11 MODELING VOLATILITY DYNAMICS

Francis X. Diebold
Jose A. Lopez

Introduction

Good macroeconomic and financial theorists, like all good theorists, want to get the facts straight before theorizing; hence, the explosive growth in the methodology and application of time-series econometrics in the last twenty-five years. Many factors fueled that growth, ranging from important developments in related fields (see Box and Jenkins, 1970) to dissatisfaction with the "incredible identifying restrictions" associated with traditional macroeconometric models (Sims, 1980) and the associated recognition that many tasks of interest, such as forecasting, simply do not require a structural model (see Granger and Newbold, 1979). A short list of active subfields includes vector autoregressions, index and dynamic factor models, causality, integration and persistence, cointegration, seasonality, unobserved-components models, state-space representations and the Kalman filter, regime-switching models, nonlinear dynamics, and optimal nonlinear filtering. Any such list must also include models of volatility dynamics. Models of autoregressive conditional heteroskedasticity (ARCH), in particular, provide parsimonious approximations to volatility dynamics and have found wide use in macroeconomics and finance.[1] The family of ARCH models is the subject of this chapter.

Economists are typically introduced to heteroskedasticity in cross-sectional contexts, such as when the variance of a cross-sectional regression disturbance depends on one or more of the regressors. A classic example is the estimation of Engel curves by weighted least squares, in light of the fact that the variance of the disturbance in an expenditure equation may depend on income. Heteroskedasticity is equally pervasive in the time-series contexts prevalent in macroeconomics and finance. For example, in Figures 11.1 and 11.2, we plot the log of daily deutsche-mark/dollar and Swiss franc/dollar spot exchange rates, as well as the daily returns and squared returns, 1974–1991. Volatility clustering (that is, contiguous periods of high or low volatility) is apparent. However, models of cross-sectional heteroskedasticity are not useful in such cases because they are not dynamic. ARCH models, on the other hand, were developed to model such time-series volatility fluctuations. Engle (1982) used them to model the variance of inflation, and more recently they have enjoyed widespread use in modeling asset return volatility.

Exhaustive surveys of the ARCH literature already exist, including Engle and Bollerslev (1986), Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993) and Bollerslev, Engle and Nelson (1994), and it is not our intention to produce another. Rather, we shall provide a selective account of certain aspects of conditional volatility modeling that are of particular relevance in macroeconomics and finance. In the following section we sketch the rudiments of a rather general univariate time-series model, allowing for dynamics in both the conditional mean and variance. We introduce the ARCH and generalized ARCH (GARCH) models there. Then we provide motivation for the models, discuss the properties of the models in depth, and discuss issues related to estimation and testing. Finally, we detail various important extensions and applications of the model and conclude with speculations on productive directions for future research.

A Time-Series Model with Conditional Mean and Variance Dynamics

Wold's (1938) celebrated decomposition theorem establishes that any covariance stationary stochastic process $\{x_t\}$ may be written as the sum of a linearly deterministic component and a linearly indeterministic component with a square-summable, one-sided moving average representation.[2] We write $x_t = d_t + y_t$, where d_t is linearly deterministic and y_t is a linearly regular (or indeterministic) covariance stationary stochastic process (LRCSSP) given by

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

$$B(L) = \sum_{i=0}^{\infty} b_i L^i, \quad \sum_{i=0}^{\infty} b_i^2 < \infty, \quad b_0 = 1,$$

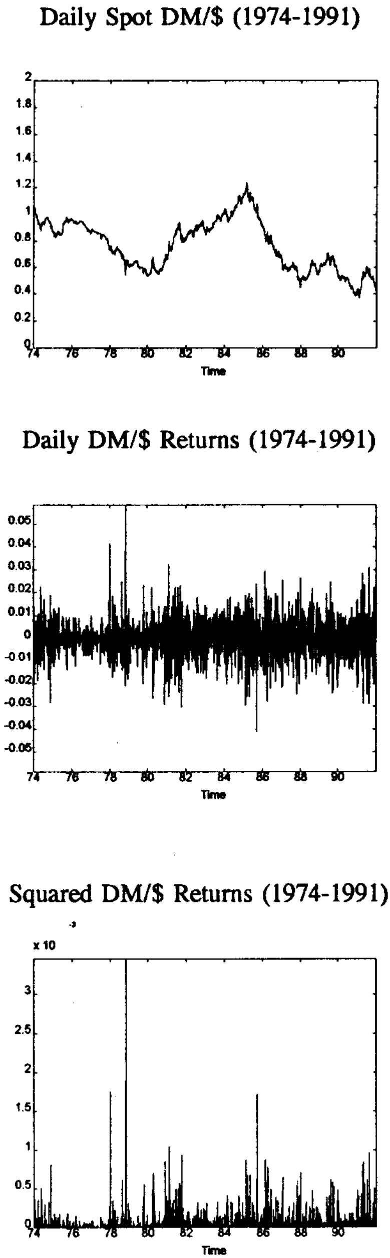
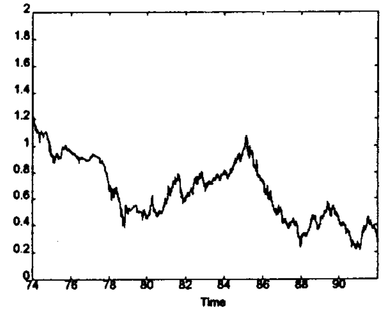
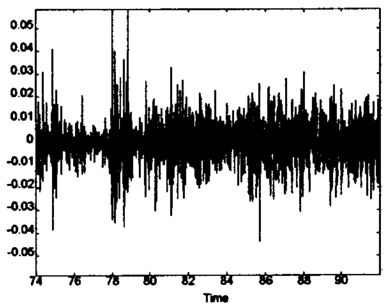


Figure 11.1. Daily logged deutschemark/dollar exchange rates (1974–1991) along with the corresponding daily returns and squared daily returns

Daily Spot SF/$ (1974-1991)



Daily SF/$ Returns (1974-1991)



Squared SF/$ Returns (1974-1991)

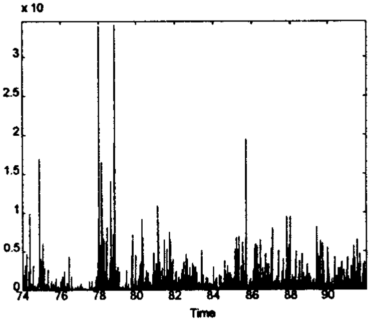


Figure 11.2. Daily logged swiss franc/dollar exchange rate (1974–1991) along with the corresponding daily returns and squared daily returns

$$E[\varepsilon_t \varepsilon_\tau] = \begin{cases} \sigma_\varepsilon^2 < \infty, & \text{if } t = \tau \\ 0, & \text{otherwise.} \end{cases}$$

The uncorrelated innovation sequence $\{\varepsilon_t\}$ need not be Gaussian and therefore need not be independent. Nonindependent innovations are characteristic of nonlinear time series in general and conditionally heteroskedastic time series in particular.

In this section, we introduce the ARCH process within Wold's framework by contrasting the polar extremes of the LRCSSP with independent and identically distributed (iid) innovations, which allows only conditional mean dynamics, and the pure ARCH process, which allows only conditional variance dynamics. We then combine these extremes to produce a generalized model that permits variation in both the first and second conditional moments. Finally, we introduce the generalized ARCH (GARCH) process, which is very useful in practice.

Conditional Mean Dynamics

Suppose that y_t is a LRCSSP with iid, as opposed to merely white-noise, innovations.[3] The ability of the LRCSSP to capture conditional mean dynamics is the source of its power. The *unconditional* mean and variance are $E[y_t] = 0$ and $E[y_t^2] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$, which are both time invariant. However, the *conditional* mean is time varying and is given by $E[y_t \mid \Omega_{t-1}] = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i}$, where the information set is $\Omega_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$.

Because the volatility of many economic time series seems to vary, one would hope that the LRCSSP could capture conditional variance dynamics as well, but such is not the case for the model as presently specified. The conditional variance of y_t is constant at $E[(y_t - E[y_t \mid \Omega_{t-1}])^2 \mid \Omega_{t-1}] = \sigma_\varepsilon^2$. This potentially unfortunate restriction manifests itself in the properties of the k-step-ahead conditional prediction error variance. The least-squares forecast is the conditional expectation

$$E[y_{t+k} \mid \Omega_t] = \sum_{i=0}^{\infty} b_{k+i} \varepsilon_{t-i},$$

and the associated prediction error is

$$y_{t+k} - E[y_{t+k} \mid \Omega_t] = \sum_{i=0}^{k-1} b_i \varepsilon_{t+k-i},$$

which has a conditional prediction error variance of

$$E[(y_{t+k} - E[y_{t+k} \mid \Omega_t])^2 \mid \Omega_t] = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} b_i^2.$$

As $k \to \infty$, the conditional prediction error variance converges to the unconditional variance $\sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$. Note that for any k, the conditional prediction error variance depends only on k and not on Ω_{t-1}; thus, readily available and potentially useful information is discarded.

Conditional Variance Dynamics

By way of contrast, we now introduce a pure ARCH process, which displays *only* conditional variance dynamics. We write

$$y_t = \varepsilon_t,$$

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

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

$$\omega > 0, \ \gamma(L) = \sum_{i=1}^{\infty} \gamma_i L^i, \quad \gamma_i \geq 0 \ \forall \ i, \quad \gamma(1) < 1.$$

The process is parameterized in terms of the conditional density of $\varepsilon_t \mid \Omega_{t-1}$, which is assumed to be normal with a zero conditional mean and a conditional variance that depends linearly on past squared innovations. Note that even though the ε_t's are serially uncorrelated, they are not independent. The stated conditions are sufficient to ensure that the conditional and unconditional variances are positive and finite as well as that y_t is covariance stationary.

The unconditional moments are constant and are given by $E[y_t] = 0$ and $E[(y_t - E[y_t])^2] = \dfrac{\omega}{1 - \gamma(1)}$. As for the conditional moments, by construction, the conditional mean of the process is zero, and the conditional variance is potentially time varying. That is, $E[y_t \mid \Omega_{t-1}] = 0$ and $E[(y_t - E[y_t \mid \Omega_{t-1}])^2 \mid \Omega_{t-1}] = \omega + \gamma(L)\varepsilon_t^2$.

Conditional Mean and Variance Dynamics

We can incorporate *both* conditional mean and conditional variance dynamics by introducing ARCH innovations into the standard LRCSSP. We write

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

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

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

subject to the conditions discussed earlier. Both the unconditional mean and variance are constant—that is, $E[y_t] = 0$ and

$$E[(y_t - E[y_t])^2] = \left(\sum_{i=0}^{\infty} b_i^2\right) E[\varepsilon_t^2] = \frac{\omega}{1-\gamma(1)} \sum_{i=0}^{\infty} b_i^2.$$

However, the conditional mean and variance are time-varying—that is,

$$E[y_t \mid \Omega_{t-1}] = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i},$$

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

Thus, this model treats the conditional mean and variance dynamics in a symmetric fashion by allowing for movement in each, a common characteristic of economic time series.

