier test for ARCH components.

```
. estat archlm, lags(1 2 3 4)
LM test for autoregressive conditional heteroskedasticity (ARCH)
```

lags(p)	chi2	df	Prob > chi2
1	4.789	1	0.0286
2	9.913	2	0.0070
3	9.957	3	0.0189
4	13.541	4	0.0089

```
H0: no ARCH effects      vs.  H1: ARCH(p) disturbance
```

The `lags(#)` option specifies the order of the ARCH AR component under the alternative hypothesis. In this example, we have tested the null of no ARCH effects against four separate alternatives: ARCH models of orders 1 through 4. The null is rejected against all four alternatives. Determining lag lengths is always a tricky business, so we are not going to focus yet on the appropriate order of the ARCH process. However, we do interpret these results as indicating the presence of ARCH disturbances of some order.

Note: As we mentioned above, the `estat archlm` command is available only after the `regress` command. This may appear to be a limitation when our original time-series model involves MA parameters, but, as a practical matter, we can still test for ARCH effects. Remember that a model with a finite MA component can be written as an infinite-order AR model. While it's not possible to use `regress` to estimate a regression with an infinite number of terms, for a stationary model, the coefficients on the AR terms die out relatively quickly. So you can estimate a finite-order regression with just-enough AR terms in place of a mixed ARMA model. Determining how many AR terms is "just enough" is a bit of an art, but basically you want to make sure the residual is indistinguishable from white noise.

Another approach, proposed by Bollerslev (1988), is to apply the Box–Jenkins methodology to the $\widehat{\epsilon}_t^2$, the squared residuals from the ARIMA model, for the observable variable. This approach may indicate that a pure AR model for ϵ_t^2 may not provide the best characterization of the variance process. More on that in the next section.

Stata provides the **arch** command to fit ARCH (and many related) models. The syntax is similar to the **arima** syntax:

arch *depvar* [**indepvars**] [*if*] [*in*] [*weight*] [, **arch(***numlist***)** *other_options*]

The **arch** command has many, many options to take account of the many different models of time-varying volatility that have been proposed since Engle's (1982) article. We will cover some of the more useful ones shortly. The **arch** command also allows the **arima()**, **ar()**, and **ma()** options we used in the previous chapter with the **arima** command.

Because the results of the **estat** command suggest the presence of first-order ARCH effects, let's begin by fitting that model.

```
. arch growth, ar(1) arch(1) nolog
ARCH family regression -- AR disturbances

Sample: 1947:2 - 2012:1                          Number of obs   =        260
Distribution: Gaussian                           Wald chi2(1)    =      31.01
Log likelihood = -701.1802                       Prob > chi2     =     0.0000
```

growth	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
growth						
_cons	3.097168	.3469909	8.93	0.000	2.417078	3.777257
ARMA						
ar						
L1.	.3993747	.0717198	5.57	0.000	.2588064	.539943
ARCH						
arch						
L1.	.2608906	.0750396	3.48	0.001	.1138158	.4079655
_cons	10.23213	.8356407	12.24	0.000	8.594301	11.86995

Compare these estimates with the **arima** estimates above. The coefficient estimates are very similar in the two models. In the ARCH model, the standard error of the constant in the structural equation is slightly smaller than before, while the standard error of ϕ_1 is somewhat larger than before. In the ARCH equation, both ζ and α_1 are significant, and $0 < \alpha_1 < 1$ as is required for stationarity of ϵ_t^2.

We have assumed the errors follow an ARCH(1) process so far, but, of course, they may follow a higher-order process. In the previous chapter, we reviewed the autocorrelations and partial autocorrelations of real GDP growth for clues to the order of the ARMA model. For ARCH models, we have no similar set of statistics to guide us, so we will resort to the old-fashioned approach of estimating multiple specifications.[6]

```
. arch growth, ar(1) arch(1 2) nolog
ARCH family regression -- AR disturbances
Sample: 1947:2 - 2012:1                                 Number of obs   =        260
Distribution: Gaussian                                  Wald chi2(1)    =      38.91
Log likelihood = -691.3014                              Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
      growth |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
growth       |
       _cons |   3.467476   .3396975    10.21   0.000     2.801681     4.13327
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .4383648   .0702733     6.24   0.000     .3006316     .576098
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .3308078   .0839041     3.94   0.000     .1663588    .4952568
         L2. |   .3870112   .1109962     3.49   0.000     .1694627    .6045598
             |
       _cons |   5.508447    .852282     6.46   0.000     3.838005    7.178889
------------------------------------------------------------------------------
```

