# 21 Time Series Models of Heteroskedasticity

## 21.1. *Autoregressive Conditional Heteroskedasticity* (ARCH)

An autoregressive process of order $p$ (denoted $AR(p)$) for an observed variable $y_t$ takes the form

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t, \qquad [21.1.1]$$

where $u_t$ is white noise:

$$E(u_t) = 0 \qquad [21.1.2]$$

$$E(u_t u_\tau) = \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases} \qquad [21.1.3]$$

The process is covariance-stationary provided that the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

are outside the unit circle. The optimal linear forecast of the level of $y_t$ for an $AR(p)$ process is

$$\hat{E}(y_t|y_{t-1}, y_{t-2}, \ldots) = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}, \qquad [21.1.4]$$

where $\hat{E}(y_t|y_{t-1}, y_{t-2}, \ldots)$ denotes the linear projection of $y_t$ on a constant and $(y_{t-1}, y_{t-2}, \ldots)$. While the conditional mean of $y_t$ changes over time according to [21.1.4], provided that the process is covariance-stationary, the unconditional mean of $y_t$ is constant:

$$E(y_t) = c/(1 - \phi_1 - \phi_2 - \cdots - \phi_p).$$

Sometimes we might be interested in forecasting not only the level of the series $y_t$ but also its variance. For example, Figure 21.1 plots the federal funds rate, which is an interest rate charged on overnight loans from one bank to another. This interest rate has been much more volatile at some times than at others. Changes in the variance are quite important for understanding financial markets, since investors require higher expected returns as compensation for holding riskier assets. A variance that changes over time also has implications for the validity and efficiency of statistical inference about the parameters $(c, \phi_1, \phi_2, \ldots, \phi_p)$ that describe the dynamics of the level of $y_t$.

Although [21.1.3] implies that the unconditional variance of $u_t$ is the constant $\sigma^2$, the conditional variance of $u_t$ could change over time. One approach is to

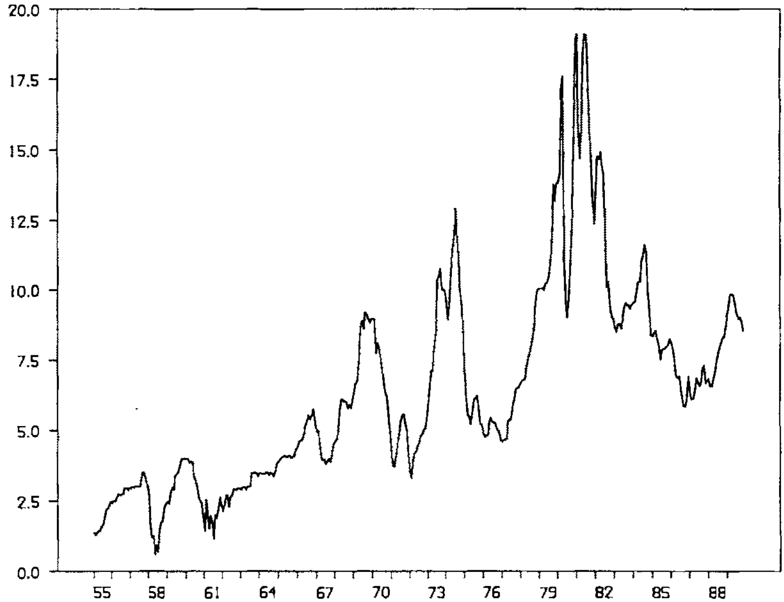


**FIGURE 21.1** U.S. federal funds rate (monthly averages quoted at an annual rate), 1955–89.

describe the square of $u_t$ as itself following an $AR(m)$ process:

$$u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + w_t, \qquad [21.1.5]$$

where $w_t$ is a new white noise process:

$$E(w_t) = 0$$

$$E(w_t w_\tau) = \begin{cases} \lambda^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Since $u_t$ is the error in forecasting $y_t$, expression [21.1.5] implies that the linear projection of the squared error of a forecast of $y_t$ on the previous $m$ squared forecast errors is given by

$$\hat{E}(u_t^2 | u_{t-1}^2, u_{t-2}^2, \ldots) = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \qquad [21.1.6]$$

A white noise process $u_t$ satisfying [21.1.5] is described as an *autoregressive conditional heteroskedastic* process of order $m$, denoted $u_t \sim ARCH(m)$. This class of processes was introduced by Engle (1982).[1]

Since $u_t$ is random and $u_t^2$ cannot be negative, this can be a sensible representation only if [21.1.6] is positive and [21.1.5] is nonnegative for all realizations of $\{u_t\}$. This can be ensured if $w_t$ is bounded from below by $-\zeta$ with $\zeta > 0$ and if $\alpha_j \geq 0$ for $j = 1, 2, \ldots, m$. In order for $u_t^2$ to be covariance-stationary, we further require that the roots of

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0$$

[1]A nice survey of *ARCH*-related models was provided by Bollerslev, Chou, and Kroner (1992).

lie outside the unit circle. If the $\alpha_j$ are all nonnegative, this is equivalent to the requirement that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m < 1. \qquad [21.1.7]$$

When these conditions are satisfied, the unconditional variance of $u_t$ is given by

$$\sigma^2 = E(u_t^2) = \zeta/(1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m). \qquad [21.1.8]$$

Let $\hat{u}^2_{t+s|t}$ denote an $s$-period-ahead linear forecast:

$$\hat{u}^2_{t+s|t} = \hat{E}(u^2_{t+s}|u_t^2, u_{t-1}^2, \ldots).$$

This can be calculated as in [4.2.27] by iterating on

$$(\hat{u}^2_{t+j|t} - \sigma^2) = \alpha_1(\hat{u}^2_{t+j-1|t} - \sigma^2) + \alpha_2(\hat{u}^2_{t+j-2|t} - \sigma^2) + \cdots + \alpha_m(\hat{u}^2_{t+j-m|t} - \sigma^2)$$

for $j = 1, 2, \ldots, s$ where

$$\hat{u}^2_{\tau|t} = u_\tau^2 \qquad \text{for } \tau \leq t.$$

The $s$-period-ahead forecast $\hat{u}^2_{t+s|t}$ converges in probability to $\sigma^2$ as $s \to \infty$, assuming that $w_t$ has finite variance and that [21.1.7] is satisfied.

It is often convenient to use an alternative representation for an $ARCH(m)$ process that imposes slightly stronger assumptions about the serial dependence of $u_t$. Suppose that

$$u_t = \sqrt{h_t} \cdot v_t, \qquad [21.1.9]$$

where $\{v_t\}$ is an i.i.d. sequence with zero mean and unit variance:

$$E(v_t) = 0 \qquad E(v_t^2) = 1.$$

If $h_t$ evolves according to

$$h_t = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2, \qquad [21.1.10]$$

then [21.1.9] implies that

$$E(u_t^2|u_{t-1}, u_{t-2}, \ldots) = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \qquad [21.1.11]$$

Hence, if $u_t$ is generated by [21.1.9] and [21.1.10], then $u_t$ follows an $ARCH(m)$ process in which the linear projection [21.1.6] is also the conditional expectation.

Notice further that when [21.1.9] and [21.1.10] are substituted into [21.1.5], the result is

$$h_t \cdot v_t^2 = h_t + w_t.$$

Hence, under the specification in [21.1.9], the innovation $w_t$ in the $AR(m)$ representation for $u_t^2$ in [21.1.5] can be expressed as

$$w_t = h_t \cdot (v_t^2 - 1). \qquad [21.1.12]$$

Note from [21.1.12] that although the unconditional variance of $w_t$ was assumed to be constant,

$$E(w_t^2) = \lambda^2, \qquad [21.1.13]$$

the conditional variance of $w_t$ changes over time.