The Generalized ARCH Process

In the previous subsections, we used an infinite-ordered ARCH process to model conditional variance dynamics. We now introduce the GARCH process, which we shall subsequently focus on almost exclusively. The finite-ordered GARCH model approximates infinite-ordered conditional variance dynamics in the same way that finite-ordered ARMA models approximate infinite-ordered conditional mean dynamics.[4]

The GARCH (p, q) process, introduced by Bollerslev (1986), is given by

$$y_t = \varepsilon_t,$$

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

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

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

$$\omega > 0; \quad \alpha_i \geq 0, \quad \beta_i \geq 0 \ \forall i; \quad \alpha(1) + \beta(1) < 1.$$

The stated conditions ensure that the conditional variance is positive and that y_t is covariance stationary.[5] The ARCH model of Engle (1982) emerges when $\beta(L) = 0$. If *both* $\alpha(L)$ and $\beta(L)$ are zero, then the model is simply iid noise with variance ω. The GARCH (p, q) model can be represented as a restricted infinite-ordered ARCH model:

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

The first two unconditional moments of the pure GARCH model are constant and given by $E[y_t] = 0$ and

$$E[(y_t - E[y_t])^2] = \frac{\omega}{1 - \alpha(1) - \beta(1)}$$

The conditional moments are $E[y_t \mid \Omega_{t-1}] = 0$ and

$$E[(y_t - E[y_t \mid \Omega_{t-1}])^2 \mid \Omega_{t-1}] = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t.$$

Motivating GARCH Processes

GARCH models have been used extensively in macroeconomics and finance because of their attractive approximation-theoretic properties. However, these models do not arise directly from economic theory, and various efforts have been made to imbue them with economic rationale. Here, we discuss both approximation-theoretic and economic motivations for the GARCH framework.

Approximation-Theoretic Considerations

The primary and most powerful justification for the GARCH model is approximation-theoretic. That is, the GARCH model provides a flexible and parsimonious approximation to conditional variance dynamics, in exactly the same way that ARMA models provide a flexible and parsimonious approximation to conditional mean dynamics. In each case, an infinite-ordered distributed lag is approximated as the ratio of two finite, low-ordered lag operator polynomials. The power and usefulness of ARMA and GARCH models come entirely from the fact that ratios of such lag operator polynomials can accurately approximate a variety of infinite-ordered lag operator polynomials.[6] In short, ARMA models with GARCH innovations offer a natural, parsimonious, and flexible way to capture the conditional mean and variance dynamics observed in a time series.

Economic Considerations

Economic considerations may also lead to GARCH effects, although the precise links have proved difficult to establish. Any of the myriad economic forces that produce *persistence* in economic dynamics may be responsible for the appearance of GARCH effects in volatility. In such cases, the persistence happens to be in the conditional second moment rather than the first.

To take one example, conditional heteroskedasticity may arise in situations in which "economic time" and "calendar time" fail to move together. A well-known example from financial economics is the subordinated stochastic process model of Clark (1973). In this model and its subsequent extensions, the number of trades occurring per unit of calendar time (I_t) is a random variable, and the price change per unit of calendar time (ε_t) is the sum of the I_t intraperiod price changes (δ_i), which are assumed to be normally distributed:

$$\varepsilon_t = \sum_{i=1}^{I_t} \delta_i, \quad \delta_i \overset{\text{i.i.d.}}{\sim} N(0, \eta).$$

Using a simple transformation, ε_t can be written more directly as a function of I_t,

$$\varepsilon_t = (\eta I_t)^{1/2} z_t, \quad z_t \overset{\text{i.i.d.}}{\sim} N(0, 1).$$

Thus, ε_t is characterized by conditional heteroskedasticity linked to trading volume. If the number of trades per unit of calendar time displays serial correlation, as in Gallant, Hsieh, and Tauchen (1991), the serial correlation induced in the conditional variance of returns (measured in calendar time) results in GARCH-like behavior. Similar ideas arise in macroeconomics. The divergence between economic time and calendar time accords with the tradition of "phase-averaging" (see Friedman and Schwartz, 1963) and is captured by the time-deformation models of Stock (1987, 1988).

Several other explanations for the existence of GARCH effects have been advanced, including parameter variation (Tsay, 1987), differences in the interpretability of information (Diebold and Nerlove, 1989), market microstructure (Bollerslev and Domowitz, 1991), and agents' "slow" adaptation to news (Brock and LeBaron, 1993). Currently, a consensus economic model producing persistence in

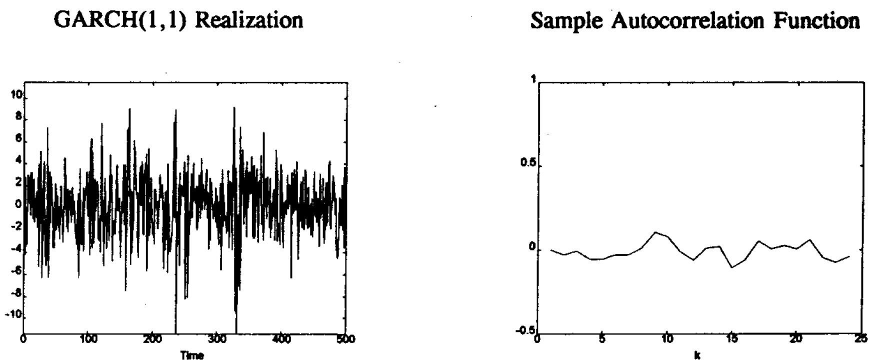


Figure 11.3. GARCH(1, 1) realization and its sample autocorrelation function

conditional volatility does not exist, but it would be foolish to deny the existence of such persistence; measurement is simply ahead of theory.

Properties of GARCH Processes

Here we highlight some important properties of GARCH processes. To facilitate the discussion, we generate a realization of a pure GARCH(1, 1) process of length 500 that we will use repeatedly for illustration.[7] The parameter values are $\omega = 1$, $\alpha = .2$, and $\beta = .7$, and the underlying shocks are $N(0, 1)$.[8] This parameterization delivers a persistent conditional variance and has finite unconditional variance and kurtosis.[9] We plot the realization and its first twenty-five sample autocorrelations in Figure 11.3. The sample autocorrelations are indicative of white noise, as expected.

The Conditional Variance is a Serially Correlated Random Variable

The conditional variance associated with the GARCH model is

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

Recall that the unconditional variance of the process is given by

$$\sigma_y^2 = \frac{\omega}{1 - \alpha(1) - \beta(1)}.$$

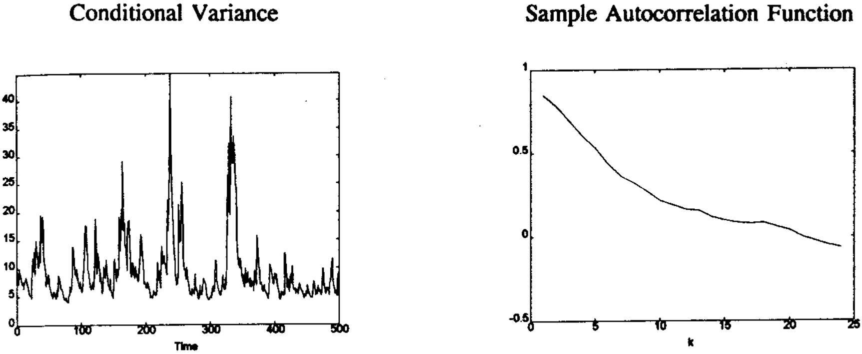


Figure 11.4. Conditional variance of the GARCH(1, 1) realization and its sample autocorrelation function

Replacing ω with $\sigma_y^2(1 - \alpha(1) - \beta(1))$ yields

$$h_t = \sigma_y^2(1 - \alpha(1) - \beta(1)) + \alpha(L)\varepsilon_t^2 + \beta(L)h_t,$$

so that

$$\begin{aligned} h_t - \sigma_y^2 &= \alpha(L)\varepsilon_t^2 - \sigma_y^2\alpha(1) + \beta(L)h_t - \sigma_y^2\beta(1) \\ &= \alpha(L)(\varepsilon_t^2 - \sigma_y^2) + \beta(L)(h_t - \sigma_y^2). \end{aligned}$$

Thus, the conditional variance is itself a serially correlated random variable.

We plot the conditional variance of the simulated GARCH(1, 1) process and its sample autocorrelation function in Figure 11.4. The high persistence of the conditional variance is due to the large sum of the coefficients, $\alpha + \beta = 0.90$.

ε_t^2 Has an ARMA Representation

If ε_t is a GARCH(p, q) process, ε_t^2 has the ARMA representation

$$\varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)v_t + v_t,$$

where $v_t = \varepsilon_t^2 - h_t$ is the difference between the squared innovation and the conditional variance at time t. To see this, note that, by supposition, $h_t = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t$. Adding and subtracting $\beta(L)\varepsilon_t^2$ from the right side gives

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

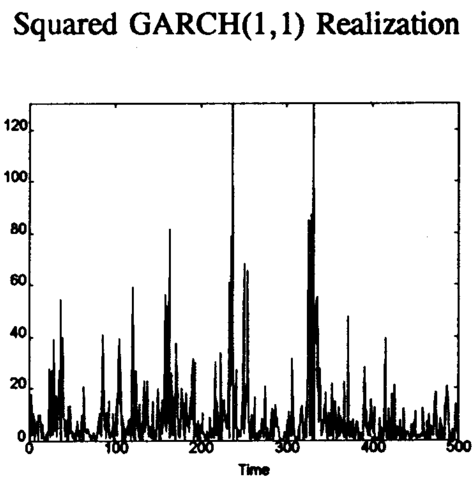


Figure 11.5. Squared GARCH(1, 1) realization

Adding ε_t^2 to each side gives

$$h_t + \varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)[\varepsilon_t^2 - h_t] + \varepsilon_t^2,$$

so that

$$\begin{aligned}\varepsilon_t^2 &= \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)[\varepsilon_t^2 - h_t] + [\varepsilon_t^2 - h_t],\\ &= \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)v_t + v_t.\end{aligned}$$

Thus, ε_t^2 is an ARMA([max(p, q)], p) process with innovation v_t, where $v_t \in [-h_t, \infty)$, and it is covariance stationary if the roots of $\alpha(L) + \beta(L) = 1$ are outside the unit circle.

The square of our GARCH(1, 1) realization is presented in Figure 11.5; the persistence in ε_t^2, which is essentially a proxy for the unobservable h_t, is apparent. *Differences* in the behavior of ε_t^2 and h_t are also apparent, however. In particular, ε_t^2 appears "noisy." To see why, use the multiplicative form of the GARCH model, $\varepsilon_t = h_t^{1/2} z_t$ with $z_t \sim N(0, 1)$. It is easy to see that ε_t^2 is an unbiased estimator of h_t,

$$E[\varepsilon_t^2 \mid \Omega_{t-1}] = E[h_t \mid \Omega_{t-1}]E[z_t^2 \mid \Omega_{t-1}] = E[h_t \mid \Omega_{t-1}],$$

because $z_t^2 \mid \Omega_{t-1} \sim \chi^2_{(1)}$. However, because the median of a $\chi^2_{(1)}$ is .455, $P\left(\varepsilon_t^2 < \frac{1}{2}h_t\right) > 1/2$. Thus, the ε_t^2 proxy introduces a potentially significant error into the analysis of small samples of h_t, $t = 1, \ldots, T$, although the error diminishes as T increases.

