On Fractionally Integrated Autoregressive Moving-Average Time Series Models With Conditional Heteroscedasticity

Shiqing LING and W. K. LI

This article considers fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity, which combines the popular generalized autoregressive conditional heteroscedastic (GARCH) and the fractional (ARMA) models. The fractional differencing parameter d can be greater than $1/2$, thus incorporating the important unit root case. Some sufficient conditions for stationarity, ergodicity, and existence of higher-order moments are derived. An algorithm for approximate maximum likelihood (ML) estimation is presented. The asymptotic properties of ML estimators, which include consistency and asymptotic normality, are discussed. The large-sample distributions of the residual autocorrelations and the square-residual autocorrelations are obtained, and two portmanteau test statistics are established for checking model adequacy. In particular, nonstationary FARIMA(p, d, q)–GARCH(r, s) models are also considered. Some simulation results are reported. As an illustration, the proposed model is also applied to the daily returns of the Hong Kong Hang Seng index (1983–1984).

KEY WORDS: Fractional differencing; Maximum likelihood estimation; Portmanteau tests: Stationarity and ergodicity.

1. INTRODUCTION

Long-memory time series and, in particular, fractional autoregressive integrated moving-average (FARIMA) time series have attracted the attention of more and more statisticians and economists since the early works of Granger and Joyeux (1980), Hosking (1981), and McLeod and Hipel (1978). Geweke and Porter-Hudak (1983) considered a least squares procedure based on the periodogram. Fox and Taqqu (1986) proposed a frequency-domain maximum likelihood (ML) estimation procedure for the FARIMA models. Hosking (1984) and Li and McLeod (1986) considered a time domain ML estimation procedure for FARIMA models. Li and McLeod also obtained the asymptotic distribution of the ML estimation and a portmanteau test for checking model adequacy. Yajima (1988) considered estimation of a regression model with long-memory stationary errors. Diebold and Rudebusch (1989) used the FARIMA model to examine persistence in U.S. aggregate output. Sowell (1992) proposed an exact ML estimation procedure (see also Li 1981). Agiakloglou and Newbold (1993) developed a Lagrange multiplier test for fractional differencing. Hassler and Wolters (1994) investigated the power of unit root tests against fractional root alternatives. Chung and Baillie (1994) studied the small-sample properties of a conditional sum-of-squares estimator for FARIMA models. Beran (1995) considered nonstationary fractionally differenced models. All of the aforementioned works assume that the conditional variance of time series is a constant over time, however.

Time series models with a time-varying conditional variance was first proposed by Engle (1982). This class of models has important applications, particularly in finance and economics (see, e.g., Bollerslev 1986, Bollerslev, Engle, and Woodridge 1988, and Weiss 1984.) Bollerslev (1992) later gave a more complete review on the subject. Li and Mak (1994) derived a formal diagnostic tool for nonlinear time series with conditional heteroscedasticity. Ling (1995) found some simple sufficient conditions for the strict stationarity and the existence of higher-order moments for several well-known classes of autoregressive conditional heteroscedastic (ARCH) models. Given the recent interests in time series with fractional unit root and changing conditional variances, it is only natural to consider time series that exhibit both features. Baillie, Chung, and Tiles (1995) used a special case, a FARIMA $(0, d, 1)$–GARCH(1, 1) model, to analyze the monthly post–World War II consumer price index (CPI) inflation series of 10 different countries. They found evidence of long memory with conditional heteroscedasticity. However, apart from the estimation procedure, a complete statistical inference methodology has not been developed for the above model. In this article, a unified approach that combines both the conditional heteroscedastic and fractional ARMA models is proposed. An important feature of the new model is that the fractional differencing parameter can be greater than $1/2$. Thus the present approach will incorporate also models with unit roots. In the nonstationary case, our estimate procedure is a direct extension of that of Beran (1995).