The unconditional variance of $w_t$ reflects the fourth moment of $u_t$, and this fourth moment does not exist for all stationary $ARCH$ models. One can see this by squaring [21.1.12] and calculating the unconditional expectation of both sides:

$$E(w_t^2) = E(h_t^2) \cdot E(v_t^2 - 1)^2. \qquad [21.1.14]$$

Taking the $ARCH(1)$ specification for illustration, we find with a little manipulation of the formulas for the mean and variance of an $AR(1)$ process that

$$\begin{aligned} E(h_t^2) &= E(\zeta + \alpha_1 u_{t-1}^2)^2 \\ &= E\{(\alpha_1^2 \cdot u_{t-1}^4) + (2\alpha_1\zeta \cdot u_{t-1}^2) + \zeta^2\} \\ &= \alpha_1^2 \cdot [\mathrm{Var}(u_{t-1}^2) + [E(u_{t-1}^2)]^2] + 2\alpha_1\zeta \cdot E(u_{t-1}^2) + \zeta^2 \qquad [21.1.15] \\ &= \alpha_1^2 \cdot \left[\frac{\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}\right] + \frac{2\alpha_1\zeta^2}{1-\alpha_1} + \zeta^2 \\ &= \frac{\alpha_1^2\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}. \end{aligned}$$

Substituting [21.1.15] and [21.1.13] into [21.1.14], we conclude that $\lambda^2$ (the unconditional variance of $w_t$) must satisfy

$$\lambda^2 = \left[\frac{\alpha_1^2\lambda^2}{1-\alpha_1^2} + \frac{\zeta^2}{(1-\alpha_1)^2}\right] \times E(v_t^2 - 1)^2. \qquad [21.1.16]$$

Even when $|\alpha_1| < 1$, equation [21.1.16] may not have any real solution for $\lambda$. For example, if $v_t \sim N(0, 1)$, then $E(v_t^2 - 1)^2 = 2$ and [21.1.16] requires that

$$\frac{(1 - 3\alpha_1^2)\lambda^2}{1 - \alpha_1^2} = \frac{2\zeta^2}{(1 - \alpha_1)^2}.$$

This equation has no real solution for $\lambda$ whenever $\alpha_1^2 \geq \frac{1}{3}$. Thus, if $u_t \sim ARCH(1)$ with the innovations $v_t$ in [21.1.9] coming from a Gaussian distribution, then the second moment of $w_t$ (or the fourth moment of $u_t$) does not exist unless $\alpha_1^2 < \frac{1}{3}$.

### *Maximum Likelihood Estimation with Gaussian* $v_t$

Suppose that we are interested in estimating the parameters of a regression model with $ARCH$ disturbances. Let the regression equation be

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t. \qquad [21.1.17]$$

Here $\mathbf{x}_t$ denotes a vector of predetermined explanatory variables, which could include lagged values of $y$. The disturbance term $u_t$ is assumed to satisfy [21.1.9] and [21.1.10]. It is convenient to condition on the first $m$ observations ($t = -m + 1, -m + 2, \ldots, 0$) and to use observations $t = 1, 2, \ldots, T$ for estimation. Let $\mathcal{Y}_t$ denote the vector of observations obtained through date $t$:

$$\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{-m+1}, \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots, \mathbf{x}_1', \mathbf{x}_0', \ldots, \mathbf{x}_{-m+1}')'.$$

If $v_t \sim$ i.i.d. $N(0, 1)$ with $v_t$ independent of both $\mathbf{x}_t$ and $\mathcal{Y}_{t-1}$, then the conditional distribution of $y_t$ is Gaussian with mean $\mathbf{x}_t'\boldsymbol{\beta}$ and variance $h_t$:

$$f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{-(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{2h_t}\right), \qquad [21.1.18]$$

where

$$\begin{aligned} h_t &= \zeta + \alpha_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots \\ &\quad + \alpha_m(y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2 \qquad [21.1.19] \\ &= [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta} \end{aligned}$$

for

$$\delta \equiv (\zeta, \alpha_1, \alpha_2, \ldots, \alpha_m)'$$
$$[\mathbf{z}_t(\boldsymbol{\beta})]' \equiv [1, (y_{t-1} - \mathbf{x}'_{t-1}\boldsymbol{\beta})^2, (y_{t-2} - \mathbf{x}'_{t-2}\boldsymbol{\beta})^2, \ldots, (y_{t-m} - \mathbf{x}'_{t-m}\boldsymbol{\beta})^2].$$

Collect the unknown parameters to be estimated in an $(a \times 1)$ vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \boldsymbol{\delta}')'.$$

The sample log likelihood conditional on the first $m$ observations is then

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{t=1}^{T} \log f(y_t | \mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) \\ &= -(T/2)\log(2\pi) - (1/2)\sum_{t=1}^{T}\log(h_t) \\ &\quad - (1/2)\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2/h_t. \end{aligned} \qquad [21.1.20]$$

For a given numerical value for the parameter vector $\boldsymbol{\theta}$, the sequence of conditional variances can be calculated from [21.1.19] and used to evaluate the log likelihood function [21.1.20]. This can then be maximized numerically using the methods described in Section 5.7. The derivative of the log of the conditional likelihood of the $t$th observation with respect to the parameter vector $\boldsymbol{\theta}$, known as the $t$th score, is shown in Appendix 21.A to be given by

$$\begin{aligned} \underset{(a\times 1)}{\mathbf{s}_t(\boldsymbol{\theta})} &= \frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \{(u_t^2 - h_t)/(2h_t^2)\}\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix} + \begin{bmatrix} (\mathbf{x}_t u_t)/h_t \\ \mathbf{0} \end{bmatrix}. \end{aligned} \qquad [21.1.21]$$

The likelihood function can be maximized using the method of scoring as in Engle (1982, p. 997) or using the Berndt, Hall, Hall, and Hausman (1974) algorithm as in Bollerslev (1986, p. 317). Alternatively, the gradient of the log likelihood function can be calculated analytically from the sum of the scores,

$$\nabla\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} s_t(\boldsymbol{\theta}),$$

or numerically by numerical differentiation of the log likelihood [21.1.20]. The analytically or numerically evaluated gradient could then be used with any of the numerical optimization procedures described in Section 5.7.

Imposing the stationarity condition $(\Sigma_{j=1}^{m}\alpha_j < 1)$ and the nonnegativity condition $(\alpha_j \geq 0$ for all $j)$ can be difficult in practice. Typically, either the value of $m$ is very small or else some ad hoc structure is imposed on the sequence $\{\alpha_j\}_{j=1}^{m}$ as in Engle (1982, equation (38)).

### *Maximum Likelihood Estimation with Non-Gaussian* $v_t$

The preceding formulation of the likelihood function assumed that $v_t$ has a Gaussian distribution. However, the unconditional distribution of many financial time series seems to have fatter tails than allowed by the Gaussian family. Some of this can be explained by the presence of *ARCH*; that is, even if $v_t$ in [21.1.9] has a Gaussian distribution, the unconditional distribution of $u_t$ is non-Gaussian with heavier tails than a Gaussian distribution (see Milhøj, 1985, or Bollerslev, 1986, p. 313). Even so, there is a fair amount of evidence that the conditional distribution of $u_t$ is often non-Gaussian as well.

The same basic approach can be used with non-Gaussian distributions. For example, Bollerslev (1987) proposed that $v_t$ in [21.1.9] might be drawn from a $t$ distribution with $\nu$ degrees of freedom, where $\nu$ is regarded as a parameter to be estimated by maximum likelihood. If $u_t$ has a $t$ distribution with $\nu$ degrees of freedom and scale parameter $M_t$, then its density is given by

$$f(u_t) = \frac{\Gamma[(\nu + 1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} M_t^{-1/2}\left[1 + \frac{u_t^2}{M_t\nu}\right]^{-(\nu+1)/2}, \qquad [21.1.22]$$

where $\Gamma(\cdot)$ is the gamma function described in the discussion following equation [12.1.18]. If $\nu > 2$, then $v_t$ has mean zero and variance[2]

$$E(u_t^2) = M_t\nu/(\nu - 2).$$

Hence, a $t$ variable with $\nu$ degrees of freedom and variance $h_t$ is obtained by taking the scale parameter $M_t$ to be

$$M_t = h_t(\nu - 2)/\nu,$$

for which the density [21.1.22] becomes

$$f(u_t) = \frac{\Gamma[(\nu + 1)/2]}{\pi^{1/2}\Gamma(\nu/2)} (\nu - 2)^{-1/2} h_t^{-1/2}\left[1 + \frac{u_t^2}{h_t(\nu - 2)}\right]^{-(\nu+1)/2}. \qquad [21.1.23]$$