The Conditional Prediction Error Variance Depends on the Conditioning Information Set

Because the conditional variance of a GARCH process is a serially correlated random variable, it is of interest to examine the optimal k-step-ahead prediction, prediction error and conditional prediction error variance. Immediately, the k-step-ahead prediction is $E[y_{t+k} \mid \Omega_t] = 0$, and the prediction error is

$$y_{t+k} - E[y_{t+k} \mid \Omega_t] = \varepsilon_{t+k}.$$

This implies that the conditional variance of the prediction error,

$$E[(y_{t+k} - E[y_{t+k} \mid \Omega_t])^2 \mid \Omega_t] = E[\varepsilon_{t+k}^2 \mid \Omega_t],$$

depends on both k *and* Ω_t because of the dynamics in the conditional variance. Simple calculations reveal that the expression for the GARCH(p, q) process is given by

$$E[\varepsilon_{t+k}^2 \mid \Omega_t] = \omega\left[\sum_{i=1}^{k-1}[\alpha(1) + \beta(1)]^i\right] + [\alpha(1) + \beta(1)]^{k-1}h_{t+1}$$

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

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

For finite k, the dependence of the prediction error variance on the current information set Ω_t can be exploited to produce better interval forecasts, as illustrated in Figure 11.6 for $k = 1$. We plot the one-step-ahead 90 percent conditional and unconditional interval forecasts of our simulated GARCH(1, 1) process along with the actual realization. We construct the conditional prediction intervals using the conditional variance

$$E[\varepsilon_{t+1}^2 \mid \Omega_t] = h_{t+1} = \omega + \alpha\varepsilon_t^2 + \beta h_t = 1 + .2\varepsilon_t^2 + .7h_t;$$

thus, the conditional prediction intervals are $\left\{\pm 1.64\sqrt{h_t}\right\}_{t=1}^{500}$. The 90 percent unconditional interval, on the other hand, is simply $[f_{.05}, f_{.95}]$, where f_α denotes the α percentile of the unconditional distribution of the GARCH process. The ability of the conditional prediction intervals to adapt to changes in volatility is clear.

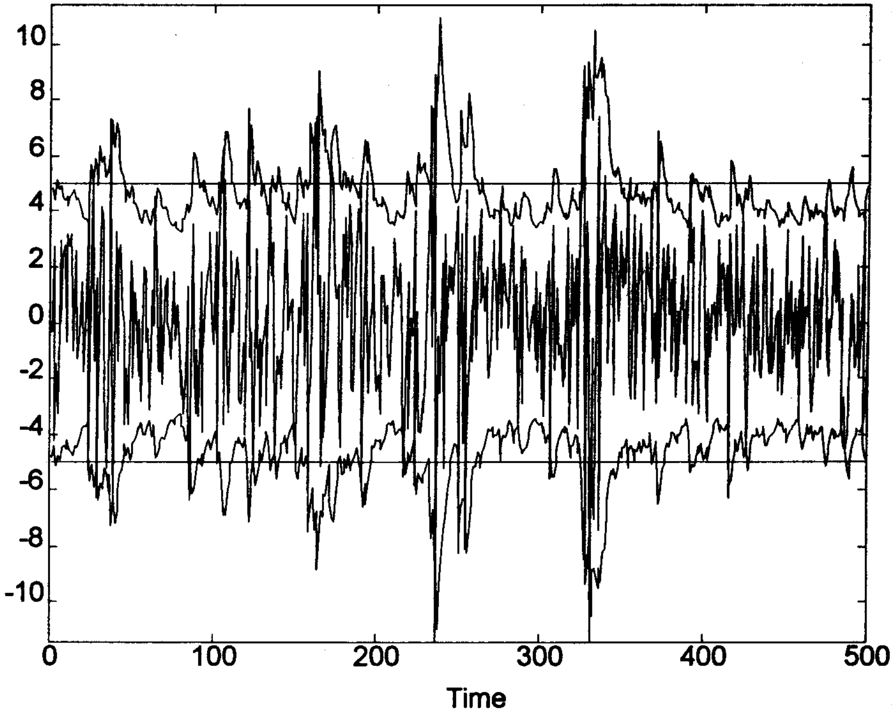


Figure 11.6. GARCH(1, 1) realization with one-step-ahead 90 percent conditional and unconditional confidence intervals

The Implied Unconditional Distribution is Symmetric and Leptokurtic

The moment structure of GARCH processes is a complicated affair. In addition to the earlier-referenced surveys, Milhoj (1985) and Bollerslev (1988) are good sources. However, straightforward calculation reveals that the unconditional distribution of a GARCH process is symmetric and leptokurtic, a characteristic that agrees nicely with a variety of financial market data. The unconditional leptokurtosis of GARCH processes follows from the persistence in conditional variance, which produces the clusters of low-volatility and high-volatility episodes associated with observations in the center and in the tails of the unconditional distribution.

GARCH processes are not constrained to have finite unconditional moments, as shown in Bollerslev (1986). In fact, the only conditionally Gaussian GARCH process with unconditional moments of all orders occurs when $\alpha(L) = \beta(L) = 0$, which is the degenerate case of iid innovations. Otherwise, depending on the precise parameterization, unconditional moments will cease to exist beyond some point. For example, most parameter estimates for financial data indicate an infinite fourth moment, and some even indicate an infinite second moment. Our illustrative process has population mean 0, variance 10, skewness 0, and kurtosis 5.2.

Temporal Aggregation Produces Convergence to Normality

Convergence to normality under temporal aggregation is a key feature of much economic data and is also a property of covariance stationary GARCH processes. The key insight is that a low-frequency change is simply the sum of the corresponding high-frequency changes; for example, an annual change is the sum of the internal quarterly changes, each of which is the sum of its internal monthly changes, and so forth. Thus, if a Gaussian central limit theorem can be invoked for sums of GARCH processes, convergence to normality under temporal aggregation is assured. Such theorems can be invoked so long as the process is covariance stationary, as shown by Diebold (1988) using a central limit argument from White (1984) that requires only the existence of an unconditional second moment. Drost and Nijman (1993) extend Diebold's result by showing that a particular generalization of the GARCH class is closed under temporal aggregation and by characterizing the precise way in which temporal aggregation leads to reduced GARCH effects.[10]

Estimation and Testing of GARCH Models

Following the majority of the literature, we focus primarily on maximum-likelihood estimation (MLE) and associated testing procedures.[11]

Approximate Maximum Likelihood Estimation

As always, the likelihood function is simply the joint density of the observations,

$$L(\theta;\ y_1, \ldots, y_T) = f(y_1, \ldots, y_T;\ \theta).$$

This joint density is non-Gaussian and does not have a known closed-form expression, but it can be factored into the product of conditional densities,

$$\begin{aligned} L(\theta;\ y_1, \ldots, y_T) = f(y_T \mid \Omega_{T-1};\ \theta) f(y_{T-1} \mid \Omega_{T-2};\ \theta) \ldots \\ f(y_{p+1} \mid \Omega_p;\ \theta) f(y_p, \ldots, y_1;\ \theta), \end{aligned}$$

where, if the conditional densities are Gaussian,

$$f(y_t \mid \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi}} h_t(\theta)^{-1/2} \exp\left[-\frac{1}{2}\frac{y_t^2}{h_t(\theta)}\right].$$

The $f(y_p, \ldots, y_1; \theta)$ term is often ignored because a closed-form expression for it does not exist and because its deletion is asymptotically inconsequential. Thus, the approximate log likelihood is

$$ln\ L(\theta; y_{p+1}, \ldots, y_T) = -\frac{T-p}{2} ln(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T} ln\ h_t(\theta) - \frac{1}{2}\sum_{t=p+1}^{T}\frac{y_t^2}{h_t(\theta)}.$$

It may be maximized numerically using iterative procedures and is easily generalized to models richer than the pure univariate GARCH process, such as regression models with GARCH disturbances. In that case, the likelihood is the same with $\varepsilon_t \equiv y_t - E[y_t \mid \Omega_{t-1}; \theta]$ in place of y_t. The unobserved conditional variances $\{h_t(\theta)\}_{t=p+1}^{T}$ that enter the likelihood function are calculated at iteration j using $\theta^{(j-1)}$, the estimated parameter vector at iteration $j-1$. The necessary initial values of the conditional variance are set at the first iteration to the sample variance of the observed data and at all subsequent iterations to the sample variance of a simulated realization with parameters $\theta^{(j-1)}$.

The assumption of conditional normality is not always appropriate. Nevertheless, Weiss (1986) and Bollerslev and Wooldridge (1992) show that even when normality is inappropriately assumed, the resulting quasi-MLE estimates are asymptotically normally distributed and consistent if the conditional mean and variance functions are specified correctly. Bollerslev and Wooldridge (1992), moreover, derive asymptotic standard errors for the quasi-MLE estimates that are robust to conditional nonnormality and are easily calculated as functions of the estimated parameters and the first derivatives of the conditional mean and variance functions.

Exact Maximum Likelihood Estimation

Diebold and Schuermann (1993) propose a numerical procedure for constructing the exact likelihood function of an ARCH process using simulation techniques in conjunction with nonparametric density estimation, thereby retaining the information contained in $\{y_p, \ldots, y_1\}$.[12] Consider the ARCH(p) process, $y_t = \varepsilon_t$, where $\varepsilon_t \mid \Omega_{t-1} \sim N(0, h_t)$, $h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2$, $\omega > 0$, $\alpha_i \geq 0$, $\forall\ i = 1, \ldots, p$, and $\sum_{i=1}^{p}\alpha_i < 1$. The conditional normality assumption is adopted only because it is the most common; alternative distributions can be used with no change in the procedure. Let $\theta = (\omega, \alpha_1, \ldots, \alpha_p)$.

The initial likelihood term $f(y_p, \ldots, y_1; \theta)$ for any given parameter configuration θ is simply the unconditional density of the first p observations evaluated at $\{y_p, \ldots, y_1\}$, which can be estimated to any desired degree of accuracy using well-known techniques of simulation and consistent nonparametric density estimation. At any iteration j, a current "best guess" of the parameter vector $\theta^{(j)}$ exists. Therefore, a very long realization of the process with parameter vector $\theta^{(j)}$ can be simulated and the value of the joint unconditional density evaluated at $\{y_p, \ldots, y_1\}$ can be consistently estimated and denoted as $\hat{f}(y_p, \ldots, y_1; \theta^{(j)})$. This estimated unconditional density can then be substituted into the likelihood where the true unconditional density appears. By simulating a large sample, the difference between $\hat{f}(y_p, \ldots, y_1; \theta^{(j)})$ and $f(y_p, \ldots, y_1; \theta^{(j)})$ is made arbitrarily small, given the consistency of the density estimation technique. The full conditionally Gaussian likelihood, evaluated at $\theta^{(j)}$, is then

$$L(\theta^{(j)}; y_T, \ldots, y_1) \approx \hat{f}(y_p, \ldots, y_1; \theta^{(j)}) \times$$

$$\prod_{t=p+1}^{T} \left[\frac{1}{\sqrt{2\pi}} h_t(\theta^{(j)})^{-1/2} \exp\left[-\frac{1}{2} \frac{y_t^2}{h_t(\theta^{(j)})} \right] \right],$$

which may be maximized with respect to θ using standard numerical techniques.

Testing

Standard likelihood-ratio procedures may be used to test the hypothesis that no ARCH effects are present in a time series, but the numerical estimation required under the ARCH alternative makes that a rather tedious approach. Instead, the Lagrange-multiplier (LM) approach, which requires estimation only under the null, is preferable. Engle (1982) proposes a simple LM test for ARCH under the assumption of conditional normality that involves only a least-squares regression of squared residuals on an intercept and lagged squared residuals. Under the null of no ARCH, TR^2 from that regression is asymptotically distributed as $x^2_{(q)}$, where q is the number of lagged squared residuals included in the regression.

A minor limitation of the LM test for ARCH is the underlying assumption of conditional normality, which is sometimes restrictive.[13] A more important limitation is that the test is difficult to generalize to the GARCH case. Lee (1991) and Lee and King (1993) present such a generalization, but as discussed in Bollerslev, Engle, and Nelson (1994), the GARCH parameters cannot be separately identified in models close to the null—the LM test for GARCH(1, 1) is the same as that for ARCH(1).