The article is organized as follows. Section 2 gives the model definition and some properties of the model. Section 3 discusses the ML estimation procedure and the asymptotic properties of the estimators. Section 4 derives the asymptotic distributions of residual autocorrelations and squared residual autocorrelations and obtains two portmanteau tests for checking model adequacy. Sections 5 and 6 report the results of several simulation experiments. Section 7 applies the FARIMA(p, d, q)–GARCH(r, s) model to the daily return of the Hong Kong Hang Seng index (1983–1984). The Appendix contains all proofs of theorems.

Shiqing Ling is Ph.D candidate and W. K. Li is Professor, Department of Statistics, The University of Hong Kong, Hong Kong. The research of W. K. Li was supported by Research Grants Council grant 338 017 00003. The authors thank the editor, the associate editor, and two referees for useful comments that helped correct some errors and greatly improve the article.

Journal of the American Statistical Association
September 1997, Vol. 92, No. 439, Theory and Methods

2. MODELS AND PROPERTIES

2.1 The FARIMA(0, *d*, 0)-GARCH(*r*, *s*) Model

For easy of presentation, we first consider a special case. We define the FARIMA$(0,d,0)$–GARCH(r,s) model to be a discrete time series $\{Y_t\}$ that satisfies the following equation:

$$(1-B)^d Y_t = \varepsilon_t, \tag{1}$$

$$\varepsilon_t | F_{t-1} \sim N(0,h_t), \qquad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i}, \tag{2}$$

where $\alpha_0 > 0$, $\alpha_1,\ldots,\alpha_r$, $\beta_1,\ldots,\beta_s \ge 0$, r and s are positive integers, d is a real number, B is the backward-shift operator, $(1-B)^d$ is defined by the binomial series

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} B^k, \tag{3}$$

and F_{t-1} is the σ field generated by the past information $\{\varepsilon_{t-1},\varepsilon_{t-2},\ldots\}$. Of course, the GARCH model defined by (2) can be replaced by other conditional heteroscedastic models; for example, the threshold ARCH model of Li and Li (1996).

The following theorem gives some basic properties of the model (1)–(2).

Theorem 2.1. Suppose that $\alpha_0 > 0$, α_i's and β_i's ≥ 0, $\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$, and $d < 1/2$. Then for model (1)–(2), there exists a F_t-measurable second-order stationary solution $\{\varepsilon_t, Y_t\}$. It is the only second-order stationary solution given the Z_t's, which are defined later. The solution $\{\varepsilon_t, Y_t\}$ has the following causal representations:

$$\varepsilon_t = Z_t \left[\alpha_0 + \sum_{j=1}^{\infty} \delta^T \left(\prod_{i=1}^{j} A_{t-i} \right) \xi_{t-j} \right]^{1/2} \quad \text{a.s.} \tag{4}$$

and

$$Y_t = \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} \varepsilon_{t-k} \quad \text{a.s.}, \tag{5}$$

where $\xi_t = (\alpha_0 Z_t^2, 0, \ldots, 0, \alpha_0, 0, \ldots, 0)^T_{(r+s)\times 1}$ (i.e., the first component is $\alpha_0 Z_t^2$ and the $(r+1)$th component is α_0), $\{Z_t\}$ are independently normally distributed with mean 0 and variance 1, $\delta = (\alpha_1,\ldots,\alpha_r,\beta_1,\ldots,\beta_s)^T$, and

$$A_t = \left(\begin{array}{ccc|ccc} \alpha_1 Z_t^2 & \cdots & \alpha_r Z_t^2 & \beta_1 Z_t^2 & \cdots & \beta_s Z_t^2 \\ \multicolumn{2}{c}{\mathbf{I}_{(r-1)\times(r-1)}} & O_{(r-1)\times 1} & \multicolumn{3}{c}{O_{(r-1)\times s}} \\ \hline \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ \multicolumn{3}{c|}{O_{(s-1)\times r}} & \multicolumn{2}{c}{\mathbf{I}_{(s-1)\times(s-1)}} & O_{(s-1)\times 1} \end{array} \right), \tag{6}$$

where $\mathbf{I}_{r\times r}$ is the $r \times r$ identity matrix. Hence $\{\varepsilon_t\}$ and $\{Y_t\}$ are strictly stationary and ergodic.