6. We realize we are committing a cardinal statistical sin. The appropriate approach is to fit an encompassing model first, then trim insignificant coefficients. By reversing this order, we complicate the interpretation of significance tests, probably beyond hope of untangling. Our aim here is to illustrate the `arch` command. Please do not take our example as encouragement of bad research habits.

```
. arch growth, ar(1) arch(1 2 3) nolog
ARCH family regression -- AR disturbances
Sample: 1947:2 - 2012:1                          Number of obs   =       260
Distribution: Gaussian                           Wald chi2(1)    =     37.02
Log likelihood = -691.1207                       Prob > chi2     =    0.0000

                             OPG
      growth      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]

growth
       _cons    3.48945   .3438003    10.15   0.000     2.815614    4.163286

ARMA
          ar
         L1.   .4309567   .0708268     6.08   0.000     .2921388    .5697746

ARCH
        arch
         L1.   .3216774   .0831899     3.87   0.000     .1586282    .4847265
         L2.   .4019769   .1126535     3.57   0.000     .1811802    .6227737
         L3.   .0245882   .0494727     0.50   0.619    -.0723766    .1215529

       _cons   5.216906   .8617869     6.05   0.000     3.527834    6.905977
```

It appears that a second-order ARCH process may fit the data best. Note that the ARCH(2) model produces slightly different estimates of μ and ϕ_1.

Recall that we fit two time-series models to real GDP growth, an AR(1) model and an MA(2) model. For comparison, here are estimates of the original MA(2) model and the same model with first-, second-, and third-order ARCH effects added.

	MA(2)		ARCH(1)		ARCH(2)		ARCH(3)	
	Coef.	$P > \|z\|$	Coef.	$P > \|z\|$	Coef.	$P > \|z\|$	Coef.	$P > \|z\|$
μ	3.11	0.00	3.18	0.00	3.53	0.00	3.52	0.00
θ_1	0.332	0.00	0.362	0.00	0.376	0.00	0.377	0.00
θ_2	0.219	0.00	0.251	0.00	0.259	0.00	0.261	0.00
ζ			9.70	0.00	5.61	0.00	5.67	0.00
α_1			0.291	0.00	0.338	0.00	0.341	0.00
α_2					0.355	0.00	0.351	0.00
α_3							−0.006	0.90

As before, an ARCH(2) process seems to fit best. Note that the estimate of α_3 in the ARCH(3) model is negative in this case, although it is not significant.

After fitting an ARCH model, Stata's `predict` command can generate an estimate of the conditional variances implied by the model. As before, either dynamic or static predictions are available. Let's use that command to get a look at the conditional variances of the AR(1) model with ARCH(2) disturbances.

```
. arch growth, ar(1) arch(1 2)
  (output omitted)
. predict variance, variance
. format date %dCCYY
. tsline variance
```

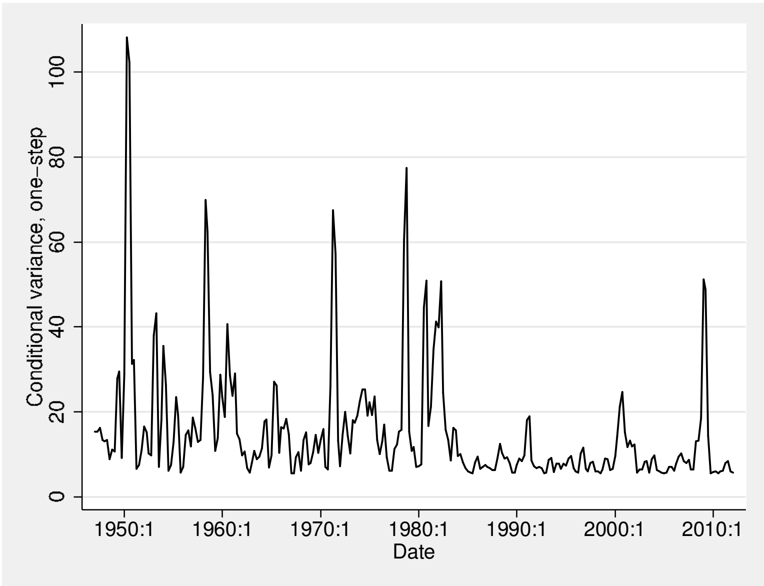

Figure 8.6. The conditional variance of ϵ_t

The conditional variance shown in figure 8.6 exhibits repeated spikes in the years prior to the mid-1980s followed by two decades of relative quiet until the spike in 2008–2009. This pattern echoes the phenomena we observed in figure 8.4 and figure 8.5 above.

It turns out that an alternative formulation of the ARCH model provides a convenient basis for developing extensions to this framework. Suppose that

$$\epsilon_t = \sqrt{h_t} \times \omega_t$$

where ω_t is a white-noise process but now with the additional restriction $E\omega^2 = 1$. In this representation, the variable $\sqrt{h_t}$ scales the white-noise innovations ω_t. Hence, the process for h_t determines the time path of ϵ^2. If that process is

$$h_t = \zeta + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_m \epsilon_{t-1}^2$$

then, as before,

$$E(\epsilon_t^2 | \epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots) = \zeta + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_m \epsilon_{t-m}^2$$

that is, ϵ_t follows an ARCH(m) process. With a little bit of algebra, we can derive a compact representation for ν_t, the innovation in the ARCH(m) process.[7,8]

$$\nu_t = h_t \times (\omega_t^2 - 1)$$

With this preparation, we can proceed now to some extensions of Engle's original proposal.