Thus, less formal diagnostics are often used, such as the sample autocorrelation function of squared residuals. McLeod and Li (1983) show that under the null hypothesis of no nonlinear dependence among the residuals from an ARMA model, the vector of normalized sample autocorrelations of the squared residuals,

$$\sqrt{T}\hat{\rho}_{\varepsilon^2}(\tau) = \sqrt{T}\frac{\sum_{t=\tau+1}^{T}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{t-\tau}^2 - \hat{\sigma}^2)}{\sum_{t=1}^{T}(\hat{\varepsilon}_t^2 - \hat{\sigma}^2)^2},$$

where $\hat{\sigma}^2$ is the estimated residual variance and $\tau = 1, \ldots, m$, is asymptotically distributed as a multivariate normal with a zero mean and a unit covariance matrix. Moreover, the associated Ljung-Box statistic,

$$\hat{Q}_{\varepsilon^2}(m) = T(T+2)\sum_{\tau=1}^{m}\frac{\hat{\rho}_{\varepsilon^2}(\tau)^2}{T-\tau},$$

is asymptotically $\chi^2_{(m)}$ under the null. If the null is rejected, then nonlinear dependence, such as GARCH, may be present.[14]

After fitting a GARCH model, it is often of interest to test the null hypothesis that the standardized residuals are conditionally homoskedastic. Bollerslev and Mikkelsen (1993) argue that one may use the Ljung-Box statistic on the squared standardized residual autocorrelations, but that the significance of the statistic should be tested using a $\chi^2_{(m-k)}$ distribution, where k is the number of estimated GARCH parameters. This adjustment is necessary due to the deflation associated with fitting the conditional variance model.

A related testing issue concerns the effect of GARCH innovations on tests for *other* deviations from classical behavior. Diebold (1987, 1988) examines the impact of GARCH effects on two standard serial correlation diagnostics, the Bartlett standard errors and the Ljung-Box statistic. As is well-known, in the large-sample Gaussian white-noise case,

$$\hat{\rho}(\tau) \overset{\text{i.i.d.}}{\sim} N\left[0, \frac{1}{T}\right], \tau = 1, 2, \ldots$$

and

$$\hat{Q}(m) = T(T+2)\sum_{\tau=1}^{m}\frac{1}{(T-\tau)}\hat{\rho}(\tau)^2 \overset{a}{\sim} \chi^2_{(m)},$$

where $\hat{\rho}(\tau)$ denotes the sample autocorrelation at lag τ. In the GARCH case, however, an adjustment must be made,

$$\hat{\rho}(\tau) \overset{\text{i.i.d.}}{\sim} N\left(0, \frac{1}{T}\left[1 + \frac{\gamma_{y^2}(\tau)}{\sigma^4}\right]\right), \tau = 1, 2, \ldots,$$

where $\gamma_{y^2}(\tau)$ denotes the autocovariance function of y_t^2 at lag τ and σ^4 is the squared unconditional variance of y_t. The adjustment is largest for small τ and decreases monotonically as $\tau \to \infty$ if the process is covariance stationary. Similarly, the robust Ljung-Box statistic is

$$\hat{Q}(m) = T(T+2)\sum_{\tau=1}^{m}\frac{1}{(T-\tau)}\left[\frac{\sigma^4}{\sigma^4 + \gamma_{y^2}(\tau)}\right]\hat{\rho}(\tau)^2 \overset{a}{\sim} \chi^2_{(m)}.$$

The formulas are made operational by replacing the unknown population parameters with the usual consistent estimators.

It is important to note that the standard error adjustment serves to *increase* the standard errors; failure to perform the adjustment results in standard error bands that are "too tight." Similarly, failure to adjust the Ljung-Box statistic causes empirical test size to be larger then nominal size—often *much* larger, due to the cumulation of distortions through summation. Thus, failure to use robust serial correlation diagnostics for GARCH effects may produce a spurious impression of serial correlation.

A more general approach that yields robust sample autocovariances and related statistics is obtained by adopting a generalized method of moments (GMM) perspective, as proposed by West and Cho (1994).[15] Define $X_t = (\varepsilon_t^2, \varepsilon_t\varepsilon_{t-1}, \ldots, \varepsilon_t\varepsilon_{t-m})'$, $\theta = (E[\varepsilon_t^2], E[\varepsilon_t\varepsilon_{t-1}], \ldots, E[\varepsilon_t\varepsilon_{t-m}])'$ and $g_t(\theta) = X_t - \theta$ as $((m+1) \times 1)$ vectors and $\hat{\theta}_{GMM}$ as the value of θ that satisfies the condition

$$\frac{1}{T-m}\sum_{t=m+1}^{T} g_t(\hat{\theta}_{GMM}) = 0.$$

Note that, because there are as many parameters being estimated as there are orthogonality conditions, GMM simply yields the standard point estimates of the autocovariances. Their standard errors and related test statistics are asymptotically robust, because as shown by Hansen (1982) under general conditions allowing for heteroskedasticity and serial correlation of unknown form, $\sqrt{T}(\hat{\theta}_{GMM} - \theta) \, \mathbf{a} N(0, V)$ where

$$V = \left\{E\left[\frac{\partial g_t(\hat{\theta}_{GMM})}{\partial \theta}\right] S^{-1} E\left[\frac{\partial g_t(\hat{\theta}_{GMM})}{\partial \theta}\right]'\right\}^{-1}$$

and S is the spectral density matrix of $g_t(\theta)$ at frequency zero. This expression for V is made operational by replacing all population objects with consistent estimates. The GMM-estimated autocovariances of y_t and their standard errors will be robust to possible conditional heteroskedasticity in ε_t, as will the Ljung-Box statistic computed using the GMM-estimated autocovariances.

Applications and Extensions

There are numerous applications and extensions of the basic GARCH model. In this section, we highlight those that we judge most important in macroeconomic and financial contexts. It is natural to discuss applications and extensions simultaneously because many of the extensions are motivated by applications.

Functional Form and Density Form

Numerous alternative functional forms for the conditional variance have been suggested in the literature.[16] One of the most interesting is Nelson's (1991) exponential GARCH(p, q) or EGARCH(p, q) model,

$$y_t = \varepsilon_t = h_t^{1/2} z_t,$$

$$z_t \overset{\text{i.i.d.}}{\sim} N(0, 1),$$

$$ln(h_t) = \omega + \sum_{i=1}^{p} \alpha_i g(z_{t-i}) + \sum_{i=1}^{q} \beta_i ln(h_{t-i}),$$

$$g(z_t) = \theta z_t + \gamma(|z_t| - E[|z_t|]).$$

The log specification ensures that the conditional variance is positive, and the model allows for an asymmetric response to the z_t innovations depending on their sign. Thus, the effect of a negative innovation on volatility may differ from that of a positive innovation. This allowance for asymmetric response has proved useful for modeling the "leverage effect" in the stock market described by Black (1976).[17]

With respect to density from, non-Gaussian conditional distributions are easily incorporated into the GARCH model. This is important, because it is commonly found that the Gaussian GARCH model does not explain all of the leptokurtosis in asset returns. With this in mind, Bollerslev (1987) proposes a conditionally student-t GARCH model, in which the degrees-of-freedom is treated as another parameter to be estimated. Alternatively, Engle and González-Rivera (1991) propose a semiparametric methodology in which the conditional variance function is parametrically specified in the usual fashion, but the conditional density is estimated nonparametrically.

GARCH-M: Time-Varying Risk Premia

Consider a regression model with GARCH disturbances of the usual sort, with one additional twist: the conditional variance enters as a regressor, thereby affecting the conditional mean. Write the model as

$$y_t = x_t'\beta + \gamma h_t + \varepsilon_t,$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, h_t).$$

This GARCH-in-Mean (GARCH-M) model is useful in modeling the relationship between risk and return when risk (as measured by the conditional variance) varies. Engle, Lillien, and Robins (1987) introduce the model and use it to examine time-varying risk premia in the term structure of interest rates.

IGARCH: Persistence in Variance

A special case of the GARCH model is the integrated GARCH (IGARCH) model, introduced by Engle and Bollerslev (1986). A GARCH(p, q) process is integrated of order one in variance if $1 - \alpha(L) - \beta(L) = 0$ has a root on the unit circle. The IGARCH process is potentially important because, as an empirical matter, GARCH roots near unity are common in high-frequency financial data.

The earlier ARMA result for the squared GARCH process now becomes an ARIMA result for the squared IGARCH process. As before, $\varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)v_t + v_t$; thus, $[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega - \beta(L)v_t + v_t$. When the autoregressive polynomial contains a unit root, it can be rewritten as

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \phi(L)(1-L)\varepsilon_t^2 = \omega - \beta(L)v_t + v_t.$$

Thus, the differenced squared process is of stationary ARMA form.

Unlike the conditional prediction error variance for the covariance stationary GARCH process, the IGARCH conditional prediction error variance does not converge as the forecast horizon lengthens; instead, it grows linearly with the length of the forecast horizon. Formally, $E[\varepsilon_{t+k}^2 \mid \Omega_t] = (k - 1)\omega + h_{t+1}$ so that $\lim_{k\to\infty} E\left[\varepsilon_{t+k}^2 \mid \Omega_t\right] = \infty$. Thus, the IGARCH process has an infinite unconditional variance.

Clearly, a parallel exists between the IGARCH process and the vast literature on unit roots in conditional mean dynamics (see Stock, 1994). This parallel, however, is partly superficial. In particular, Nelson (1990b) shows that the IGARCH(1, 1) process (with $\omega \neq 0$) is nevertheless strictly stationary and ergodic, which leads one to suspect that likelihood-based inference may proceed in the standard fashion. This conjecture is verified in the theoretical and Monte Carlo work of Lee and Hansen (1994) and Lumsdaine (1992, 1995).

Although conditional variance dynamics are often empirically found to be highly persistent, it is difficult to ascertain whether they are actually integrated. (Again, this difficulty parallels the unit root literature.) Circumstantial evidence against IGARCH arises from several sources, such as temporal aggregation. Little is known about the temporal aggregation of IGARCH processes, but due to the infinite unconditional second moment, we conjecture that a Gaussian central limit theorem is unattainable. (To the best of our knowledge, no existing Gaussian central limit theorems are applicable.) If so, this bodes poorly for the IGARCH model, because actual series displaying GARCH effects seem to approach normality when temporally aggregated. It would then appear likely that highly persistent covariance-stationary GARCH models, not IGARCH models, provide a better approximation to conditional variance dynamics.

The possibility also arises that some findings of IGARCH may be due to misspecification of the conditional variance function. In particular, Diebold (1986) suggests that the appearance of IGARCH could be an artifact resulting from failure to allow for structural breaks in the unconditional variance, if in fact such breaks exist. This is borne out in various contexts by Lastrapes (1989), Lamoureux and Lastrapes (1990), and Hamilton and Susmel (1994). Accordingly, Chu (1993) suggests procedures for testing parameter instability in GARCH models.

Stochastic Volatility Models

A simple first-order stochastic volatility model is given by

$$\varepsilon_t = \sigma_t z_t = \exp\left[\frac{h_t}{2}\right] z_t,$$

$$z_t \sim N(0, 1),$$

$$h_t = \omega + \beta h_{t-1} + \eta_t,$$

$$\eta_t \sim N(0, \sigma_\eta^2).$$

Thus, as opposed to standard GARCH models, h_t is not deterministic conditional on Ω_{t-1}; the conditional variance evolves as a first-order autoregressive process driven by a separate innovation. Moreover, the exponential specification ensures that the conditional variance remains positive. It is clear that the stochastic volatility model is intimately related to Clark's (1973) subordinated stochastic process model: in fact, for all practical purposes, it *is* Clark's model. For further details, see Harvey, Ruiz, and Shephard (1994), and for alternative approaches to estimation, which can be challenging, see Jacquier, Polson, and Rossi (1994) and Kim and Shephard (1994). Although there has been substantial recent interest in stochastic volatility models, their empirical success relative to GARCH models has yet to be established.