This density can be used in place of the Gaussian specification [21.1.18] along with the same specification of the conditional mean and conditional variance used in [21.1.17] and [21.1.19]. The sample log likelihood conditional on the first $m$ observations then becomes

$$\begin{aligned}
&\sum_{t=1}^{T} \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta}) \\
&\quad = T \log\left\{\frac{\Gamma[(\nu + 1)/2]}{\pi^{1/2}\Gamma(\nu/2)} (\nu - 2)^{-1/2}\right\} - (1/2) \sum_{t=1}^{T} \log(h_t) \qquad [21.1.24] \\
&\quad\quad - [(\nu + 1)/2] \sum_{t=1}^{T} \log\left[1 + \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{h_t(\nu - 2)}\right],
\end{aligned}$$

where

$$\begin{aligned}
h_t &= \zeta + \alpha_1(y_{t-1} - \mathbf{x}_{t-1}'\boldsymbol{\beta})^2 + \alpha_2(y_{t-2} - \mathbf{x}_{t-2}'\boldsymbol{\beta})^2 + \cdots + \alpha_m(y_{t-m} - \mathbf{x}_{t-m}'\boldsymbol{\beta})^2 \\
&= [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta}.
\end{aligned}$$

The log likelihood [21.1.24] is then maximized numerically with respect to $\nu$, $\boldsymbol{\beta}$, and $\boldsymbol{\delta}$ subject to the constraint $\nu > 2$.

The same approach can be used with other distributions for $v_t$. Other distributions that have been employed with *ARCH*-related models include a Normal-Poisson mixture distribution (Jorion, 1988), power exponential distribution (Baillie and Bollerslev, 1989), Normal–log normal mixture (Hsieh, 1989), generalized exponential distribution (Nelson, 1991), and serially dependent mixture of Normals (Cai, forthcoming) or $t$ variables (Hamilton and Susmel, forthcoming).

[2]See, for example, DeGroot (1970, p. 42).

*Quasi-Maximum Likelihood Estimation*

Even if the assumption that $v_t$ is i.i.d. $N(0, 1)$ is invalid, we saw in [21.1.6] that the *ARCH* specification can still offer a reasonable model on which to base a linear forecast of the squared value of $v_t$. As shown in Weiss (1984, 1986), Bollerslev and Wooldridge (1992), and Glosten, Jagannathan, and Runkle (1989), maximization of the Gaussian log likelihood function [21.1.20] can provide consistent estimates of the parameters $\zeta, \alpha_1, \alpha_2, \ldots, \alpha_m$ of this linear representation even when the distribution of $u_t$ is non-Gaussian, provided that $v_t$ in [21.1.9] satisfies

$$E(v_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = 0$$

and

$$E(v_t^2|\mathbf{x}_t, \mathcal{Y}_{t-1}) = 1.$$

However, the standard errors have to be adjusted. Let $\hat{\boldsymbol{\theta}}_T$ be the estimate that maximizes the Gaussian log likelihood [21.1.20], and let $\boldsymbol{\theta}$ be the true value that characterizes the linear representations [21.1.9], [21.1.17], and [21.1.19]. Then even when $v_t$ is actually non-Gaussian, under certain regularity conditions

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1}),$$

where

$$\mathbf{S} = \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} [\mathbf{s}_t(\boldsymbol{\theta})]\cdot[\mathbf{s}_t(\boldsymbol{\theta})]'$$

for $\mathbf{s}_t(\boldsymbol{\theta})$ the score vector as calculated in [21.1.21], and where

$$\begin{aligned}
\mathbf{D} &= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} -E\left\{\frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\bigg| \mathbf{x}_t, \mathcal{Y}_{t-1}\right\} \\
&= \operatorname*{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left\{[1/(2h_t^2)]\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}\right. \qquad [21.1.25] \\
&\qquad \times \left.\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}'_{t-j} & [\mathbf{z}_t(\boldsymbol{\beta})]' \end{bmatrix} + (1/h_t)\begin{bmatrix} \mathbf{x}_t\mathbf{x}'_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right\},
\end{aligned}$$

where

$$\mathcal{Y}_t = (y_t, y_{t-1}, \ldots, y_1, y_0, \ldots, y_{-m+1}, \mathbf{x}'_t, \mathbf{x}'_{t-1}, \ldots, \mathbf{x}'_1, \mathbf{x}'_0, \ldots, \mathbf{x}'_{-m+1})'.$$

The second equality in [21.1.25] is established in Appendix 21.A. The matrix $\mathbf{S}$ can be consistently estimated by

$$\hat{\mathbf{S}}_T = T^{-1} \sum_{t=1}^{T} [\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)]\cdot[\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)]',$$

where $\mathbf{s}_t(\hat{\boldsymbol{\theta}}_T)$ indicates the vector given in [21.1.21] evaluated at $\hat{\boldsymbol{\theta}}_T$. Similarly, the matrix $\mathbf{D}$ can be consistently estimated by

$$\begin{aligned}
\hat{\mathbf{D}}_T = T^{-1} \sum_{t=1}^{T} &\left\{[1/(2\hat{h}_t^2)]\begin{bmatrix} \sum_{j=1}^{m} -2\hat{\alpha}_j \hat{u}_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\hat{\boldsymbol{\beta}}) \end{bmatrix}\right. \\
&\times \left.\begin{bmatrix} \sum_{j=1}^{m} -2\hat{\alpha}_j \hat{u}_{t-j}\mathbf{x}'_{t-j} & [\mathbf{z}_t(\hat{\boldsymbol{\beta}})]' \end{bmatrix} + (1/\hat{h}_t)\begin{bmatrix} \mathbf{x}_t\mathbf{x}'_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right\}.
\end{aligned}$$

Standard errors for $\hat{\theta}_T$ that are robust to misspecification of the family of densities can thus be obtained from the square root of diagonal elements of

$$T^{-1}\hat{\mathbf{D}}_T^{-1}\hat{\mathbf{S}}_T\hat{\mathbf{D}}_T^{-1}.$$

Recall that if the model is correctly specified so that the data were really generated by a Gaussian model, then $\mathbf{S} = \mathbf{D}$, and this simplifies to the usual asymptotic variance matrix for maximum likelihood estimation.

## *Estimation by Generalized Method of Moments*

The *ARCH* regression model of [21.1.17] and [21.1.19] can be characterized by the assumptions that the residual in the regression equation is uncorrelated with the explanatory variables,

$$E[(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t] = \mathbf{0},$$

and that the implicit error in forecasting the squared residual is uncorrelated with lagged squared residuals,

$$E[(u_t^2 - h_t)\mathbf{z}_t] = \mathbf{0}.$$

As noted by Bates and White (1988), Mark (1988), Ferson (1989), Simon (1989), or Rich, Raymond, and Butler (1991), this means that the parameters of an *ARCH* model could be estimated by generalized method of moments,[3] choosing $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\delta}')'$ so as to minimize

$$[\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T)]'\hat{\mathbf{S}}_T^{-1}[\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T)],$$

where

$$\mathbf{g}(\boldsymbol{\theta}; \mathcal{Y}_T) = \begin{bmatrix} T^{-1}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t \\ T^{-1}\sum_{t=1}^{T}\{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 - [\mathbf{z}_t(\boldsymbol{\beta})]'\boldsymbol{\delta}\}\mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}.$$

The matrix $\hat{\mathbf{S}}_T$, standard errors for parameter estimates, and tests of the model can be constructed using the methods described in Chapter 14. Any other variables believed to be uncorrelated with $u_t$ or with $(u_t^2 - h_t)$ could be used as additional instruments.

## *Testing for* ARCH

Fortunately, it is simple to test whether the residuals $u_t$ from a regression model exhibit time-varying heteroskedasticity without actually having to estimate the *ARCH* parameters. Engle (1982, p. 1000) derived the following test based on the Lagrange multiplier principle. First the regression of [21.1.17] is estimated by *OLS* for observations $t = -m + 1, -m + 2, \ldots, T$ and the *OLS* sample residuals $\hat{u}_t$ are saved. Next, $\hat{u}_t^2$ is regressed on a constant and $m$ of its own lagged values:

$$\hat{u}_t^2 = \zeta + \alpha_1\hat{u}_{t-1}^2 + \alpha_2\hat{u}_{t-2}^2 + \cdots + \alpha_m\hat{u}_{t-m}^2 + e_t, \qquad [21.1.26]$$

for $t = 1, 2, \ldots, T$. The sample size $T$ times the uncentered $R_u^2$ from the regression

[3]As noted in Section 14.4, maximum likelihood estimation can itself be viewed as estimation by *GMM* in which the orthogonality condition is that the expected score is zero.

of [21.1.26] then converges in distribution to a $\chi^2$ variable with $m$ degrees of freedom under the null hypothesis that $u_t$ is actually i.i.d. $N(0, \sigma^2)$.