8.3 Extensions to the ARCH model

The phenomenon of time-varying volatility had been recognized for decades, but until the publication of Engle's original proposal, researchers lacked a satisfactory technique for modeling this form of heteroskedasticity. Engle's innovation generated a deluge of empirical articles applying this new technique. However, the initial ARCH model addresses only the fundamental observation that volatility often varies over time. Other aspects of time-varying volatility—other long-recognized empirical regularities in financial and economic time series—are not captured by the "vanilla" ARCH model. As a consequence, econometricians developed numerous, often competing extensions to the ARCH model.

Some of the more important empirical regularities are the following:

- The uneven, intermittent, or random arrival of "news" (impactful new information) is commonly cited as cause of time-varying volatility. However, many series appear to react asymmetrically to positive and negative news.
- The conditional mean of y_t, the observable time series, often appears to depend on the current level of volatility. For example, the level of stock prices often declines during periods of unusually high uncertainty.

7. Squaring both sides of the equation for ϵ_t produces

$$\epsilon_t^2 = h_t \times \omega_t^2$$

but

$$\epsilon_t^2 = \zeta + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_m \epsilon_{t-m}^2 + \nu_t = h_t + \nu_t$$

hence,

$$h_t + \nu_t = h_t \times \omega_t^2$$

or

$$\nu_t = h_t \times (\omega_t^2 - 1)$$

8. This formulation, although convenient, is more restrictive than the first version of the ARCH model we presented. In particular, somewhat unintuitive restrictions on the α_i are required to guarantee a solution for σ_ν^2, the unconditional variance of the innovation in the AR process for ϵ_t^2. See Hamilton (1994) for details.

- Asset prices, such as stock prices, tend to have distributions with "fat tails". In other words, extreme events (unusually large price increases or decreases) occur more frequently than in a normal distribution. The ARCH model introduces some leptokurtosis[9] in the conditional variance but less than is typically observed in financial time series.

The options to Stata's `arch` command provide these and many other extensions to Engle's original formulation. Covering them all would take us too far afield, but we highlight the most important below.[10]

8.3.1 GARCH: Limiting the order of the model

In many situations, it may appear that a high-order ARCH process is required to provide an adequate description of the time-varying volatility. This possibility does not present any theoretical obstacles. All stationary time series can be represented by an AR process.[11] But high-order processes are unwieldy and difficult to estimate precisely. A better approach is to combine low-order AR and MA components to model the dynamic behavior of the series.

A similar situation can arise when modeling time-varying volatility. The data may suggest that many ARCH terms are required. The GARCH introduced by Bollerslev (1986) can provide a good fit with a lower-order parameterization.

Using the notation we developed at the end of section 8.2, we characterized the time-series properties of the random error in the ARIMA model as an AR process in h_t, the variable that scales the standard normal innovation. We assumed that the ARCH process had some finite order, call it m, but more generally, we can assume that the ARCH process depends on an infinite-order autoregression in h_t,

$$h_t = \zeta + \pi(L)\epsilon_t^2$$

where

$$\pi(L) = \sum_{j=1}^{\infty} \pi_j^* L^j$$

As we discussed in chapter 6, this infinite-order lag polynomial can be written more compactly as the ratio of two finite-order lag polynomials:

$$\pi(L) = \frac{\alpha(L)}{1-\delta(L)} = \frac{\alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_m L^m}{1 - \delta_1 L - \delta_2 L^2 - \cdots - \delta_r L^r}$$

9. The typical definition of kurtosis (or, more precisely, excess kurtosis) is $\mu_4/\sigma^4 - 3$, where μ_4 is the fourth moment about the mean and σ is the standard deviation. A distribution, like the normal, with zero excess kurtosis is called mesokurtic. A distribution with positive excess kurtosis (and, hence, "fat tails") is called leptokurtic.
10. See Bollerslev, Chou, and Kroner (1992) for a useful survey of ARCH and its extensions. They also discuss many of the applications of these extensions to financial time series.
11. An infinite-order autoregression may be required to provide an exact representation.

Multiplying both sides of the equation for h_t by $1 - \delta(L)$ yields

$$\{1 - \delta(L)\}\, h_t = \kappa + \alpha(L)\epsilon_t^2$$

where

$$\kappa = (1 - \delta_1 - \cdots - \delta_r)\zeta$$

This parameterization is the GARCH(r, m) model.[12] As in the basic ARCH model, the requirement that $h_t > 0$ implies that all the coefficients of both lag polynomials are positive.