Multivariate GARCH Models

Cross-variable interactions are key in macroeconomics and finance. Multivariate GARCH models are used to capture cross-variable conditional volatility interactions. The first multivariate GARCH model, developed by Kraft and Engle (1982), is a multivariate generalization of the pure ARCH model. The multivariate GARCH (p, q) model is proposed in Bollerslev, Engle, and Wooldridge (1988). The N-dimensional Gaussian GARCH(p, q) process is $\varepsilon_t \mid \Omega_{t-1} \sim N(0, H_t)$, where H_t is the $(N \times N)$ conditional covariance matrix given by

$$\text{vech}(H_t) = W + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}') + \sum_{j=1}^{p} B_i \text{vech}(H_t),$$

vech(.) is the vector-half operator that converts $(N \times N)$ matrices into $(N(N+1)/2 \times 1)$ vectors of their lower triangular elements, W is an $(N(N+1)/2 \times 1)$ parameter vector, and A_i and B_j are $((N(N+1)/2) \times (N(N+1)/2))$ parameter matrices. Likelihood-based estimation and inference are conceptually straightforward and parallel the univariate case. The approximate log likelihood function for the conditionally Gaussian multivariate GARCH(p, q) process, aside from a constant, is

$$ln\ L(\theta; \varepsilon_j, \ldots, \varepsilon_T) = -\frac{1}{2}\sum_{t=j}^{T} ln|H_t| - \frac{1}{2}\sum_{t=j}^{T} \varepsilon_t' H_t^{-1} \varepsilon_t\,;\ j \geq 1.$$

In practice, however, two complications arise. First, the conditions needed to ensure that H_t is positive definite are complex and difficult to verify. Second, the model lacks parsimony; an unrestricted parameterization of H_t is too profligate to be of much empirical use. As written above, the model has $(N(N+1)/2)[1+(p+q)N(N+1)/2] = O(N^4)$ parameters, which makes numerical maximization of the likelihood function extremely difficult, even for low values of N, p, and q.

Various strategies have been proposed to deal with the positive definiteness and parsimony complications. Engle and Kroner (1993) propose restrictions that guarantee positive definiteness without entirely ignoring these cross-variable interactions. Bollerslev, Engle, and Wooldridge (1988) enforce further parsimony by requiring that the A_i and B_i matrices be diagonal, reducing the number of parameters to $(N(N + 1)/2)[1 + p + q] = O(N^2)$. However, the parsimony of this "diagonal" model comes at potentially high cost because much of the potential cross-variable volatility interaction, a key point of multivariate analysis, is assumed away.

Common Volatility Patterns: Multivariate Models with Factor Structure

Multivariate models with factor structure, such as the latent-factor GARCH model (Diebold and Nerlove, 1989) and the factor GARCH model (Engle, 1987; Bollerslev and Engle, 1993), capture the idea of commonality of volatility shocks, which appears empirically relevant in systems of asset returns in the stock, foreign exchange, and bond markets.[18] Models with factor structure are also parsimonious and are easily constrained to maintain positive definiteness of the conditional covariance matrix.

In the latent-factor model, movements in each of the N time series are driven by an idiosyncratic shock and a set of $k < N$ common latent shocks or "factors."[19] The latent factors display GARCH effects, whereas the idiosyncratic shocks are i.i.d. and orthogonal at all leads and lags. The one-factor model is important in practice, and we describe it in some detail. The model is written as $\varepsilon_t = \lambda F_t + v_t$, where ε_t, λ, and v_t are $(N \times 1)$ vectors and F_t is a scalar. F_t and v_t have zero conditional means and are orthogonal at all leads and lags. The factor F_t follows a GARCH(p, q) process,

$$F_t \mid \Omega_{t-1} \sim N(0, h_t)$$

$$h_t = \omega + \alpha(L)F_t^2 + \beta(L)h_t,$$

so that the conditional distribution of the observed vector is

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

$$H_t = \lambda\lambda' h_t + \Gamma,$$

where $\Gamma = \mathrm{cov}(v_t) = \mathrm{diag}(\gamma_1, \ldots, \gamma_N)$. Thus, the j^{th} time-t conditional variance is

$$H_{jj,t} = \lambda_j^2 h_t + \gamma_j = \lambda_j^2 \left[\omega + \sum_{i=1}^{p} \alpha_i F_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \right] + \gamma_j,$$

and the j, k^{th} time-t conditional covariance is

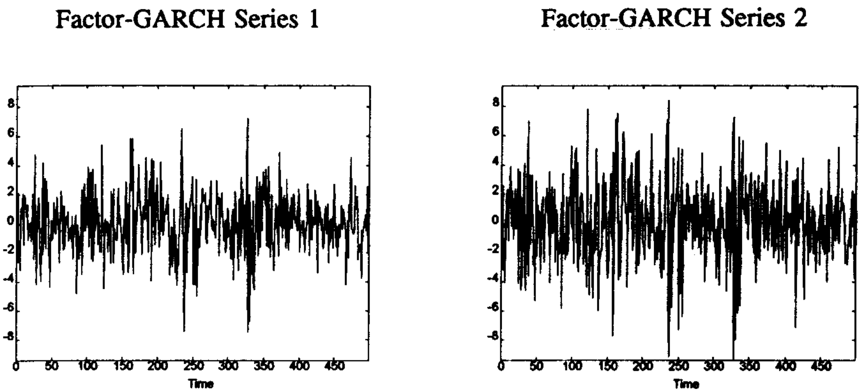


Figure 11.7. The two factor-GARCH series

$$H_{jk,t} = \lambda_j \lambda_k \, h_t = \lambda_j \lambda_k \left[\omega + \sum_{i=1}^{p} \alpha_i F_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \right].$$

Note that the latent factor F_t is unobservable and not directly included in $\Omega_{t-1} = \{\varepsilon_{t-1}, \ldots, \varepsilon_1\}$. Effectively, the latent-factor model is a stochastic volatility model.

In general, the number of parameters in the k-factor model is $N(k + 1) + k^2(1 + p + q) = O(N)$, so the number of parameters in the one-factor case is $2N + (1 + p + q)$, a drastic reduction relative to the general multivariate case. Moreover, the conditional covariance matrix is guaranteed to be positive definite, so long as the conditional variances of the common and idiosyncratic factors are constrained to be positive.

A simulated realization from a bivariate model with one common GARCH(1, 1) factor is shown in Figures 11.7, 11.8, and 11.9. The model is parameterized as

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.9 \end{bmatrix} F_t + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix},$$

$$F_t \mid \Omega_{t-1} \sim N(0, h_t),$$

$$h_t = 1 + .2F_{t-1}^2 + .7h_{t-1},$$

$$(v_{1t}, v_{2t})' \overset{\text{i.i.d.}}{\sim} N(0, I).$$

The realization of the common factor underlying the system is precisely the one presented in our earlier discussion of univariate GARCH models. The latent-factor GARCH series exhibit the volatility clustering present in the common factor. As

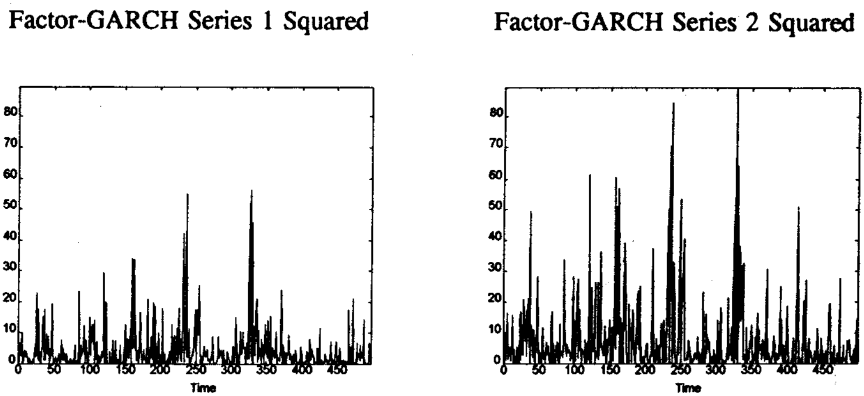


Figure 11.8. The squared factor-GARCH series

before, the squared realizations of the two series indicate a degree of persistence in volatility. Furthermore, as expected, the conditional second moments of the two series are similar to that of F_t because, as shown above, they are simply multiples of h_t.

Diebold and Nerlove (1989) suggest a two-step estimation procedure. The first step entails performing a standard factor analysis—that is, factoring the unconditional covariance matrix as $H = \lambda\lambda'\sigma^2 + \Gamma$, where σ^2 is the unconditional variance of F_t, and extracting an estimate of the time series of factor values $\{\hat{F}_t\}_{t=1}^{T}$. The second step entails estimating the latent-factor GARCH model treating the extracted factor series $\hat{F}_t$ as if it were the actual series F_t.

The Diebold-Nerlove procedure is clearly suboptimal relative to fully simultaneous maximum likelihood estimation because the $\hat{F}_t$ series is not equal to the F_t series, even asymptotically. Harvey, Ruiz, and Sentana (1992) provide a better approximation to the exact likelihood function that involves a correction factor to account for the fact that the F_t series is unobservable.[20] For example, using an ARCH(1) specification, the conditional variance of the latent factor F_t in the Diebold-Nerlove model is

$$h_t = \mathrm{var}(F_t \mid \Omega_{t-1}) = \omega + \alpha F_{t-1}^2 = \omega + \alpha E[F_{t-1}^2 \mid \Omega_{t-1}].$$

Using the identity $F_{t-1} = \hat{F}_{t-1} + (F_{t-1} - \hat{F}_{t-1})$,

$$\begin{aligned} E[F_{t-1}^2 \mid \Omega_{t-1}] &= E[(\hat{F}_{t-1} + (F_{t-1} - \hat{F}_{t-1}))^2 \mid \Omega_{t-1}] = E[\hat{F}_{t-1}^2 \mid \Omega_{t-1}] + p_{t-1} \\ &= \hat{F}_{t-1}^2 + p_{t-1}, \end{aligned}$$