Recalling that the $ARCH(m)$ specification can be regarded as an $AR(m)$ process for $u_t^2$, another approach developed by Bollerslev (1988) is to use the Box-Jenkins methods described in Section 4.8 to analyze the autocorrelations of $u_t^2$. Other tests for $ARCH$ are described in Bollerslev, Chou, and Kroner (1992, p. 8).

## 21.2. *Extensions*

### *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH)

Equations [21.1.9] and [21.1.10] described an $ARCH(m)$ process $(u_t)$ characterized by

$$u_t = \sqrt{h_t} \cdot v_t,$$

where $v_t$ is i.i.d. with zero mean and unit variance and where $h_t$ evolves according to

$$h_t = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2.$$

More generally, we can imagine a process for which the conditional variance depends on an infinite number of lags of $u_{t-j}^2$,

$$h_t = \zeta + \pi(L) u_t^2, \qquad [21.2.1]$$

where

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

A natural idea is to parameterize $\pi(L)$ as the ratio of two finite-order polynomials:

$$\pi(L) = \frac{\alpha(L)}{1 - \delta(L)} = \frac{\alpha_1 L^1 + \alpha_2 L^2 + \cdots + \alpha_m L^m}{1 - \delta_1 L^1 - \delta_2 L^2 - \cdots - \delta_r L^r}, \qquad [21.2.2]$$

where for now we assume that the roots of $1 - \delta(z) = 0$ are outside the unit circle. If [21.2.1] is multiplied by $1 - \delta(L)$, the result is

$$[1 - \delta(L)]h_t = [1 - \delta(1)]\zeta + \alpha(L) u_t^2$$

or

$$\begin{aligned} h_t = \kappa &+ \delta_1 h_{t-1} + \delta_2 h_{t-2} + \cdots + \delta_r h_{t-r} \\ &+ \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \end{aligned} \qquad [21.2.3]$$

for $\kappa \equiv [1 - \delta_1 - \delta_2 - \cdots - \delta_r]\zeta$. Expression [21.2.3] is the *generalized autoregressive conditional heteroskedasticity* model, denoted $u_t \sim GARCH(r, m)$, proposed by Bollerslev (1986).

One's first guess from expressions [21.2.2] and [21.2.3] might be that $\delta(L)$ describes the "autoregressive" terms for the variance while $\alpha(L)$ captures the "moving average" terms. However, this is not the case. The easiest way to see why is to add $u_t^2$ to both sides of [21.2.3] and rewrite the resulting expression as

$$\begin{aligned} h_t + u_t^2 = \kappa &- \delta_1(u_{t-1}^2 - h_{t-1}) - \delta_2(u_{t-2}^2 - h_{t-2}) - \cdots \\ &- \delta_r(u_{t-r}^2 - h_{t-r}) + \delta_1 u_{t-1}^2 + \delta_2 u_{t-2}^2 + \cdots \\ &+ \delta_r u_{t-r}^2 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 + u_t^2 \end{aligned}$$

or

$$u_t^2 = \kappa + (\delta_1 + \alpha_1)u_{t-1}^2 + (\delta_2 + \alpha_2)u_{t-2}^2 + \cdots + (\delta_p + \alpha_p)u_{t-p}^2 + w_t - \delta_1 w_{t-1} - \delta_2 w_{t-2} - \cdots - \delta_r w_{t-r}, \qquad [21.2.4]$$

where $w_t \equiv u_t^2 - h_t$ and $p \equiv \max\{m, r\}$. We have further defined $\delta_j \equiv 0$ for $j > r$ and $\alpha_j \equiv 0$ for $j > m$. Notice that $h_t$ is the forecast of $u_t^2$ based on its own lagged values and thus $w_t \equiv u_t^2 - h_t$ is the error associated with this forecast. Thus, $w_t$ is a white noise process that is fundamental for $u_t^2$. Expression [21.2.4] will then be recognized as an $ARMA(p, r)$ process for $u_t^2$, in which the $j$th autoregressive coefficient is the sum of $\delta_j$ plus $\alpha_j$ while the $j$th moving average coefficient is the negative of $\delta_j$. If $u_t$ is described by a $GARCH(r, m)$ process, then $u_t^2$ follows an $ARMA(p, r)$ process, where $p$ is the larger of $r$ and $m$.

The nonnegativity requirement is satisfied if $\kappa > 0$ and $\alpha_j \geq 0$, $\delta_j \geq 0$ for $j = 1, 2, \ldots, p$. From our analysis of $ARMA$ processes, it then follows that $u_t^2$ is covariance-stationary provided that $w_t$ has finite variance and that the roots of

$$1 - (\delta_1 + \alpha_1)z - (\delta_2 + \alpha_2)z^2 - \cdots - (\delta_p + \alpha_p)z^p = 0$$

are outside the unit circle. Given the nonnegativity restriction, this means that $u_t^2$ is covariance-stationary if

$$(\delta_1 + \alpha_1) + (\delta_2 + \alpha_2) + \cdots + (\delta_p + \alpha_p) < 1.$$

Assuming that this condition holds, the unconditional mean of $u_t^2$ is

$$E(u_t^2) = \sigma^2 = \kappa/[1 - (\delta_1 + \alpha_1) - (\delta_2 + \alpha_2) - \cdots - (\delta_p + \alpha_p)].$$

Nelson and Cao (1992) noted that the conditions $\alpha_j \geq 0$ and $\delta_j \geq 0$ are sufficient but not necessary to ensure nonnegativity of $h_t$. For example, for a $GARCH(1, 2)$ process, the $\pi(L)$ operator implied by [21.2.2] is given by

$$\begin{aligned}\pi(L) &= (1 - \delta_1 L)^{-1}(\alpha_1 L + \alpha_2 L^2) \\ &= (1 + \delta_1 L + \delta_1^2 L^2 + \delta_1^3 L^3 + \cdots)(\alpha_1 L + \alpha_2 L^2) \\ &= \alpha_1 L + (\delta_1\alpha_1 + \alpha_2)L^2 + \delta_1(\delta_1\alpha_1 + \alpha_2)L^3 \\ &\quad + \delta_1^2(\delta_1\alpha_1 + \alpha_2)L^4 + \cdots.\end{aligned}$$

The $\pi_j$ coefficients are all nonnegative provided that $0 \leq \delta_1 < 1$, $\alpha_1 \geq 0$, and $(\delta_1\alpha_1 + \alpha_2) \geq 0$. Hence, $\alpha_2$ could be negative as long as $-\alpha_2$ is less than $\delta_1\alpha_1$.

The forecast of $u_{t+s}^2$ based on $u_t^2, u_{t-1}^2, \ldots$, denoted $\hat{u}_{t+s|t}^2$, can be calculated as in [4.2.45] by iterating on

$$\hat{u}_{t+s|t}^2 - \sigma^2 = \begin{cases} (\delta_1 + \alpha_1)(\hat{u}_{t+s-1|t}^2 - \sigma^2) + (\delta_2 + \alpha_2)(\hat{u}_{t+s-2|t}^2 - \sigma^2) \\ + \cdots + (\delta_p + \alpha_p)(\hat{u}_{t+s-p|t}^2 - \sigma^2) - \delta_s\hat{w}_t - \delta_{s+1}\hat{w}_{t-1} \\ - \cdots - \delta_r\hat{w}_{t+s-r} \qquad \text{for } s = 1, 2, \ldots, r \\ (\delta_1 + \alpha_1)(\hat{u}_{t+s-1|t}^2 - \sigma^2) + (\delta_2 + \alpha_2)(\hat{u}_{t+s-2|t}^2 - \sigma^2) \\ + \cdots + (\delta_p + \alpha_p)(\hat{u}_{t+s-p|t}^2 - \sigma^2) \qquad \text{for } s = r+1, r+2, \ldots, \end{cases}$$

$$\begin{aligned}\hat{u}_{\tau|t}^2 &= u_\tau^2 && \text{for } \tau \leq t \\ \hat{w}_\tau &= u_\tau^2 - \hat{u}_{\tau|\tau-1}^2 && \text{for } \tau = t, t-1, \ldots, t-r+1.\end{aligned}$$

See Baillie and Bollerslev (1992) for further discussion of forecasts and mean squared errors for $GARCH$ processes.