Remark. The representation (4) is simple and clear, comparing to that of Bollerslev (1986). For the first-order ARCH model, Pantula (1989) gave a similar representation under the assumption that $\{\varepsilon_t\}$ begins in the infinite past with a finite initial variance. This restriction is no longer required here, because we adopt a different method of proof.

Let $A_t^{\otimes n}$ be the Kronecker product of n matrices A_t. To derive the condition for the finiteness of moments of FARIMA–GARCH models, we first present a lemma. (This lemma is theorem 6.2 in Ling 1995.)

Lemma. Suppose that $\{\varepsilon_t\}$ is a process generated by (1)–(2). If $\rho[E(A_t^{\otimes m})] < 1$, then $\{\varepsilon_t\}$ is strictly stationary, and the $2m$th order moment of ε_t is finite. In particular, as $r = s = 1$ and $m = 2$, the condition is equivalent to $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$; that is, Bollerslev's (1986) condition.

Remark. For the GARCH$(r,0)$ model (i.e., the rth order ARCH model), Milhøj (1985) gave the condition for the existence of the $2m$th order moment. However, as $m > 2$, Milhøj's condition is too complicated to check. The foregoing lemma gives a simple sufficient condition for the GARCH(r,s) model.

Theorem 2.2. Suppose that $\{\varepsilon_t\}$ and $\{Y_t\}$ are generated by model (1)–(2). Then, under the assumption of Theorem 2.1:

a. If $\rho[E(A_t^{\otimes 2})] < 1$, then the fourth-order moments of $\{\varepsilon_t\}$ and $\{Y_t\}$ are finite.
b. If $\rho[E(A_t^{\otimes 4})] < 1$, then the eighth-order moments of $\{\varepsilon_t\}$ and $\{Y_t\}$ are finite.

Here, A_t is defined by (6).

2.2 The General FARIMA(*p*, *d*, *q*)-GARCH(*r*, *s*) Model

Given the extensive literature on FARIMA and GARCH models, the FARIMA$(0,d,0)$–GARCH(r,s) model is expected to find applications in time series modeling. Yet in practice, more general models may be required (see, e.g., Baillie et al. 1995). A natural extension of the FARIMA$(0,d,0)$–GARCH(r,s) model is the FARIMA(p,d,q)–GARCH(r,s) model. Let $\{Y_t\}$ satisfy the equation

$$\phi(B)(1-B)^d (Y_t - \mu) = \theta(B)\varepsilon_t, \tag{7}$$

$$\varepsilon_t | F_{t-1} \sim N(0,h_t), \qquad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i}, \tag{8}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ are polynomials in B with no common factors; p, q, r, and s are positive integers; and d, B, α_0, $\alpha_1, \ldots, \alpha_r$, $\beta_1, \ldots, \beta_s$ are as defined in Section 2.1. The following theorem shows some properties of FARIMA(p, d, q)–GARCH(r, s) models.

Theorem 2.3. Let $\{Y_t\}$ be generated by (7)–(8). Suppose that all roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle and $\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$.

a. If $d < 1/2$, then $\{Y_t\}$ is second-order stationary and has the following representation:

$$Y_t - \mu = \phi^{-1}(B)\theta(B) \sum_{k=0}^{\infty} \frac{(k+d-1)!}{k!(d-1)!} \varepsilon_{t-k}, \quad (9)$$

where ε_t has representation (4). Hence $\{Y_t\}$ is strictly stationary and ergodic.

b. If $d > -1/2$, then $\{Y_t\}$ is invertible; that is, ε_t can be written as

$$\varepsilon_t = \phi(B)\theta^{-1}(B) \sum_{k=0}^{\infty} \frac{(k-d-1)!}{k!(-d-1)!} (Y_{t-k} - \mu). \quad (10)$$

c. If $\rho[E(A_t^{\otimes 2})] < 1$, then the fourth moments of $\{Y_t\}$ are finite, and if $\rho[E(A_t^{\otimes 4})] < 1$, then the eighth moments of $\{Y_t\}$ are finite, where A_t is defined by (6).