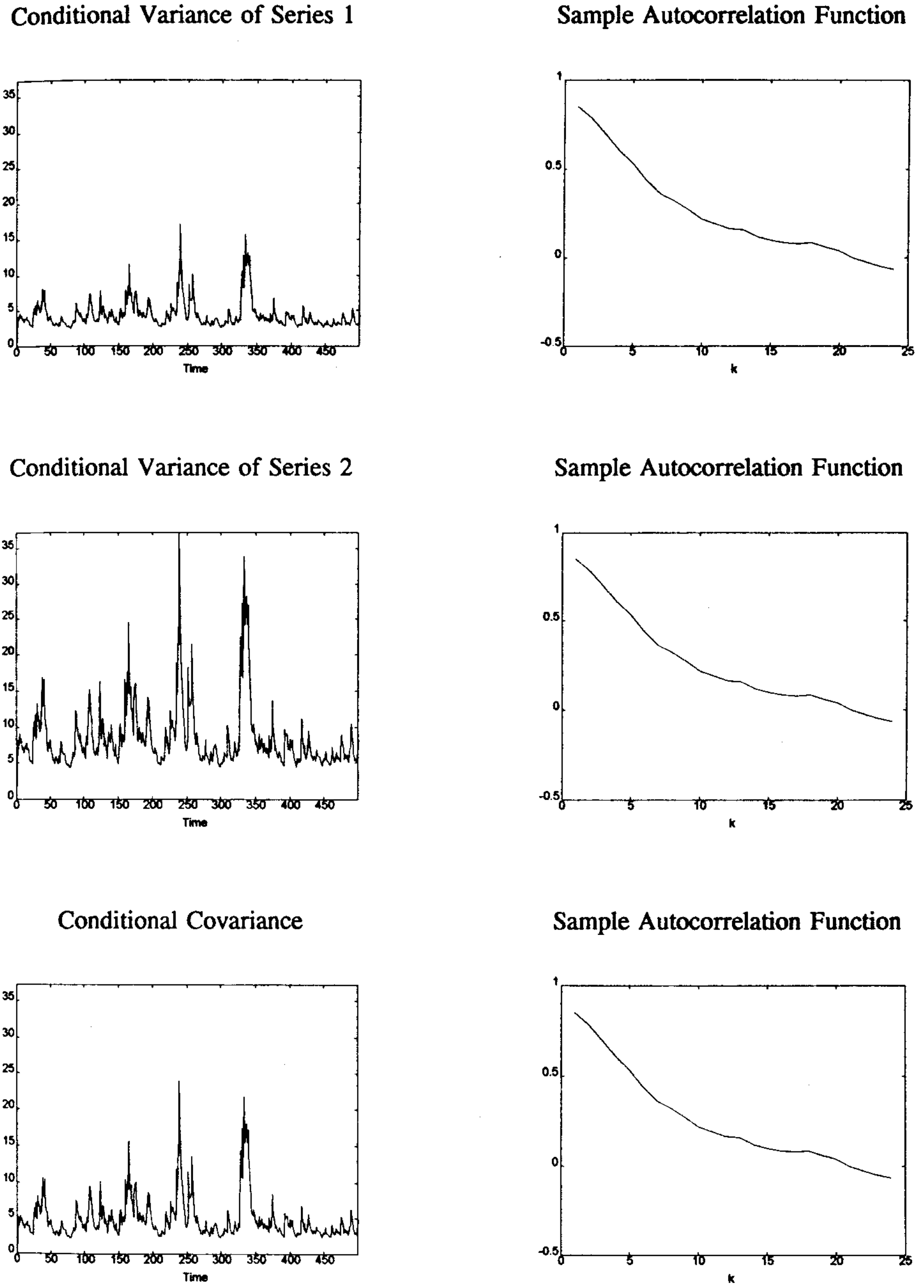


Figure 11.9. The conditional variance of the two factor-GARCH series and their conditional covariance as well on the corresponding sample autocorrelation functions

where p_{t-1} is the correction factor. Thus, h_t is expressed as $h_t = \omega + \alpha(\hat{F}_{t-1} + p_{t-1})$. The correction factor can be constructed using the appropriate elements in the conditional covariance matrix of the state vector estimated by the Kalman filter.

Finally, we note that recently developed Markov-chain Monte Carlo techniques facilitate exact maximum-likelihood estimation of the latent-factor model (or, more precisely, approximate maximum-likelihood estimation with the crucial distinction that the approximation error is under the user's control and can be made as small as possible). For details see Kim and Shephard (1994).

Optimal Prediction Under Asymmetric Loss

Volatility forecasts are readily generated from GARCH models and used for a variety of purposes, such as producing improved interval forecasts, as discussed previously. Less obvious but equally true is the fact that, under asymmetric loss, volatility dynamics can be exploited to produce improved *point* forecasts, as shown by Christoffersen and Diebold (1994). If, for example, y_{t+k} is normally distributed with conditional mean $\mu_{t+k} \mid \Omega_t$ and conditional variance $h_{t+k} \mid \Omega_t$ and $L(e_{t+k})$ is any loss function defined on the k-step-ahead prediction error $e_{t+k} = y_{t+k} - \hat{y}_{t+k}$, then the optimal predictor is $\hat{y}_{t+k} = \mu_{t+k} \mid \Omega_t + \alpha_t$, where α_t depends only on the loss function and the conditional prediction error variance $\mathrm{var}(e_{t+k} \mid \Omega_t) = \mathrm{var}(y_{t+k} \mid \Omega_t) = h_{t+k} \mid \Omega_t$. The optimal predictor under asymmetric loss is *not* the conditional mean, but rather the conditional mean shifted by a time-varying adjustment that depends on the conditional variance. The intuition for this is simple: when, for example, positive prediction errors are more costly than negative errors, a negative conditionally expected error is desirable and is induced by setting the bias $\alpha_t > 0$. The optimal amount of bias depends on the conditional prediction error variance of the process. As the conditional variation around $\mu_{t+k} \mid \Omega_t$ grows, so too does the optimal amount of bias needed to avoid large positive prediction errors.

To illustrate this idea, consider the linlin loss function, so-named for its linearity on each side of the origin (albeit with possibly different slopes):

$$L(y_{t+k} - \hat{y}_{t+k}) = \begin{cases} a \mid y_{t+k} - \hat{y}_{t+k} \mid, \text{ if } y_{t+k} - \hat{y}_{t+k} > 0 \\ b \mid y_{t+k} - \hat{y}_{t+k} \mid, \text{ if } y_{t+k} - \hat{y}_{t+k} \leq 0. \end{cases}$$

Christoffersen and Diebold (1994) show that the optimal predictor of y_{t+k} under this loss function is

$$\hat{y}_{t+k} = \mu_{t+k} \mid \Omega_t + (h_{t+k} \mid \Omega_t)^{1/2} \Phi^{-1}\left[\frac{a}{a+b}\right],$$

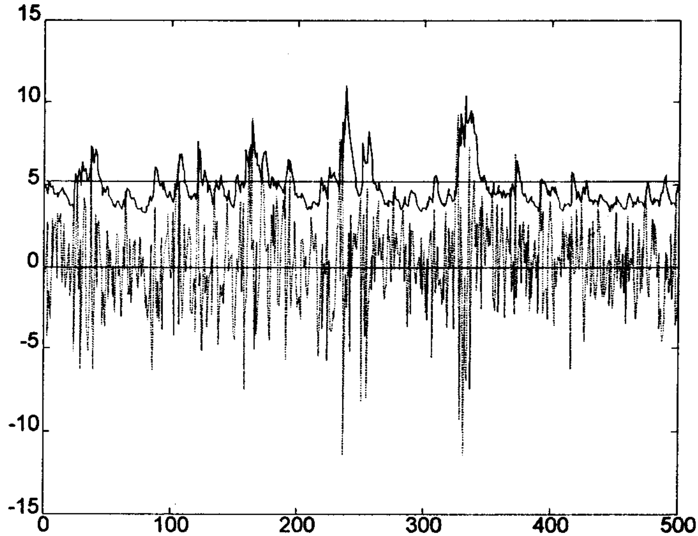


Figure 11.10. GARCH(1, 1) realization with linlin optimal, pseudo-optimal, and conditional mean predictors

Notes: The linlin loss parameters are set to a = .95 and b = .05, so that a/(a + b) = .95. The GARCH(1, 1) parameters are set to $\alpha = .2$ and $\beta = .70$. The dotted line is the GARCH(1, 1) realization. The horizontal line at zero is the conditional mean predictor, the horizontal line at 1.65 is the pseudo-optimal predictor, and the time-varying solid line is the optimal predictor.

where Φ is the Gaussian cumulative density function. In contrast, a pseudo-optimal predictor, which accounts for loss asymmetry but not conditional variance dynamics, is

$$\hat{y}_{t+k} = \mu_{t+k} \mid \Omega_t + \sigma_k \Phi^{-1}\left[\frac{a}{a+b}\right],$$

where σ_k^2 is the unconditional variance of y_{t+k}.

In Figure 11.10, we show our GARCH(1,1) realization together with the one-step-ahead linlin-optimal, pseudo-optimal, and conditional mean predictors for the loss parameters $a = .95$ and $b = .05$. Note that the optimal predictor injects more bias when conditional volatility is high, reflecting the fact that it accounts for both loss asymmetry and conditional heteroskedasticity. This conditionally optimal amount of bias may be more or less than the constant bias associated with the pseudo-optimal predictor. Of course, the conditional-mean predictor injects no bias, as it accounts for neither loss asymmetry nor conditional heteroskedasticity.

Evaluating Volatility Forecasts

Although volatility forecast accuracy comparisons are often conducted using mean-squared error, loss functions that explicitly incorporate the forecast user's economic loss function are more relevant and may lead to different rankings of models. West, Edison, and Cho (1993) and Engle et al. (1993) make important contributions along those lines, proposing economic loss functions based on utility maximization and profit maximization, respectively.

Lopez (1995) proposes a volatility forecast evaluation framework that subsumes a variety of economic loss functions. The framework is based on transforming a model's volatility forecasts into probability forecasts by integrating over the distribution of ε_t. By selecting the range of integration corresponding to an event of interest, a forecast user can incorporate elements of her loss function into the probability forecasts. For example, given $\varepsilon_t \mid \Omega_{t-1} \sim D(0, h_t)$ and a volatility forecast $\hat{h}_t$, an options trader interested in the event $\varepsilon_t \in [L_{\varepsilon,t}, U_{\varepsilon,t}]$ would generate the probability forecast

$$P_t = Pr(L_{\varepsilon,t} < \varepsilon_t < U_{\varepsilon,t}) = Pr\left(\frac{L_{\varepsilon,t}}{\sqrt{\hat{h}_t}} < z_t < \frac{U_{\varepsilon,t}}{\sqrt{\hat{h}_t}}\right) = \int_{l_{\varepsilon,t}}^{u_{\varepsilon,t}} f(z_t)dz_t,$$

where z_t is the standardized innovation, $f(z_t)$ is the functional form of the distribution $D(0, 1)$, and $[l_{\varepsilon,t}, u_{\varepsilon,t}]$ is the standardized range of integration. In contrast, a forecast user such as a portfolio manager or a central bank interested in the behavior of $y_t = \mu_t + \varepsilon_t$, where $\mu_t = E[y_t \mid \Omega_{t-1}]$, would generate the probability forecast

$$P_t = Pr(L_{y,t} < y_t < U_{y,t}) = Pr\left(\frac{L_{y,t} - \hat{\mu}_t}{\sqrt{\hat{h}_t}} < z_t < \frac{U_{y,t} - \hat{\mu}_t}{\sqrt{\hat{h}_t}}\right) = \int_{l_{y,t}}^{u_{y,t}} f(z_t)dz_t,$$

where $\hat{\mu}_t$ is the forecasted conditional mean and $(l_{y,T+t}, u_{y,T+t})$ is the standardized range of integration.

The probability forecasts so-generated can be evaluated using statistical tools tailored to the user's loss function. In particular, probability scoring rules can be used to assess the accuracy of the probability forecasts, and the significance of differences across models can be tested using a generalization of the Diebold-Mariano (1995) procedure. Moreover, the calibration tests of Seillier-Moiseiwitsch and Dawid (1993) can be used to examine the degree of equivalence between an event's predicted and observed frequencies of occurrence within subsets of the probability forecasts specified by the user.

Directions for Future Research

Fifteen years ago, little attention was paid to conditional volatility dynamics in modeling macroeconomic and financial time series; the situation has since changed dramatically. GARCH and related models have proved tremendously useful in modeling such dynamics. However, perhaps in contrast to the impression we may have created, we believe that the literature on modeling conditional volatility dynamics is far from settled and that complacency with the ubiquitous GARCH(1, 1) model is not justified.

Almost without exception, low-ordered (and hence potentially restrictive) GARCH models are used in applied work. For example, among hundreds of empirical applications of the GARCH model, almost all casually and uncritically adopt the GARCH(1, 1) specification. EGARCH applications have followed suit with the vast majority adopting the EGARCH(1, 1) specification. Similarly, applications of the stochastic volatility model typically use an AR(1) specification. However, recent findings suggest that such specifications—as well as the models themselves, regardless of the particular specification—are often too restrictive to maintain fidelity to the data.