Calculation of the sequence of conditional variances $\{h_t\}_{t=1}^T$ from [21.2.3] requires presample values for $h_{-p+1}, \ldots, h_0$ and $u_{-p+1}^2, \ldots, u_0^2$. If we have observations on $y_t$ and $\mathbf{x}_t$ for $t = 1, 2, \ldots, T$, Bollerslev (1986, p. 316) suggested setting

$$h_j = u_j^2 = \hat{\sigma}^2 \qquad \text{for } j = -p + 1, \ldots, 0,$$

where

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2.$$

The sequence $\{h_t\}_{t=1}^{T}$ can be used to evaluate the log likelihood from the expression given in [21.1.20]. This can then be maximized numerically with respect to $\boldsymbol{\beta}$ and the parameters $\kappa, \delta_1, \ldots, \delta_r, \alpha_1, \ldots, \alpha_m$ of the *GARCH* process; for details, see Bollerslev (1986).

## *Integrated* GARCH

Suppose that $u_t = \sqrt{h_t} \cdot v_t$, where $v_t$ is i.i.d. with zero mean and unit variance and where $h_t$ obeys the *GARCH*$(r, m)$ specification

$$\begin{aligned} h_t = \kappa &+ \delta_1 h_{t-1} + \delta_2 h_{t-2} + \cdots + \delta_r h_{t-r} \\ &+ \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2. \end{aligned}$$

We saw in [21.2.4] that this implies an *ARMA* process for $u_t^2$ where the $j$th autoregressive coefficient is given by $(\delta_j + \alpha_j)$. This *ARMA* process for $u_t^2$ would have a unit root if

$$\sum_{j=1}^{r} \delta_j + \sum_{j=1}^{m} \alpha_j = 1. \qquad [21.2.5]$$

Engle and Bollerslev (1986) referred to a model satisfying [21.2.5] as an *integrated GARCH* process, denoted *IGARCH*.

If $u_t$ follows an *IGARCH* process, then the unconditional variance of $u_t$ is infinite, so neither $u_t$ nor $u_t^2$ satisfies the definition of a covariance-stationary process. However, it is still possible for $u_t$ to come from a strictly stationary process in the sense that the unconditional density of $u_t$ is the same for all $t$; see Nelson (1990).

## *The* ARCH-*in-Mean Specification*

Finance theory suggests that an asset with a higher perceived risk would pay a higher return on average. For example, let $r_t$ denote the ex post rate of return on some asset minus the return on a safe alternative asset. Suppose that $r_t$ is decomposed into a component anticipated by investors at date $t - 1$ (denoted $\mu_t$) and a component that was unanticipated (denoted $u_t$):

$$r_t = \mu_t + u_t.$$

Then the theory suggests that the mean return ($\mu_t$) would be related to the variance of the return ($h_t$). In general, the *ARCH*-in-mean, or *ARCH-M*, regression model introduced by Engle, Lilien, and Robins (1987) is characterized by

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + \delta h_t + u_t \\ u_t &= \sqrt{h_t} \cdot v_t \\ h_t &= \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2 \end{aligned}$$

for $v_t$ i.i.d. with zero mean and unit variance. The effect that higher perceived variability of $u_t$ has on the level of $y_t$ is captured by the parameter $\delta$.

## *Exponential* GARCH

As before, let $u_t = \sqrt{h_t} \cdot v_t$ where $v_t$ is i.i.d. with zero mean and unit variance. Nelson (1991) proposed the following model for the evolution of the conditional variance of $u_t$:

$$\log h_t = \zeta + \sum_{j=1}^{\infty} \pi_j \cdot \{|v_{t-j}| - E|v_{t-j}| + \aleph v_{t-j}\}. \qquad [21.2.6]$$

Nelson's model is sometimes referred to as *exponential GARCH*, or *EGARCH*. If $\pi_j > 0$, Nelson's model implies that a deviation of $|v_{t-j}|$ from its expected value causes the variance of $u_t$ to be larger than otherwise, an effect similar to the idea behind the *GARCH* specification.

The $\aleph$ parameter allows this effect to be asymmetric. If $\aleph = 0$, then a positive surprise ($v_{t-j} > 0$) has the same effect on volatility as a negative surprise of the same magnitude. If $-1 < \aleph < 0$, a positive surprise increases volatility less than a negative surprise. If $\aleph < -1$, a positive surprise actually reduces volatility while a negative surprise increases volatility. A number of researchers have found evidence of asymmetry in stock price behavior—negative surprises seem to increase volatility more than positive surprises.[4] Since a lower stock price reduces the value of equity relative to corporate debt, a sharp decline in stock prices increases corporate leverage and could thus increase the risk of holding stocks. For this reason, the apparent finding that $\aleph < 0$ is sometimes described as the *leverage effect.*

One of the key advantages of Nelson's specification is that since [21.2.6] describes the log of $h_t$, the variance itself ($h_t$) will be positive regardless of whether the $\pi_j$ coefficients are positive. Thus, in contrast to the *GARCH* model, no restrictions need to be imposed on [21.2.6] for estimation. This makes numerical optimization simpler and allows a more flexible class of possible dynamic models for the variance. Nelson (1991, p. 351) showed that [21.2.6] implies that $\log h_t$, $h_t$, and $u_t$ are all strictly stationary provided that $\Sigma_{j=1}^{\infty} \pi_j^2 < \infty$.

A natural parameterization is to model $\pi(L)$ as the ratio of two finite-order polynomials as in the *GARCH(r, m)* specification:

$$\begin{aligned} \log h_t = \kappa &+ \delta_1 \log h_{t-1} + \delta_2 \log h_{t-2} + \cdots \\ &+ \delta_r \log h_{t-r} + \alpha_1\{|v_{t-1}| - E|v_{t-1}| + \aleph v_{t-1}\} \\ &+ \alpha_2\{|v_{t-2}| - E|v_{t-2}| + \aleph v_{t-2}\} + \cdots \\ &+ \alpha_m\{|v_{t-m}| - E|v_{t-m}| + \aleph v_{t-m}\}. \end{aligned} \qquad [21.2.7]$$

The *EGARCH* model can be estimated by maximum likelihood by specifying a density for $v_t$. Nelson proposed using the *generalized error distribution*, normalized to have zero mean and unit variance:

$$f(v_t) = \frac{\nu \exp[-(1/2)|v_t/\lambda|^{\nu}]}{\lambda \cdot 2^{[(\nu+1)/\nu]}\Gamma(1/\nu)}. \qquad [21.2.8]$$

Here $\Gamma(\cdot)$ is the gamma function, $\lambda$ is a constant given by

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

[4]See Pagan and Schwert (1990), Engle and Ng (1991), and the studies cited in Bollerslev, Chou, and Kroner (1992, p. 24).

and $\nu$ is a positive parameter governing the thickness of the tails. For $\nu = 2$, the constant $\lambda = 1$ and expression [21.2.8] is just the standard Normal density. If $\nu < 2$, the density has thicker tails than the Normal, whereas for $\nu > 2$ it has thinner tails. The expected absolute value of a variable drawn from this distribution is

$$E|v_t| = \frac{\lambda \cdot 2^{1/\nu}\Gamma(2/\nu)}{\Gamma(1/\nu)}.$$

For the standard Normal case ($\nu = 2$), this becomes

$$E|v_t| = \sqrt{2/\pi}.$$

As an illustration of how this model might be used, consider Nelson's analysis of stock return data. For $r_t$ the daily return on stocks minus the daily interest rate on Treasury bills, Nelson estimated a regression model of the form

$$r_t = a + br_{t-1} + \delta h_t + u_t.$$

The residual $u_t$ was modeled as $\sqrt{h_t}\cdot v_t$, where $v_t$ is i.i.d. with density [21.2.8] and where $h_t$ evolves according to

$$\begin{aligned}\log h_t - \zeta_t = {} & \delta_1(\log h_{t-1} - \zeta_{t-1}) + \delta_2(\log h_{t-2} - \zeta_{t-2}) \\ & + \alpha_1\{|v_{t-1}| - E|v_{t-1}| + \aleph v_{t-1}\} \\ & + \alpha_2\{|v_{t-2}| - E|v_{t-2}| + \aleph v_{t-2}\}.\end{aligned} \qquad [21.2.9]$$