We will use the time series of inflation in U.S. consumer prices from January 1960 through March 2012 to illustrate how Stata fits a GARCH model. We calculate the inflation rate as the change in the log of the consumer price index (CPI). We will model the monthly rate of inflation, but let's calculate the annual rate of inflation (overlapping 12-month log differences) as a simple, MA smoother to highlight the systematic component of inflation.

```
. use ${ITSUS_DATA}/monthly, clear
(Monthly data for ITSUS)
. tsset
        time variable:  date, Jan 13 to Mar 12
                delta:  1 month
. generate lcpi = log(cpi)
. generate anncpi = 100*S12.lcpi
(12 missing values generated)
. label variable anncpi "Annual CPI inflation"
. generate inflation = 1200*D.lcpi
(1 missing value generated)
. label variable inflation "Monthly CPI inflation"
```

12. It's tempting to interpret the $\delta(L)$ as the AR terms for the variance and the $\alpha(L)$ as the MA terms, but that turns out not to be the case. In fact, if ϵ_t follows a GARCH(r, m) process, then ϵ_t^2 follows an ARMA(p, r) process where p is the larger of r and m. See Hamilton (1994) for details.

```
. twoway (line inflation date) (connected anncpi date, msymbol(+) msize(small)),
> yline(0)
```

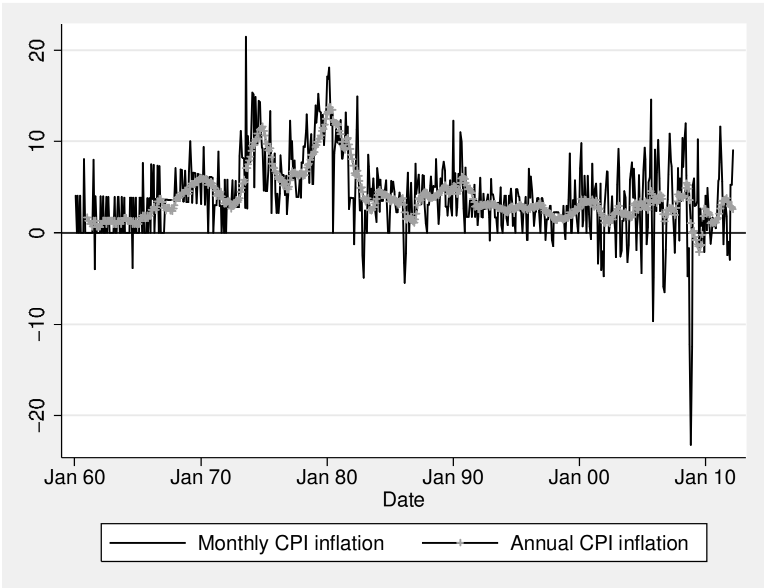

Figure 8.7. Monthly and annual consumer price inflation in the United States

The oscillations in monthly inflation provide an impressionistic sense of the volatility in the series. The level of inflation is highest in the 1970s and early 1980s, during the era of stagflation, but it appears that the volatility may be highest near the end of the series, around the time that deflation occurs.

To save time, we jump straight to the ARIMA and ARCH models for monthly inflation without demonstrating how we landed on this parameterization. (You should work through the Box–Jenkins steps of identifying, fitting, and testing this model to see if you agree with our choices.)

```
. arch inflation, arima(2,0,1) arch(1 2 3) nolog
ARCH family regression -- ARMA disturbances
Sample: Feb 60 - Mar 12                          Number of obs    =        626
Distribution: Gaussian                           Wald chi2(3)     =   65444.26
Log likelihood = -1612.692                       Prob > chi2      =     0.0000

                              OPG
   inflation       Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]

inflation
       _cons    3.522882   .6551663     5.38   0.000      2.23878    4.806985

ARMA
          ar
         L1.    1.264746   .0568226    22.26   0.000     1.153376    1.376116
         L2.   -.2781428   .0536782    -5.18   0.000    -.3833501   -.1729355

          ma
         L1.   -.9146949   .0226525   -40.38   0.000     -.959093   -.8702968

ARCH
        arch
         L1.    .3222899   .0724758     4.45   0.000     .1802399      .46434
         L2.    .1454913   .0430307     3.38   0.001     .0611527    .2298299
         L3.    .0218814   .0388328     0.56   0.573    -.0542295    .0979922

       _cons    6.168393   .5877223    10.50   0.000     5.016479    7.320308

. predict avar, variance
. label variable avar "ARCH conditional variance"
```

We have fit an ARMA(2,1) model with an ARCH(3) process for the conditional variance. (A fourth-order ARCH term is insignificant.) We have also used Stata's `predict` command to save the ARCH estimate of the conditional variance, in the variable `avar`, for later reference.

As a practical matter, we tend to avoid processes higher than second order, so we are a little suspicious of this ARCH(3) model.[13] It turns out that a GARCH(1,1) model fits these data equally well.