It appears, for example, that the conditional volatility dynamics of stock market returns (as well as certain other asset returns) contain long memory. Ding, Engle, and Granger (1993) find positive and significant sample autocorrelations for daily S&P 500 returns at up to 2,500 lags and that their rate of decay is slower than exponential. A model consistent with such long-memory volatility findings is the fractionally integrated GARCH (FIGARCH) model developed by Baillie, Bollerslev, and Mikkelsen (1993), building on earlier work by Robinson (1991). FIGARCH is a model of fractionally integrated conditional variance dynamics, in parallel to the well-known fractionally integrated ARMA models of conditional mean dynamics (see Granger and Joyeux, 1980). The FIGARCH model implies a hyperbolic rate of decay for the autocorrelations of the squared process that is slower than exponential.

To motivate the FIGARCH process, begin with the GARCH(1, 1) process,

$$y_t = \varepsilon_t,$$

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

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

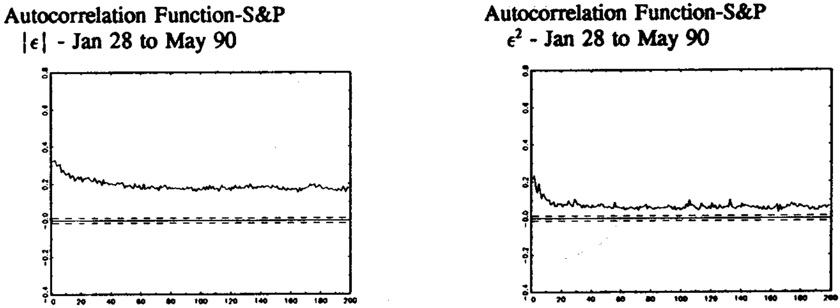


Figure 11.11. Autocorrelation function—S&P ($|\varepsilon|$ Jan. 1928 to May 1990), and autocorrelation function—S&P (ε^2, Jan. 1928 to May 1990)

Rearranging the conditional variance into ARMA form, the FIGARCH (p, d, q) equation is

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \phi(L)(1 - L)^d\varepsilon_t^2 = \omega + (1 - \beta(L))v_t.$$

That is, the $[1 - \alpha(L) - \beta(L)]$ polynomial can be factored into a stationary ARMA component and a long-memory difference operator. If $0 < d < 1$, the process is FIGARCH(p, d, q). If $d = 0$, then the standard GARCH(p, q) model obtains; if $d = 1$, then the IGARCH(p, q) model obtains. Bollerslev and Mikkelsen (1993) conjecture that the coefficients in the ARCH representation of a FIGARCH process ($d < 1$) are dominated by those of an IGARCH process. If so, then FIGARCH ($d < 1$) would be strictly stationary (though not covariance stationary) because IGARCH is strictly stationary.

Long memory is only one of many previously unnoticed features of volatility. Interestingly, as we study volatility more carefully, more and more anomalies emerge. Volatility patterns turn out to differ across assets, time periods, and transformations of the data. The complacency with the "standard" CARCH model is being shattered, and we think it unlikely that any one consensus model will take its place. The implications of this development are twofold. First, real care must be taken in tailoring volatility models to the relevant data, as in Engle and Ng (1993). Second, because all volatility models are likely to be misspecified, care should be taken in assessing models' robustness to misspecification.

To illustrate the deviations from classical GARCH models that turn out to be routinely present in real data, we present in Figure 11.11 the sample autocorrelation functions of the absolute and squared change in the log daily closing value of the S&P 500 stock index, 1928–1990. The autocorrelation functions are shown to displacement $\tau = 200$ in order to assess the evidence for long memory, and dashed lines indicate the Bartlett 95 percent confidence interval for white noise. Note that

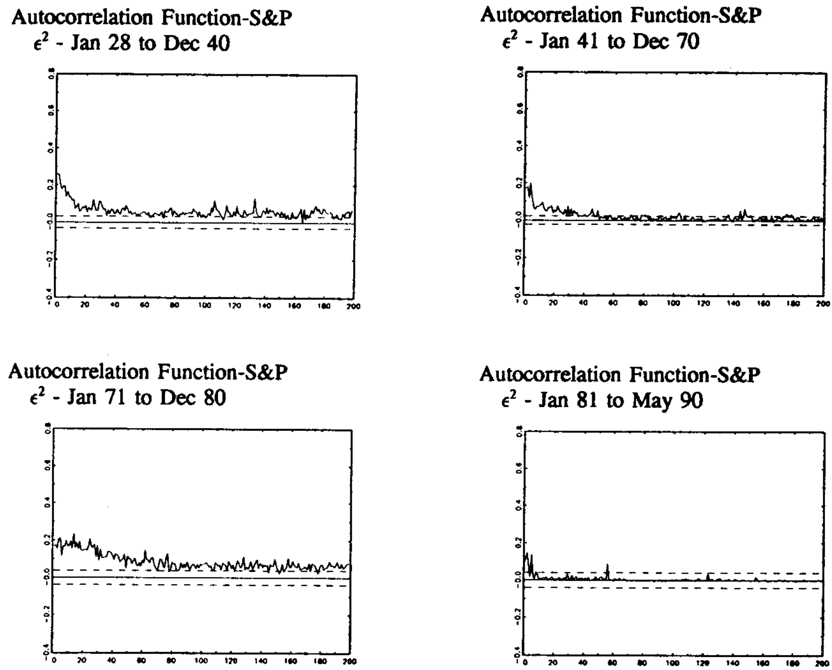


Figure 11.12. Autocorrelation function—S&P (ε^2, Jan. 1928 to Dec. 1940), autocorrelation function—S&P(ε^2, Jan. 1941 to Dec. 1970), autocorrelation function—S&P (ε^2, Jan. 1971 to Dec. 1980), and autocorrelation function—S&P (ε^2, Jan. 1981 to May 1990)

substantially more persistence is found in absolute returns than in squared returns, in keeping with Ding, Engle, and Granger (1993), and that both absolute and squared returns appear too persistent to accord with any of the "standard" volatility models. In addition, these patterns are different over time. In Figure 11.12, we show squared returns over various subperiods: 1928–1940, 1941–1970, 1971–1980, and 1981–1990. It seems clear that most of the long memory is driven by the 1928–1940 period. To the extent that there is any long memory in the post-1940 period, it seems to be coming from the 1970s. Interestingly, there seems to be *no* GARCH effects in the 1980s as shown by the negligible autocorrelations for ε_t^2.

Other assets, including interest rates, foreign exchange rates, and other stock indexes, display a bewildering variety of volatility patterns, as discussed in Mor (1994). Sometimes there seems to be long memory; sometimes not. Sometimes the autocorrelation patterns of ε_t^2 match those of $|\varepsilon_t|$, and sometimes the autocorrelation patterns of $|\varepsilon_t|$ appear much more persistent. The patterns differ across assets and often seem to indicate structural change. For example, the long memory seemingly present in exchange rate volatility seems concentrated in the 1970s, while long memory in interest rate volatility is typically concentrated in the 1980s. These observed phenomena, as well as occasional long-horizon spikes in autocorrelations and the appearance of oscillatory autocorrelation behavior, are again inconsistent with standard specifications.

An additional illustration of the inadequacies of GARCH models is provided by West and Cho (1994). Using weekly exchange rates, they show that for horizons longer than one week, out-of-sample GARCH volatility forecasts loose their value, even though volatility seems highly persistent. The good in-sample performance of GARCH models breaks down rapidly out-of-sample.[21] In addition, standard tests of forecast optimality, such as regressions of realized squared returns on an intercept and the GARCH forecast, strongly reject the null of the optimality of the GARCH forecast with respect to available information. West and Cho (1994) suggest time-varying parameters and discrete shifts in the mean level of volatility as possible explanations.

In light of the emerging evidence that GARCH models are likely misspecified and the unlikely occurrence of happening upon a "correct" specification, it is of interest to consider whether GARCH models might still perform adequately in tracking and forecasting volatility—that is, whether their good properties are robust to misspecification. In a series of papers (Nelson, 1990a, 1992, 1993; Nelson and Foster, 1991, 1994), Nelson and Foster find that the usefulness of GARCH models in volatility tracking and short-term volatility forecasting is robust to a variety of types of misspecification; thus, in spite of misspecification, GARCH models can consistently extract conditional variances from high-frequency time series. More specifically, if a process is well approximated by a continuous-time diffusion, then broad classes of GARCH models provide consistent estimates of the instantaneous conditional variance as the sampling frequency increases. This occurs because the sequence of GARCH(1, 1) models used to form estimates of next period's conditional variance average increasing numbers of squared residuals from the increasingly recent past. In this way, a sequence of GARCH(1, 1) models can consistently estimate next period's conditional variance despite potentially severe misspecification.

Acknowledgments

Helpful comments were received from Richard Baillie, Tim Bollerslev, Pedro de Lima, Wayne Ferson, Kevin Hoover, Peter Robinson, and Til Schuermann. We gratefully acknowledge the support of the National Science Foundation, the Sloan Foundation, and the University of Pennsylvania Research Foundation.

Notes

1. ARCH is short for autoregressive conditional heteroskedasticity.

2. A process is linearly deterministic if it can be predicted to any desired degree of accuracy by linear projection on sufficiently many past observations.

3. Recall that the defining characteristic of white noise is a lack of serial correlation, which is a weaker condition than serial independence.

4. The obvious empirically useful approximation to an LRCSSP (which is an infinite-ordered moving average) with infinite-ordered ARCH errors is an ARMA process with GARCH errors. See Weiss (1984), who studies ARMA processes with finite-ordered ARCH errors. (The GARCH process had not yet been invented.)

5. Nelson and Cao (1992) show that, for higher-order GARCH processes, the nonnegativity constraints are sufficient, but not necessary, for the conditional variance to be positive.

6. See, for example, Jorgenson (1966).

7. Setting $y_0 = 0$ and $h_0 = E(y_t^2)$, we generate 1,500 observations, and we discard the first 1,000 to eliminate the effects of the start-up values.

8. The parameter values for α and β are typical of the parameter estimates reported in the empirical literature.

9. For a precise statement of the necessary and sufficient condition for finite kurtosis, see Bollerslev (1986).

10. Their results, however, require a finite fourth unconditional moment, a condition likely to be violated in financial contexts.

11. Alternative approaches may of course be taken. Geweke (1989), for example, discusses Bayesian procedures.

12. Generalization to the GARCH case has not yet been done.

13. However, Bollersley, and Wooldridge (1992) introduce a modified LM test robust to non-normal conditional distributions.

14. As always, rejection of the null does not imply acceptance of the alternative. Tests for conditional heteroskedasticity, for example, often have power against alternatives of serial correlation as well; see Engle, Hendry, and Trumble (1985).

15. Robinson (1991) also treats the issue of robustness by proposing general classes of heteroskedasticity-robust serial correlation tests and serial correlation-robust heteroskedasticirty tests.

16. In fact, Robinson (1987) goes so far as to propose nonparametric estimation of the conditional variance function, thereby eliminating the need for parametric specification of functional form.

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

18. Models of "copersistence" in variance and cointegration in variance are based on similar ideas (see Bollerslev and Engle, 1993).

19. Despite the similarity in their names, the latent-factor GARCH model discussed here is different from the factor GARCH model. In the latent-factor GARCH case, the observed variables are linear combinations of latent GARCH processes, whereas in the factor GARCH case, linear combinations of the observed variables follow univariate GARCH processes. As pointed out by Sentana (1992), the difference between the two models is similar to the difference between standard factor analysis and principal components analysis.

20. See also King, Sentana, and Wadhwani (1994), Demos and Sentana (1991), and Sentana (1992).

21. Note, however, that West and Cho (1994) evaluate volatility forecasts using the mean-squared error criterion, which may not be the most appropriate. For further discussion, see Bollerslev, Engle, and Nelson (1994) and Lopez (1995).

References

Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen. (1993). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity." Manuscript, J.L. Kellogg School of Management, Northwestern University.