Nelson allowed $\zeta_t$, the unconditional mean of $\log h_t$, to be a function of time:

$$\zeta_t = \zeta + \log(1 + \rho N_t),$$

where $N_t$ denotes the number of nontrading days between dates $t - 1$ and $t$ and $\zeta$ and $\rho$ are parameters to be estimated by maximum likelihood. The sample log likelihood is then

$$\begin{aligned}\mathcal{L} = {} & T\{\log(\nu/\lambda) - (1 + \nu^{-1})\log(2) - \log[\Gamma(1/\nu)]\} \\ & - (1/2)\sum_{t=1}^{T} |(r_t - a - br_{t-1} - \delta h_t)/(\lambda\cdot\sqrt{h_t})|^{\nu} - (1/2)\sum_{t=1}^{T}\log(h_t).\end{aligned}$$

The sequence $\{h_t\}_{t=1}^{T}$ is obtained by iterating on [21.2.7] with

$$v_t = (r_t - a - br_{t-1} - \delta h_t)/\sqrt{h_t}$$

and with presample values of $\log h_t$ set to their unconditional expectations $\zeta_t$.

### *Other Nonlinear* ARCH *Specifications*

Asymmetric consequences of positive and negative innovations can also be captured with a simple modification of the linear *GARCH* framework. Glosten, Jagannathan, and Runkle (1989) proposed modeling $u_t = \sqrt{h_t}\cdot v_t$, where $v_t$ is i.i.d. with zero mean and unit variance and

$$h_t = \kappa + \delta_1 h_{t-1} + \alpha_1 u_{t-1}^2 + \aleph u_{t-1}^2 \cdot I_{t-1}. \qquad [21.2.10]$$

Here, $I_{t-1} = 1$ if $u_{t-1} \geq 0$ and $I_{t-1} = 0$ if $u_{t-1} < 0$. Again, if the leverage effect holds, we expect to find $\aleph < 0$. The nonnegativity condition is satisfied provided that $\delta_1 \geq 0$ and $\alpha_1 + \aleph \geq 0$.

A variety of other nonlinear functional forms relating $h_t$ to $\{u_{t-1}, u_{t-2}, \ldots\}$ have been proposed. Geweke (1986), Pantula (1986), and Milhøj (1987) suggested a specification in which the log of $h_t$ depends linearly on past logs of the squared residuals. Higgins and Bera (1992) proposed a power transformation of the form

$$h_t = [\zeta^\delta + \alpha_1(u_{t-1}^2)^\delta + \alpha_2(u_{t-2}^2)^\delta + \cdots + \alpha_m(u_{t-m}^2)^\delta]^{1/\delta},$$

with $\zeta > 0$, $\delta > 0$, and $\alpha_i \geq 0$ for $i = 1, 2, \ldots, m$. Gourieroux and Monfort (1992) used a Markov chain to model the conditional variance as a general stepwise function of past realizations.

## *Multivariate* GARCH *Models*

The preceding ideas can also be extended to an $(n \times 1)$ vector $\mathbf{y}_t$. Consider a system of $n$ regression equations of the form

$$\underset{(n\times 1)}{\mathbf{y}_t} = \underset{(n\times k)}{\boldsymbol{\Pi}'} \cdot \underset{(k\times 1)}{\mathbf{x}_t} + \underset{(n\times 1)}{\mathbf{u}_t},$$

where $\mathbf{x}_t$ is a vector of explanatory variables and $\mathbf{u}_t$ is a vector of white noise residuals. Let $\mathbf{H}_t$ denote the $(n \times n)$ conditional variance-covariance matrix of the residuals:

$$\mathbf{H}_t = E(\mathbf{u}_t\mathbf{u}_t'|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots).$$

Engle and Kroner (1993) proposed the following vector generalization of a *GARCH*$(r, m)$ specification:

$$\begin{aligned}\mathbf{H}_t = \mathbf{K} &+ \boldsymbol{\Delta}_1\mathbf{H}_{t-1}\boldsymbol{\Delta}_1' + \boldsymbol{\Delta}_2\mathbf{H}_{t-2}\boldsymbol{\Delta}_2' + \cdots + \boldsymbol{\Delta}_r\mathbf{H}_{t-r}\boldsymbol{\Delta}_r' + \mathbf{A}_1\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\mathbf{A}_1' \\ &+ \mathbf{A}_2\mathbf{u}_{t-2}\mathbf{u}_{t-2}'\mathbf{A}_2' + \cdots + \mathbf{A}_m\mathbf{u}_{t-m}\mathbf{u}_{t-m}'\mathbf{A}_m'.\end{aligned}$$

Here $\mathbf{K}$, $\boldsymbol{\Delta}_s$, and $\mathbf{A}_s$ for $s = 1, 2, \ldots$ denote $(n \times n)$ matrices of parameters. An advantage of this parameterization is that $\mathbf{H}_t$ is guaranteed to be positive definite as long as $\mathbf{K}$ is positive definite, which can be ensured numerically by parameterizing $\mathbf{K}$ as $\mathbf{PP}'$, where $\mathbf{P}$ is a lower triangular matrix.

In practice, for reasonably sized $n$ it is necessary to restrict the specification for $\mathbf{H}_t$ further to obtain a numerically tractable formulation. One useful special case restricts $\boldsymbol{\Delta}_s$ and $\mathbf{A}_s$ to be diagonal matrices for $s = 1, 2, \ldots$. In such a model, the conditional covariance between $u_{it}$ and $u_{jt}$ depends only on past values of $u_{i,t-s} \cdot u_{j,t-s}$, and not on the products or squares of other residuals.

Another popular approach introduced by Bollerslev (1990) assumes that the conditional correlations among the elements of $\mathbf{u}_t$ are constant over time. Let $h_{ii}^{(t)}$ denote the row $i$, column $i$ element of $\mathbf{H}_t$. Thus, $h_{ii}^{(t)}$ represents the conditional variance of the $i$th element of $\mathbf{u}_t$:

$$h_{ii}^{(t)} = E(u_{it}^2|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots).$$

This conditional variance might be modeled with a univariate *GARCH*$(1, 1)$ process driven by the lagged innovation in variable $i$:

$$h_{ii}^{(t)} = \kappa_i + \delta_i h_{ii}^{(t-1)} + \alpha_i u_{i,t-1}^2.$$

We might postulate $n$ such *GARCH* specifications ($i = 1, 2, \ldots, n$), one for each element of $\mathbf{u}_t$. The conditional covariance between $u_{it}$ and $u_{jt}$, or the row $i$, column $j$ element of $\mathbf{H}_t$, is then taken to be a constant correlation $\rho_{ij}$ times the conditional standard deviations of $u_{it}$ and $u_{jt}$:

$$h_{ij}^{(t)} = E(u_{it}u_{jt}|\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots) = \rho_{ij} \cdot \sqrt{h_{ii}^{(t)}} \cdot \sqrt{h_{jj}^{(t)}}.$$

Maximum likelihood estimation of this specification turns out to be quite tractable; see Bollerslev (1990) for details.

Other multivariate models include a formulation for vech($\mathbf{H}_t$) proposed by Bollerslev, Engle, and Wooldridge (1988) and the factor *ARCH* specifications of Diebold and Nerlove (1989) and Engle, Ng, and Rothschild (1990).