13. As we emphasized in chapter 3, modeling and specification choices for time series should be founded as much as possible on an understanding of the process that generates the data. We are modeling monthly inflation estimates based on the CPI published by the United States Bureau of Labor Statistics. With monthly financial data, we often find a difference between end-of-quarter observations and intraquarter observations. End-of-quarter observations often reflect a more rigorous review process because these observations are used in audited quarterly financial statements. While the CPI is published by a government agency rather than a private corporation, there may be some feature of the data-collection process that spreads random increments to information across a quarter. Alternatively, there may be a seasonal adjustment process that affects the dynamic behavior of volatility. Our purpose here is to illustrate Stata's implementation of GARCH, so we will not explore these possibilities further. However, you should review these possibilities carefully in any real research project.

```
. arch inflation, arima(2,0,1) arch(1) garch(1) nolog
ARCH family regression -- ARMA disturbances
Sample: Feb 60 - Mar 12                         Number of obs    =       626
Distribution: Gaussian                          Wald chi2(3)     =  51266.27
Log likelihood = -1603.898                      Prob > chi2      =    0.0000

                            OPG
   inflation       Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]

inflation
       _cons    3.441357   .6031881     5.71   0.000      2.25913    4.623584

ARMA
          ar
         L1.    1.225343   .0491122    24.95   0.000     1.129085    1.321601
         L2.    -.240727   .0462019    -5.21   0.000     -.331281   -.1501731

          ma
         L1.   -.9060132   .0242611   -37.34   0.000    -.9535641   -.8584623

ARCH
        arch
         L1.    .1976492   .0391789     5.04   0.000       .12086    .2744384

       garch
         L1.    .7181568    .051498    13.95   0.000     .6172227     .819091

       _cons    1.065144   .2871044     3.71   0.000     .5024296    1.627858

. predict gvar, variance
. label variable gvar "GARCH conditional variance"
. correlate gvar avar
(obs=627)

                 gvar     avar

        gvar   1.0000
        avar   0.8674   1.0000
```

The ARCH and GARCH specifications produce similar estimates of the conditional variance. The GARCH model just uses a more parsimonious specification to obtain its estimate.[14]

14. Admittedly, in this example, we reduce the number of parameters by only one. The point is the GARCH model can provide significant reductions in the model degrees of freedom with little or no reduction in model fit in many cases.

Figure 8.8 displays a scatterplot of the ARCH and GARCH conditional variance estimates.[15] This figure provides a graphical confirmation of the high degree of association between the two estimates, although the relationship is not precisely linear.[16]

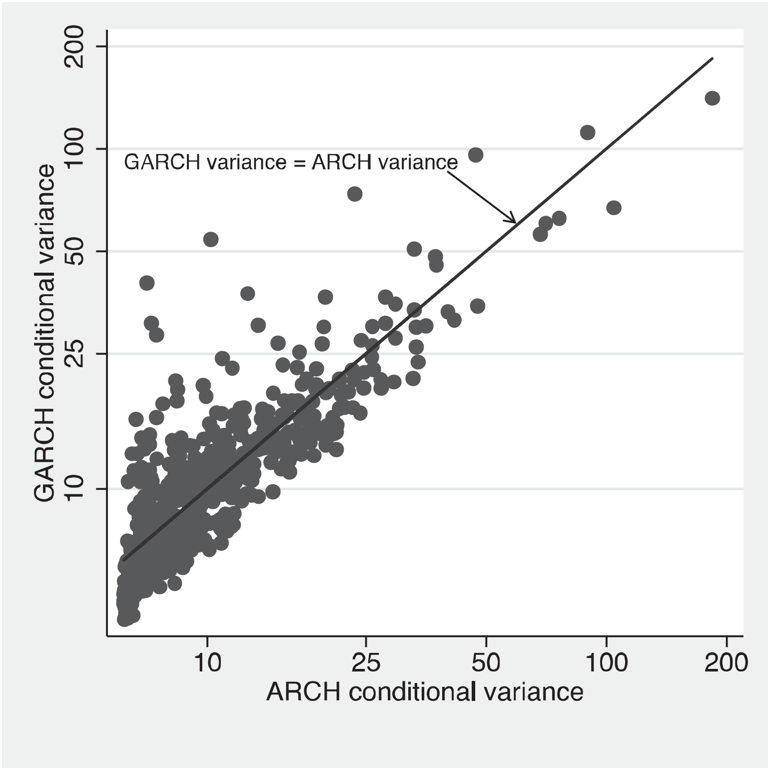

Figure 8.8. Comparison of GARCH and ARCH estimates of conditional variance

15. Both series are plotted on a log scale to improve the readability of the graph.
16. If you take a closer look at subsets of the time line, you will see that the GARCH estimates of the conditional variance have more "inertia" than the ARCH estimates; that is, they take longer to fall back from the peak levels of conditional variance.