Denoting $\rho_k = \text{cov}(Y_t Y_{t-k})/\text{var}(Y_t)$, similar to theorem 2(4) of Hosking (1981), we can also see that $\rho_k \sim c|k|^{2d-1} (k \to \infty)$ as $|d| < 1/2$ and $\{Y_t\}$ is stationary and invertible, where c is a constant number. This means that the dependence between observations decays hyperbolically. The decay is slower than the geometric decay of the stationary ARMA model. These types of long-range dependency may be modeled by the present approach.

3. ESTIMATION OF FARIMA–GARCH MODELS

3.1 The Stationary Case

Suppose that $Y_1, \ldots, Y_n$ are generated by the model (7)–(8) with known mean $\mu = 0$. Denoting $\gamma = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)^T$, $\delta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)^T$, and $\lambda = (\gamma^T, \delta^T)^T$, assume that $\lambda_0 = (\gamma_0^T, \delta_0^T)^T$ is the true value of λ and is in the interior of the compact set Θ. The approximate ML estimator (MLE) $\hat{\lambda}_n$ of λ in Θ maximizes the conditional log-likelihood on F_0 (ignoring constants),

$$L(\lambda) = \frac{1}{n} \sum_{t=1}^{n} l_t, \qquad l_t = -\frac{1}{2} \ln h_t - \frac{\varepsilon_t^2}{2h_t}. \quad (11)$$

To obtain $\hat{\lambda}_n$, we need to find the first-order derivatives and the information matrix. For each t, these are as follows:

$$\frac{\partial l_t}{\partial \gamma} = \frac{1}{2h_t} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial h_t}{\partial \gamma} - \frac{\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \gamma}$$

and

$$\frac{\partial l_t}{\partial \delta} = \frac{1}{2h_t} \frac{\partial h_t}{\partial \delta} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right), \quad (12)$$

where

$$\frac{\partial \varepsilon_t}{\partial \phi_j} = -\phi^{-1}(B)\varepsilon_{t-j}, \qquad \frac{\partial \varepsilon_t}{\partial \theta_j} = -\theta^{-1}(B)\varepsilon_{t-j}, \quad (13)$$

$$\frac{\partial \varepsilon_t}{\partial d} = -\sum_{k=1}^{\infty} \frac{1}{k} \varepsilon_{t-k}, \qquad \frac{\partial h_t}{\partial \delta} = \tilde{\varepsilon}_t + \sum_{i=1}^{s} \beta_i \frac{\partial h_{t-i}}{\partial \delta}, \quad (14)$$

and

$$\frac{\partial h_t}{\partial \gamma} = 2 \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i} \frac{\partial \varepsilon_{t-i}}{\partial \gamma} + \sum_{i=1}^{s} \beta_i \frac{\partial h_{t-i}}{\partial \gamma}, \quad (15)$$

where $\tilde{\varepsilon}_t = (1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-r}^2, h_{t-1}, \ldots, h_{t-s})^T$.

Directly differentiating (12), we have

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \gamma \partial \gamma^T} = & -\frac{1}{h_t} \frac{\partial \varepsilon_t}{\partial \gamma} \frac{\partial \varepsilon_t}{\partial \gamma^T} - \frac{1}{2h_t^2} \frac{\partial h_t}{\partial \gamma} \frac{\partial h_t}{\partial \gamma^T} \frac{\varepsilon_t^2}{h_t} \\ & + \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \gamma} \left[\frac{1}{2h_t} \frac{\partial h_t}{\partial \gamma^T} \right] \\ & - \frac{2\varepsilon_t}{h_t} \frac{\partial \varepsilon_t}{\partial \gamma} \frac{\partial h_t}{\partial \gamma^T} + \frac{\varepsilon_t}{h_t} \frac{\partial^2 \varepsilon_t}{\partial \gamma \partial \gamma^T} \end{aligned} \quad (16)$$

and

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \delta \partial \delta^T} = & -\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \delta} \frac{\partial h_t}{\partial \delta^T} \left(\frac{\varepsilon_t^2}{h_t} \right) \\ & + \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \gamma} \left(\frac{1}{h_t} \frac{\partial h_t}{\partial \delta^T} \right). \end{aligned} \quad (17)$$