## *Nonparametric Estimates*

Pagan and Hong (1990) explored a nonparametric kernel estimate of the expected value of $u_t^2$. The estimate is based on an average value of those $u_\tau^2$ whose preceding values of $u_{\tau-1}, u_{\tau-2}, \ldots, u_{\tau-m}$ were "close" to the values that preceded $u_t^2$:

$$h_t = \sum_{\substack{\tau=1\\ \tau\neq t}}^{T} w_\tau(t)\cdot u_\tau^2.$$

The weights $\{w_\tau(t)\}_{\tau=1,\tau\neq t}^{T}$ are a set of $(T - 1)$ numbers that sum to unity. If the values of $u_{\tau-1}, u_{\tau-2}, \ldots, u_{\tau-m}$ that preceded $u_\tau$ were similar to the values $u_{t-1}, u_{t-2}, \ldots, u_{t-m}$ that preceded $u_t$, then $u_\tau^2$ is viewed as giving useful information about $h_t = E(u_t^2|u_{t-1}, u_{t-2}, \ldots, u_{t-m})$. In this case, the weight $w_\tau(t)$ would be large. If the values that preceded $u_\tau$ are quite different from those that preceded $u_t$, then $u_\tau^2$ is viewed as giving little information about $h_t$ and so $w_\tau(t)$ is small. One popular specification for the weight $w_\tau(t)$ is to use a Gaussian kernel:

$$\kappa_\tau(t) = \prod_{j=1}^{m} (2\pi)^{-1/2}\lambda_j^{-1}\exp[-(u_{\tau-j} - u_{t-j})^2/(2\lambda_j^2)].$$

The positive parameter $\lambda_j$ is known as the *bandwidth*. The bandwidth calibrates the distance between $u_{\tau-j}$ and $u_{t-j}$—the smaller is $\lambda_j$, the closer $u_{\tau-j}$ must be to $u_{t-j}$ before giving the value of $u_\tau^2$ much weight in estimating $h_t$. To ensure that the weights $w_\tau(t)$ sum to unity, we take

$$w_\tau(t) = \frac{\kappa_\tau(t)}{\sum\limits_{\substack{\tau=1\\ \tau\neq t}}^{T} \kappa_\tau(t)}.$$

The key difficulty with constructing this estimate is in choosing the bandwidth parameter $\lambda_j$. One approach is known as *cross-validation*. To illustrate this approach, suppose that the same bandwidth is selected for each lag ($\lambda_j = \lambda$ for $j = 1, 2, \ldots, m$). Then the nonparametric estimate of $h_t$ is implicitly a function of the bandwidth parameter imposed, and accordingly could be denoted $h_t(\lambda)$. We might then choose $\lambda$ so as to minimize

$$\sum_{t=1}^{T} [u_t^2 - h_t(\lambda)]^2.$$

## *Semiparametric Estimates*

Other approaches to describing the conditional variance of $u_t$ include general series expansions for the function $h_t = h(u_{t-1}, u_{t-2}, \ldots)$ as in Pagan and Schwert (1990, p. 278) or for the density $f(v_t)$ itself as in Gallant and Tauchen (1989) and Gallant, Hsieh, and Tauchen (1989). Engle and Gonzalez-Rivera (1991) combined a parametric specification for $h_t$ with a nonparametric estimate of the density of $v_t$ in [21.1.9].

## *Comparison of Alternative Models of Stock Market Volatility*

A number of approaches have been suggested for comparing alternative *ARCH* specifications. One appealing measure is to see how well different models of heteroskedasticity forecast the value of $u_t^2$. Pagan and Schwert (1990) fitted a number of different models to monthly U.S. stock returns from 1834 to 1925. They found that the semiparametric and nonparametric methods did a good job in sample, though the parametric models yielded superior out-of-sample forecasts. Nelson's *EGARCH* specification was one of the best in overall performance from this comparison. Pagan and Schwert concluded that some benefits emerge from using parametric and nonparametric methods together.

Another approach is to calculate various specification tests of the fitted model. Tests can be constructed from the Lagrange mutiplier principle as in Engle, Lilien, and Robins (1987) or Higgins and Bera (1992), on moment tests and analysis of outliers as in Nelson (1991), or on the information matrix equality as in Bera and Zuo (1991). Related robust diagnostics were developed by Bollerslev and Wooldridge (1992). Other diagnostics are illustrated in Hsieh (1989). Engle and Ng (1991) suggested some particularly simple tests of the functional form of $h_t$ related to Lagrange multiplier tests, from which they concluded that Nelson's *EGARCH* specification or Glosten, Jagannathan, and Runkle's modification of *GARCH* described in [21.2.10] best describes the asymmetry in the conditional volatility of Japanese stock returns.

Engle and Mustafa (1992) proposed another approach to assessing the usefulness of a given specification of the conditional variance based on the observed prices for security options. These financial instruments give an investor the right to buy or sell the security at some date in the future at a price agreed upon today. The value of such an option increases with the perceived variability of the security. If the term for which the option applies is sufficiently short that stock prices can be approximated by Brownian motion with constant variance, a well-known formula developed by Black and Scholes (1973) relates the price of the option to investors' perception of the variance of the stock price. The observed option prices can then be used to construct the market's implicit perception of $h_t$, which can be compared with the specification implied by a given time series model. The results of such comparisons are quite favorable to simple *GARCH* and *EGARCH* specifications. Studies by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) suggest that *GARCH*(1, 1) or *EGARCH*(1, 1) models can improve on the market's implicit assessment of $h_t$. Related evidence in support of the *GARCH*(1, 1) formulation was provided by Engle, Hong, Kane, and Noh (1991) and West, Edison, and Cho (1993).

# APPENDIX 21.A. *Derivation of Selected Equations for Chapter 21*

This appendix provides the details behind several of the assertions in the text.

■ **Derivation of [21.1.21].** Observe that

$$\frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}} - \frac{1}{2}\left\{\frac{1}{h_t}\frac{\partial (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\partial \boldsymbol{\theta}} - \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{h_t^2}\frac{\partial h_t}{\partial \boldsymbol{\theta}}\right\}. \quad [21.A.1]$$

But

$$\frac{\partial(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{\partial\boldsymbol{\theta}} = \begin{bmatrix} -2\mathbf{x}_t u_t \\ \mathbf{0} \end{bmatrix} \qquad [21.A.2]$$

and

$$\begin{aligned} \frac{\partial h_t}{\partial\boldsymbol{\theta}} &= \frac{\partial\left(\zeta + \sum_{j=1}^{m} \alpha_j u_{t-j}^2\right)}{\partial\boldsymbol{\theta}} \\ &= \partial\zeta/\partial\boldsymbol{\theta} + \sum_{j=1}^{m} (\partial\alpha_j/\partial\boldsymbol{\theta})\cdot u_{t-j}^2 + \sum_{j=1}^{m} \alpha_j\cdot(\partial u_{t-j}^2/\partial\boldsymbol{\theta}) \\ &= \begin{bmatrix} \mathbf{0} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ 0 \\ u_{t-1}^2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} \mathbf{0} \\ 0 \\ 0 \\ \vdots \\ u_{t-m}^2 \end{bmatrix} + \sum_{j=1}^{m} \alpha_j \begin{bmatrix} -2u_{t-j}\mathbf{x}_{t-j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}. \end{aligned} \qquad [21.A.3]$$

Substituting [21.A.2] and [21.A.3] into [21.A.1] produces

$$\frac{\partial \log f(y_t|\mathbf{x}_t, \mathcal{Y}_{t-1}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -\left\{\frac{1}{2h_t} - \frac{u_t^2}{2h_t^2}\right\} \begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix} + \begin{bmatrix} (\mathbf{x}_t u_t)/h_t \\ \mathbf{0} \end{bmatrix},$$

as claimed. ■

■ **Derivation of [21.1.25].** Expression [21.A.1] can be written

$$\mathbf{s}_t(\boldsymbol{\theta}) = \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\} \frac{\partial \log h_t}{\partial\boldsymbol{\theta}} - \frac{1}{2h_t}\frac{\partial u_t^2}{\partial\boldsymbol{\theta}},$$

from which

$$\begin{aligned} \frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = {} & \frac{1}{2}\frac{\partial \log h_t}{\partial\boldsymbol{\theta}}\left\{\frac{1}{h_t}\frac{\partial u_t^2}{\partial\boldsymbol{\theta}'} - \frac{u_t^2}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right\} + \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\}\frac{\partial^2 \log h_t}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} \\ & - \frac{1}{2h_t}\frac{\partial^2 u_t^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} + \frac{\partial u_t^2}{\partial\boldsymbol{\theta}}\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}. \end{aligned} \qquad [21.A.4]$$