Figure 8.9 displays the GARCH estimate of the conditional variance of monthly inflation in the U.S. CPI. As we suspected, the volatility increased dramatically during the global financial crisis in late 2008 and early 2009.

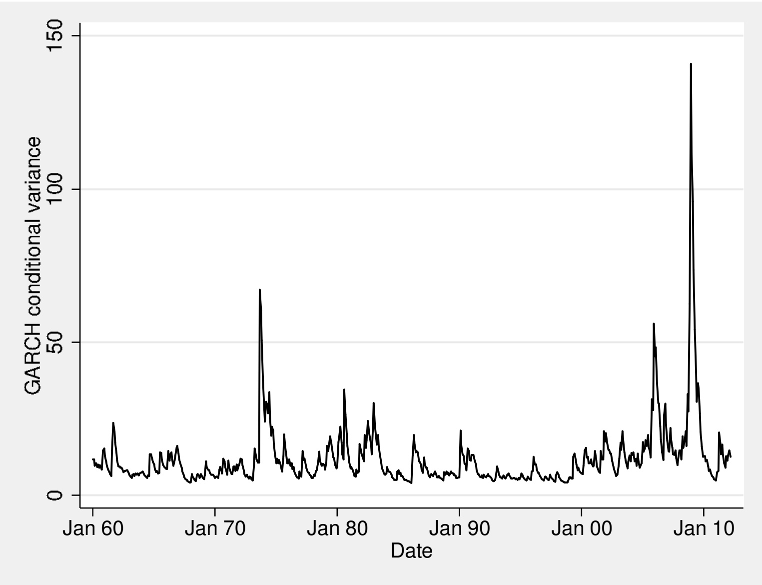

Figure 8.9. Conditional variance of monthly consumer inflation

It turns out that a GARCH(1,1) specification is flexible enough to handle a wide variety of processes. In this example, higher-order ARCH and GARCH terms are negative (which violates our assumptions), statistically insignificant, or both.

8.3.2 Other extensions

The GARCH model provides a remarkably useful technical solution (a low-order flexible functional form) to a technical problem (the need for too many terms in a pure ARCH model). While many, many other extensions to the ARCH model have been proposed, most of them are designed to capture one or another empirical regularity of financial time series that the pure ARCH model does not handle.

Asymmetric responses to "news"

In financial markets, bad news (for instance, the unexpected failure of a major financial institution) sometimes has a larger impact on volatility than good news. The models we have discussed so far cannot capture this feature. The arrival of news is represented by a nonzero value of ν_t, the innovation in the ARCH process. The magnitude of the response to news is proportional to $|\nu_t|$; that is, it depends on the size but not the sign of ν_t.

The exponential GARCH (EGARCH) model introduced by Nelson (1991) provides a specification that allows an asymmetric response to positive and negative news. In the EGARCH model, h_t, the time-varying scale factor for the conditional variance, is modeled as

$$\log h_t = \zeta + \sum_{j=1}^{\infty} \pi_j \left(|\nu_{t-j}| - E|\nu_{t-j}| + \aleph\nu_{t-j}\right)$$

The first two terms in the parentheses—$(|\nu_{t-j}| - E|\nu_{t-j}|)$—indicate that unexpected events ("news") tend to increase volatility temporarily. If $\pi_j > 0$, a larger-than-expected $|\nu_{t-j}|$ increases the variance of ϵ_t. The third term—$\aleph\nu_{t-j}$—allows positive and negative news to affect volatility differently. If $\aleph = 0$, there is no asymmetry—positive shocks have the same impact as negative shocks of the same magnitude. If $-1 < \aleph < 0$, positive shocks have a smaller impact than negative shocks. If $\aleph < -1$, the effect is reversed.

This parameterization of the EGARCH model follows Hamilton (1994). The notation and implementation in Stata is a bit different. Stata denotes the π's as γ's. And Stata expands $\aleph$ into a series of α's, one for each lag of ν_t. I know it's confusing. There is a staggering variety of ARCH-like models, and each author has proposed a slightly different way of summarizing and contrasting them.

Note that the EGARCH model specifies a process for $\log h_t$ rather than h_t. As a consequence, the variance h_t is always positive regardless of the signs of the π_j coefficients; thus no restrictions need to be imposed on the estimation of these parameters. ϵ_t, h_t, and $\log h_t$ are all stationary if $\sum_{j=1}^{\infty} \pi_j^2 < \infty$. Moreover, just as in the GARCH model, a high-order $\pi(L)$ can be reparameterized as the ratio of two lower-order lag polynomials, which results in the model

$$\{1 - \delta(L)\} \log h_t = \zeta + \alpha(L)(|\nu_{t-j}| - E|\nu_{t-j}| + \aleph\nu_{t-j})$$

Stata's implementation of EGARCH incorporates this extension. The `earch()` option specifies the terms in $\alpha(L)$, while the `egarch()` option specifies the terms in $\delta(L)$.