Bera, A.K., and M.L. Higgins. (1993). "ARCH Models: Properties, Estimation and Testing." *Journal of Economic Surveys* 7, 305–362.

Black, F. (1976). "Studies of Stock Price Volatility Changes." *Proceedings of the American Statistical Association, Business and Economic Statistic Section* (pp. 177–181).

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31, 307–327.

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

Bollerslev, T. (1988). "On the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process." *Journal of Time Series Analysis* 9, 121–131.

Bollerslev, T., R.Y. Chou, and K.F. Kroner. (1992). "ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence." *Journal of Econometrics* 52, 5–59.

Bollerslev, T., and I. Domowitz. (1991). "Price Volatility, Spread Variability and the Role of Alternative Market Mechanisms." *Review of Futures Markets* 10, 78–102.

Bollerslev, T., and R.F. Engle. (1993). "Common Persistence in Conditional Variances." *Econometrica* 61, 166–187.

Bollerslev, T., R.F. Engle, and D.B. Nelson. (1994). "ARCH Models." In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics* (Vol. 4). Amsterdam: North-Holland.

Bollerslev, T., R.F. Engle, and J.M. Wooldridge. (1988). "A Capital Asset Pricing Model with Time Varying Covariances." *Journal of Political Economy* 95, 116–131.

Bollerslev, T., and H.O. Mikkelsen. (1993). "Modeling and Pricing Long Memory in Stock Market Volatility." Manuscript, J.L. Kellogg School of Management, Northwestern University.

Bollerslev, T., and J.M. Wooldridge. (1992). "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances." *Econometric Reviws* 11, 143–179.

Box, G.E.P., and G.W. Jenkins. (1970). *Time Series Analysis Forecasting and Control.* Oakland: Holden-Day.

Brock, W.A., and B.D. LeBaron. (1993). "Using Structural Modeling in Building Statistical Models of Volatility and Volume of Stock Market Returns." Manuscript, Department of Economics, University of Wisconsin, Madison.

Christoffersen, P.F., and F.X. Diebold. (1994). "Optimal Prediction under Asymmetric Loss." Technical Working Paper 167, National Bureau of Economic Research.

Chu, C.-S.J. (1993). "Detecting Parameter Shifts in Generalized Autoregressive Conditional Heteroskedasticity Models." Manuscript, Department of Economics, University of Southern California.

Clark, P.K. (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica* 41, 135–156.

Demos, A., and E. Sentana. (1991). "An EM-Based Algorithm for Conditionally Heteroskedastic Latent Factor Models." Manuscript, Financial Markets Group, London School of Economics.

Diebold, F.X. (1986). "Modeling the Persistence of Conditional Variances: Comment." *Econometric Reviews* 5, 51–56.

Diebold, F.X. (1987). "Testing for Serial Correlation in the Presence of ARCH." *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1986* (pp. 323–328). Washington, DC: American Statistical Association.

Diebold, F.X. (1988). *Empirical Modeling of Exchange Rate Dynamics.* New York: Springer-Verlag.

Diebold, F.X., and R.S. Mariano. (1995). "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13, 253–264.

Diebold, F.X., and M. Nerlove. (1989). "The Dynamics of Exchange Rate Volatility: A Multivariate Latent-Factor ARCH Model." *Journal of Applied Econometrics* 4, 1–22.

Diebold, F.X., and T. Schuermann. (1993). "Exact Maximum Likelihood Estimation of ARCH Models." Manuscript, Department of Economics, University of Pennsylvania.

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

Drost, F.C., and T.E. Nijman. (1993). "Temporal Aggregation of GARCH Processes." *Econometrica* 61, 909–927.

Engle, R.F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50, 987–1008.

Engle, R.F. (1987). "Multivariate GARCH with Factor Structures—Cointegration in Variance." Manuscript, Department of Economics, University of California, San Diego.

Engle, R.F., and T. Bollerslev. (1986). "Modeling the Persistence of Conditional Variances." *Econometric Reviews* 5, 1–50.

Engle, R.F., and G. González-Rivera. (1991). "Semiparametric ARCH Models." *Journal of Business and Economic Statistics* 9, 345–359.

Engle, R.F., D.F. Hendry, and D. Trumble. (1985). "Small-Sample Properties of ARCH Estimators and Tests." *Canadian Journal of Economics* 18, 66–93.

Engle, R.F., C.-H. Hong, A. Kane, and J. Noh. (1993). "Arbitrage Valuation of Variance Forecasts with Simulated Options." In D. Chance and R. Tripp (eds.), *Advances in Futures and Options Research.* Greenwich, CT: JIA Press.

Engle, R.F., and K.F. Kroner. (1993). "Multivariate Simultaneous Generalized ARCH." *Econometric Theory*, forthcoming.

Engle, R.F., D.M. Lillien, and R.P. Robins. (1987). "Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model." *Econometrica* 55, 391–408.

Engle, R.F., and V.K. Ng. (1993). "Measuring and Testing the Impact of News on Volatility." *Journal of Finance* 48, 1749–1778.

Friedman, M., and A.J. Schwartz. (1963). *A Monetary History of the United States, 1867–1960*. Princeton, NJ: Princeton University Press.

Gallant, A.R., D.A. Hsieh, and G. Tauchen. (1991). "On Fitting a Recalcitrant Series: The Pound-Dollar Exchange Rate, 1974–1983." In W.A. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press.

Geweke, J. (1989). "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica* 57, 1317–1339.

Granger, C.W.J., and R. Joyeux. (1980). "An Introduction to Long-Memory Time Series Models and Fractional Differencing." *Journal of Time Series Analysis* 1, 15–39.

Granger, C.W.J., and P. Newbold. (1979). *Forecasting Economic Time Series*. New York: Academic Press.

Hamilton, J.D., and R. Susmel. (1994). "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics* 64, 307–333.

Hansen, L.P. (1982). "Large Sample Properties of the Method of Moment Estimators." *Econometrica* 50, 1029–1054.

Harvey, A., E. Ruiz, and E. Sentana. (1992). "Unobserved Component Time Series Models with ARCH Disturbances." *Journal of Econometrics* 52, 129–158.

Harvey, A., E. Ruiz, and N. Shephard. (1994). "Multivariate Stochastic Variance Models." *Review of Economic Studies* 61, 247–264.

Jacquier, E., N.G. Polson, and P.E. Rossi. (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economics Statistics* 12, 371–389.

Jorgenson, D.W. (1966). "Rational Distributed Lag Functions." *Econometrica* 34, 135–149.

Kim, S., and N. Shephard. (1994). "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models." Manuscript, Nuffield College, Oxford University.

King, M., E. Sentana, and S. Wadhwani. (1994). "Volatility and Links Between National Stock Markets." *Econometrica* 62, 901–933.

Kraft, D., and R.F. Engle. (1982). "Autoregressive Conditional Heteroskedasticity in Multiple Time Series Models." Discussion Paper 82–23, Department of Economics, University of California, San Diego.

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

Lastrapes, W.D. (1989). "Exchange Rate Volatility and U.S. Monetary Policy: An ARCH Application." *Journal of Money, Credit and Banking* 21, 66–77.

Lee, J.H.H. (1991). "A Lagrange Multiplier Test for GARCH Models." *Economics Letters* 37, 265–271.

Lee, J.H.H., and M.L. King. (1993). "A Locally Most Mean Powerful Based Score Test for ARCH and GARCH Regression Disturbances." *Journal of Business and Economics Statistics* 11, 17–27.

Lee, S.-W., and B.E. Hansen. (1994). "Asymptotic Theory for the GARCH (1, 1) Quasi-Maximum Likelihood Estimator." *Econometric Theory* 10, 29–52.

Lopez, J.A. (1995). "Evaluating the Predictive Accuracy of Volatility Models." Manuscript, Department of Economics, University of Pennsylvania.

Lumsdaine, R.L. (1992). "Asymptotic Properties of the Quasi-Maximum Likelihood

Estimator in GARCH(1, 1) and IGARCH(1, 1) Models." Manuscript, Department of Economics, Princeton University.

Lumsdaine, R.L. (1995). "Finite Sample Properties of the Maximum Likelihood Estimator In GARCH(1, 1) and IGARCH(1, 1) Models: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 13, 1–10.

McLeod, A.I., and W.K. Li. (1983). "Diagnostic Checking of ARMA Time Series Models Using Squared Residual Autocorrelations." *Journal of Time Series Analysis* 4, 269–273.

Milhoj, A. (1985). "The Moment Structure of ARCH Processes." *Scandinavian Journal of Statistics* 12, 281–292.

Mor, N.M. (1994). "Essays on Nonlinearity in Exchange Rates." Doctoral dissertation, Department of Economics, University of Pennsylvania.

Nelson, D.B. (1990a). "ARCH Models as Diffusion Approximations." *Journal of Econometrics* 45, 7–39.

Nelson, D.B. (1990b). "Stationarity and Persistence in the GARCH(1, 1) Model." *Econometric Theory* 6, 318–334.

Nelson, D.B. (1991). "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica* 59, 347–370.

Nelson, D.B. (1992). "Filtering and Forecasting with Misspecified ARCH Models: I." *Journal of Econometrics* 52, 61–90.

Nelson, D.B. (1993). "Asymptotic Filtering and Smoothing Theory for Multivariate ARCH Models." Manuscript, Graduate School of Business, University of Chicago.

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

Nelson, D.B., and D.P. Foster. (1991). "Filtering and Forecasting with Misspecified ARCH Models: II." *Journal of Econometrics*, forthcoming.

Nelson, D.B., and D.P. Foster. (1994). "Asymptotic Filtering Theory for Univariate ARCH Models." *Econometrica* 62, 1–41.

Robinson, P.M. (1987). "Adaptive Estimation of Heteroskedastic Econometric Models." *Revista de Econometria* (Rio de Janeiro) 7, 5–28.

Robinson, P.M. (1991). "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression." *Journal of Econometrics* 47, 67–84.

Sentana, E. (1992). "Identification and Estimation of Multivariate Conditionally Heteroskedastic Latent Factor Models." Manuscript, Financial Markets Group, London School of Economics.

Seillier-Moiseiwitsch, F., and A.P. Dawid. (1993). "On Testing the Validity of Sequential Probability Forecasts." *Journal of the American Statistical Association* 88, 355–359.

Sims, C.A. (1980). "Macroeconomics and Reality." *Econometrica* 48, 1–48.

Stock, J.H. (1987). "Measuring Business Cycle Time." *Journal of Political Economy* 95, 1240–1261.

Stock, J.H. (1988). "Estimating Continuous-Time Processes Subject to Time Deformation: An Application to Postwar U.S. GNP." *Journal of the American Statistical Association* 83, 77–85.

Stock, J.H. (1994). "Unit Roots and Trend Breaks." In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4. Amsterdam: North-Holland.

Tsay, R.S. (1987). "Conditional Heteroskedastic Time Series Models." *Journal of the American Statistical Association* 82, 509–604.

Weiss, A.A. (1984). "ARMA Models with ARCH Errors." *Journal of Time Series Analysis* 5, 129–143.

Weiss, A.A. (1986). "Asymptotic Theory for ARCH Models: Estimation and Testing." *Econometric Theory* 2, 107–131.

West, K.D., and D. Cho (1994). "The Predictive Ability of Several Models of Exchange Rate Volatility." Technical Working Paper 152, National Bureau of Economic Research.

West, K.D., H.J. Edison and D. Cho (1993). "A Utility-Based Comparison of Some Models of Exchange Rate Volatility." *Journal of International Economics* 35, 23–45.

White, H. (1984). *Asymptotic Theory for Econometricians*. New York: Academic Press.

Wold, H.O. (1938). *The Analysis of Stationary Time Series*. Uppsala: Almquist and Wicksell.