From expression [21.A.2],

$$\begin{aligned} \frac{\partial^2 u_t^2}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} &= \begin{bmatrix} -2\mathbf{x}_t \\ \mathbf{0} \end{bmatrix} \frac{\partial u_t}{\partial\boldsymbol{\theta}'} \\ &= \begin{bmatrix} 2\mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \end{aligned}$$

Substituting this and [21.A.2] into [21.A.4] results in

$$\begin{aligned} \frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} = {} & \frac{1}{2}\frac{\partial \log h_t}{\partial\boldsymbol{\theta}}\left\{\frac{1}{h_t}[-2u_t\mathbf{x}_t' \quad \mathbf{0}'] - \frac{u_t^2}{h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}\right\} + \frac{1}{2}\left\{\frac{u_t^2}{h_t} - 1\right\}\frac{\partial^2 \log h_t}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'} \\ & - \frac{1}{2h_t}\begin{bmatrix} 2\mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -2\mathbf{x}_t u_t \\ \mathbf{0} \end{bmatrix}\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\boldsymbol{\theta}'}. \end{aligned} \qquad [21.A.5]$$

Recall that conditional on $\mathbf{x}_t$ and on $\mathcal{Y}_{t-1}$, the magnitudes $h_t$ and $\mathbf{x}_t$ are nonstochastic and

$$E(u_t|\mathbf{x}_t, \mathcal{Y}_{t-1}) = 0$$
$$E(u_t^2|\mathbf{x}_t, \mathcal{Y}_{t-1}) = h_t.$$

Thus, taking expectations of [21.A.5] conditional on $\mathbf{x}_t$ and $\mathcal{Y}_{t-1}$ results in

$$E\left\{\frac{\partial \mathbf{s}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\,\middle|\, \mathbf{x}_t, \mathcal{Y}_{t-1}\right\} = -\frac{1}{2}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}}\frac{\partial \log h_t}{\partial \boldsymbol{\theta}'} - \frac{1}{h_t}\begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$
$$= -\frac{1}{2h_t^2}\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j} \\ \mathbf{z}_t(\boldsymbol{\beta}) \end{bmatrix}\begin{bmatrix} \sum_{j=1}^{m} -2\alpha_j u_{t-j}\mathbf{x}_{t-j}' & [\mathbf{z}_t(\boldsymbol{\beta})]' \end{bmatrix}$$
$$- \frac{1}{h_t}\begin{bmatrix} \mathbf{x}_t\mathbf{x}_t' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where the last equality follows from [21.A.3]. ■

## Chapter 21 References


Baillie, Richard T., and Tim Bollerslev. 1989. "The Message in Daily Exchange Rates: A Conditional Variance Tale." *Journal of Business and Economic Statistics* 7:297–305.

——— and ———. 1992. "Prediction in Dynamic Models with Time-Dependent Conditional Variances." *Journal of Econometrics* 52:91–113.

Bates, Charles, and Halbert White. 1988. "Efficient Instrumental Variables Estimation of Systems of Implicit Heterogeneous Nonlinear Dynamic Equations with Nonspherical Errors," in William A. Barnett, Ernst R. Berndt, and Halbert White, eds., *Dynamic Econometric Modeling*. Cambridge, England: Cambridge University Press.

Bera, Anil K., and X. Zuo. 1991. "Specification Test for a Linear Regression Model with ARCH Process." University of Illinois at Champaign-Urbana. Mimeo.

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman. 1974. "Estimation and Inference in Nonlinear Structural Models." *Annals of Economic and Social Measurement* 3:653–65.

Black, Fischer, and Myron Scholes. 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81:637–54.

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–27.

———. 1987. "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return." *Review of Economics and Statistics* 69:542–47.

———. 1988. "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process." *Journal of Time Series Analysis* 9:121–31.

———. 1990. "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model." *Review of Economics and Statistics* 72:498–505.

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

———, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. "A Capital Asset Pricing Model with Time Varying Covariances." *Journal of Political Economy* 96:116–31.

——— and Jeffrey M. Wooldridge. 1992. "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances." *Econometric Reviews* 11:143–72.

Cai, Jun. Forthcoming. "A Markov Model of Unconditional Variance in ARCH." *Journal of Business and Economic Statistics*.

Day, Theodore E., and Craig M. Lewis. 1992. "Stock Market Volatility and the Information Content of Stock Index Options." *Journal of Econometrics* 52:267–87.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

Diebold, Francis X., and Mark Nerlove. 1989. "The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model." *Journal of Applied Econometrics* 4:1–21.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50:987–1007.

——— and Tim Bollerslev. 1986. "Modelling the Persistence of Conditional Variances." *Econometric Reviews* 5:1–50.

——— and Gloria Gonzalez-Rivera. 1991. "Semiparametric ARCH Models." *Journal of Business and Economic Statistics* 9:345–59.

———, Ted Hong, Alex Kane, and Jaesun Noh. 1991. "Arbitration Valuation of Variance Forecasts Using Simulated Options Markets." *Advances in Futures and Options Research* forthcoming.

——— and Kenneth F. Kroner. 1993. "Multivariate Simultaneous Generalized ARCH." UCSD. Mimeo.

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

——— and Chowdhury Mustafa. 1992. "Implied ARCH Models from Options Prices." *Journal of Econometrics* 52:289–311.

——— and Victor K. Ng. 1991. "Measuring and Testing the Impact of News on Volatility." University of California, San Diego. Mimeo.

———, Victor K. Ng, and Michael Rothschild. 1990. "Asset Pricing with a FACTOR-ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics* 45:213–37.

Ferson, Wayne E. 1989. "Changes in Expected Security Returns, Risk, and the Level of Interest Rates." *Journal of Finance* 44:1191–1218.

Gallant, A. Ronald, David A. Hsieh, and George Tauchen. 1989. "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate 1974–83." Duke University. Mimeo.

——— and George Tauchen. 1989. "Semi Non-Parametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications." *Econometrica* 57:1091–1120.

Geweke, John. 1986. "Modeling the Persistence of Conditional Variances: A Comment." *Econometric Reviews* 5:57–61.

Glosten, Lawrence R., Ravi Jagannathan, and David Runkle. 1989. "Relationship between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." Northwestern University. Mimeo.

Gourieroux, Christian, and Alain Monfort. 1992. "Qualitative Threshold ARCH Models." *Journal of Econometrics* 52:159–99.

Hamilton, James D., and Raul Susmel. Forthcoming. "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics*.

Higgins, M. L., and A. K. Bera. 1992. "A Class of Nonlinear ARCH Models." *International Economic Review* 33:137–58.

Hsieh, David A. 1989. "Modeling Heteroscedasticity in Daily Foreign-Exchange Rates." *Journal of Business and Economic Statistics* 7:307–17.

Jorion, Philippe. 1988. "On Jump Processes in the Foreign Exchange and Stock Markets." *Review of Financial Studies* 1:427–45.

Lamoureux, Christopher G., and William D. Lastrapes. 1993. "Forecasting Stock Return Variance: Toward an Understanding of Stochastic Implied Volatilities." *Review of Financial Studies* 5:293–326.

Mark, Nelson. 1988. "Time Varying Betas and Risk Premia in the Pricing of Forward Foreign Exchange Contracts." *Journal of Financial Economics* 22:335–54.

Milhøj, Anders. 1985. "The Moment Structure of ARCH Processes." *Scandinavian Journal of Statistics* 12:281–92.

———. 1987. "A Multiplicative Parameterization of ARCH Models." Department of Statistics, University of Copenhagen. Mimeo.

Nelson, Daniel B. 1990. "Stationarity and Persistence in the GARCH(1, 1) Model." *Econometric Theory* 6:318–34.

———. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59:347–70.

——— and Charles Q. Cao. 1992. "Inequality Constraints in the Univariate GARCH Model." *Journal of Business and Economic Statistics* 10:229–35.

Pagan, Adrian R., and Y. S. Hong. 1990. "Non-Parametric Estimation and the Risk Premium," in W. Barnett, J. Powell, and G. Tauchen, eds., *Semiparametric and Nonparametric Methods in Econometrics and Statistics*. Cambridge, England: Cambridge University Press.

Pagan, Adrian R., and G. William Schwert. 1990. "Alternative Models for Conditional Stock Volatility." *Journal of Econometrics* 45:267–90.