Let's test for asymmetric response in our model of monthly consumer price inflation. We will start with a model that includes only the α and $\aleph$ terms. As in our first ARCH specification, a third-order process is required to characterize the time-varying volatility.

```
. arch inflation, arima(2,0,1) earch(1 2 3) nolog
ARCH family regression -- ARMA disturbances
Sample: Feb 60 - Mar 12                          Number of obs    =       626
Distribution: Gaussian                           Wald chi2(3)     =  72158.09
Log likelihood = -1609.768                       Prob > chi2      =    0.0000
```

inflation	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
inflation						
_cons	3.50312	.6382905	5.49	0.000	2.252094	4.754147
ARMA						
ar						
L1.	1.267458	.053295	23.78	0.000	1.163001	1.371914
L2.	-.2804102	.0500037	-5.61	0.000	-.3784156	-.1824048
ma						
L1.	-.9209155	.0218292	-42.19	0.000	-.9636998	-.8781311
ARCH						
earch						
L1.	-.0264707	.0584792	-0.45	0.651	-.1410879	.0881464
L2.	.0998417	.0426097	2.34	0.019	.0163283	.1833551
L3.	-.0431963	.0551584	-0.78	0.434	-.1513048	.0649123
earch_a						
L1.	.5837489	.0947084	6.16	0.000	.3981238	.7693739
L2.	.3568973	.0714131	5.00	0.000	.2169301	.4968644
L3.	.1469461	.0890959	1.65	0.099	-.0276785	.3215708
_cons	2.335611	.0695702	33.57	0.000	2.199256	2.471966

The `earch_a` rows display the estimated coefficients of $\alpha(L)$, that is, the symmetric impacts of the $|\nu_{t-j}| - E|\nu_{t-j}|$. The `earch` rows display the estimates of the asymmetric impacts. In this example, there is mixed evidence of asymmetric responses to volatility shocks—only the second lag is statistically significant. Just to be sure, we test them jointly.

```
. test [ARCH]:L.earch L2.earch L3.earch
 ( 1)  [ARCH]L.earch = 0
 ( 2)  [ARCH]L2.earch = 0
 ( 3)  [ARCH]L3.earch = 0
           chi2(  3) =    7.29
         Prob > chi2 =    0.0631
```

Unfortunately, not definitive.

As we saw above, in our discussion of the GARCH model, we can replace the third-order ARCH specification with a GARCH(1,1) model. We use the **egarch()** option to effect a similar change here.[17] With this specification, there clearly is no evidence of an asymmetric response to shocks.

```
. arch inflation, arima(2,0,1) earch(1) egarch(1) nolog
ARCH family regression -- ARMA disturbances

Sample: Feb 60 - Mar 12                         Number of obs   =        626
Distribution: Gaussian                          Wald chi2(3)    =   54463.80
Log likelihood = -1603.061                      Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
   inflation |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
inflation    |
       _cons |   3.641849   .6782711     5.37   0.000     2.312462    4.971235
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   1.255739   .0495531    25.34   0.000     1.158616    1.352861
         L2. |  -.2690102   .0462552    -5.82   0.000    -.3596687   -.1783518
             |
          ma |
         L1. |  -.9124459   .0247751   -36.83   0.000    -.9610042   -.8638876
-------------+----------------------------------------------------------------
ARCH         |
       earch |
         L1. |    .034305   .0289041     1.19   0.235    -.0223459     .090956
             |
     earch_a |
         L1. |    .317302   .0453016     7.00   0.000     .2285124    .4060915
             |
      egarch |
         L1. |   .9040205   .0275583    32.80   0.000     .8500072    .9580338
             |
       _cons |   .2277512   .0629302     3.62   0.000     .1044103    .3510921
------------------------------------------------------------------------------
```

Variations in volatility affect the mean of the observable series

In some cases, there may be reason to believe the mean of the observable variable, y_t, is associated with the conditional variance of the random innovation. For instance, some economists have speculated that high inflation tends to be accompanied by high levels of price variability.[18] In finance, most theories suggest that investors require a

17. In preliminary estimates of this EGARCH model, with a slightly different sample, the `arch` command failed to converge. Models in the ARCH family frequently are difficult to estimate. It may be difficult or impossible to maximize the likelihood function. If you encounter these difficulties, Stata offers some help in the form of options that can control the maximization process. These tools cannot guarantee a solution, and they require a fair degree of sophistication (and luck) to use them successfully. Fortunately, we did not require these options with our final sample.
18. Interestingly, our estimates in this chapter appear to suggest the opposite—the conditional variance is highest when deflation is a real possibility. Even so, there seems to be a relationship between the level of inflation and the conditional variance.

higher return when the perceived risk (that is, variance) of an asset is higher. These theories motivated Engle, Lilien, and Robins (1987) to propose the ARCH-in-mean, or ARCH-M, model. In this model, the conditional variance, h_t, appears as a regressor in the equation for y_t.

To fit an ARCH-M model in Stata, add the `archm` option. For instance,

```
arch inflation, arima(2,0,1) arch(1) garch(1) archm
```

would add an ARCH-M term to our GARCH(1,1) model of monthly inflation. (Try it. The ARCH-M term is insignificant.) Alternatively, Stata allows you to specify an arbitrary lag polynomial in ARCH-M terms with the `archmlags()` option. Finally, you can add an arbitrary function of the conditional variance to the equation for y_t with the `archmexp()` option.

Nonnormal errors

We noted above that the distributions of asset prices possess "fat tails"; that is, extreme outcomes occur more frequently than is predicted by the normal distribution. You can use the `distribution()` option to specify that the random errors follow a t distribution or a generalized error distribution instead of a normal distribution. Stata estimates the degrees of freedom for the t distribution and the shape parameter for the generalized error distribution, but you can override those estimates by adding a legal value to the `distribution` option; for instance, `distribution(ged 3)` forces Stata to set the shape parameter to 3.

Odds and ends

We have just scratched the surface of proposed models of time-varying volatility. We have not even covered all the models available in Stata. The table below is a concise attempt to fill that gap. It lists issues that you might be facing, such as asymmetric responses, and lists the related models Stata offers through the `arch` command. The number of choices can be overwhelming. As always, you are on solid ground if your choice of model is tied to knowledge of the dynamic process generating your data.

Table 8.1. Extensions to the ARCH model and `arch` command

Reason	Approach	Models	Options
ARCH specification alone requires too many lags	Combine ARCH terms (lags of ϵ_t^2) with GARCH terms (lags of σ_t^2), the conditional variance	GARCH	`garch(`*numlist*`)`
Volatility responds asymmetrically to positive and negative “news”	Add terms that differentiate the impact of positive and negative “news”	simple asymmetric ARCH	`saarch(`*numlist*`)`
		EGARCH	`earch(`*numlist*`)` `egarch(`*numlist*`)`
		threshold ARCH	`abarch(`*numlist*`)` `atarch(`*numlist*`)` `sdgarch(`*numlist*`)`
		Glosten, Jagannathan, and Runkle	`tarch(`*numlist*`)`
Linear combination of lagged ϵ_t^2 terms not adequate	Introduce nonlinearity by shifting the minimum impact of the ith term from $\epsilon_{t-i} = 0$ to $\epsilon_{t-i} = \kappa_i$	nonlinear ARCH	`narch(`*numlist*`)`
		nonlinear power ARCH with one shift	`narchk(`*numlist*`)`
The conditional mean of y_t depends on σ_t^2, the conditional variance	Add current and lagged values of σ_t^2 to the state equation	ARCH in mean	`archm(`*numlist*`)` or `archmlags(`*numlist*`)`
Model in a power function of ϵ_t rather than ϵ_t^2 seems more appropriate	Power function versions of most of the previous specifications	power ARCH	`parch(`*numlist*`)`
		asymmetric power ARCH	`aparch(`*numlist*`)`
		nonlinear power ARCH or nonlinear power ARCH with one shift	`nparch(`*numlist*`)` or `nparchk(`*numlist*`)`
Innovations may not follow a normal distribution	Specify a different error distribution		`distribution(`*dist* [#]`)`
Observable variables influence the conditional variance	Add observable variables to the conditional variance equation		`het(`*varlist*`)`

8.4 Points to remember

- Researchers have long noted what appear to be episodes of relatively high and relatively low volatility (that is, conditional variance), especially in economic and financial time series. The standard ARIMA model assumes constant residual variance; hence, it does not capture this phenomenon.
- The ARCH model of Engle proposed an AR structure for the conditional variance of a time series while retaining the constant unconditional variance. This model provided the first practical method for modeling time-varying volatility, and its appearance triggered an explosion in both empirical and theoretical research on time-varying volatility.
- Many extensions to Engle's ARCH specification have been proposed and used successfully. Bollerslev's GARCH model provides a flexible way of fitting the conditional variance with relatively few parameters. Other proposals (exponential GARCH, ARCH-M, etc.) extend the model to capture additional empirical regularities.
- Models in the ARCH family often are difficult to estimate, and you may need to experiment with some of the parameters that control Stata's maximization algorithm to obtain results.

Stata commands and features discussed

`arch` ([TS] **arch**): ARCH family of estimators; sections 8.2 and 8.3

`egen` ([D] **egen**): Extensions to generate; section 8.1

`estat archlm` ([R] **regress postestimation time series**): Test for ARCH effects in the residuals; section 8.2

`predict` ([TS] **arch postestimation**): Postestimation tools for arch; section 8.3

`table` ([R] **table**): Tables of summary statistics; section 8.1

`test` ([TS] **arch postestimation**): Postestimation tools for arch; section 8.3