Similarly, we find $(\partial^2 l_t)/(\partial \gamma \partial \delta^T)$. Note that only n observations are available. However, ε_t, h_t, $\partial \varepsilon_t/\partial \gamma$, $\partial \varepsilon_t/\partial \delta$, $\partial h_t/\partial \gamma$, and $\partial h_t/\partial \delta$ all depend on the theoretically infinite past history of $\{Y_t\}$ or $\{\varepsilon_t\}$. For simplicity, we assume that the presample values of Y_t and ε_t are 0 and choose the presample estimates of h_t and ε_t^2 to be $\sum_{t=1}^n \varepsilon_t^2/n$. This will not affect asymptotic efficiency and other asymptotic properties (see Bollerslev 1986 and Weiss 1986). The following theorems give the asymptotic properties of the information matrix and the MLE $\hat{\lambda}_n$.

Theorem 3.1. Suppose that $\{Y_t\}$ and $\{\varepsilon_t\}$ generated by (7)–(8) with $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_r$, $\beta_1, \ldots, \beta_s \geq 0$, $\sum_{i=1}^r \alpha_i + \sum_{i=1}^s \beta_i < 1$, $E(\varepsilon_t^4) < \infty$, and $|d| < 1/2$ and that all roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle. Then

$$-\frac{1}{n} \sum_{t=1}^{n} \begin{pmatrix} \partial^2 l_t/(\partial \gamma \partial \gamma^T) & \partial^2 l_t/(\partial \delta \partial \gamma^T) \\ \partial^2 l_t/(\partial \gamma \partial \delta^T) & \partial^2 l_t/(\partial \delta \partial \delta^T) \end{pmatrix} \xrightarrow{\text{a.s.}} \begin{pmatrix} \mathbf{\Omega}_\gamma & O \\ O & \mathbf{\Omega}_\delta \end{pmatrix}, \quad (18)$$

as $n \to \infty$ and $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\delta$ are positive matrices, where

$$\mathbf{\Omega}_\gamma = E \left[\frac{1}{h_t} \frac{\partial \varepsilon_t}{\partial \gamma} \frac{\partial \varepsilon_t}{\partial \gamma^T} + \frac{1}{2h_t^2} \frac{\partial h_t}{\partial \gamma} \frac{\partial h_t}{\partial \gamma^T} \right],$$

and

$$\mathbf{\Omega}_\delta = E \left[\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \delta} \frac{\partial h_t}{\partial \delta^T} \right], \quad (19)$$

and $\xrightarrow{\text{a.s.}}$ denotes convergence almost surely.

Theorem 3.2. Under the assumptions of Theorem 3.1, the following hold:

a. There exists a MLE $\hat{\lambda}_n$ such that it satisfies the equation $\partial L(\lambda)/\partial\lambda = 0$ and $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ as $n \to \infty$.

b. For such a sequence, $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(O, \mathbf{\Omega}_0^{-1})$ as $n \to \infty$, where $\xrightarrow{p}$ and $\xrightarrow{D}$ denote convergences in probability and in distribution, $\mathbf{\Omega}_0 = \mathrm{diag}(\mathbf{\Omega}_{\gamma 0}, \mathbf{\Omega}_{\delta 0})$, and $\mathbf{\Omega}_{\gamma 0}$ and $\mathbf{\Omega}_{\delta 0}$ are values of $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\delta$ at $\lambda = \lambda_0$. Further, the information matrices $\mathbf{\Omega}_{\gamma 0}$ and $\mathbf{\Omega}_{\delta 0}$ can be estimated consistently by

$$\hat{\mathbf{\Omega}}_\gamma = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{h_t}\frac{\partial\varepsilon_t}{\partial\gamma}\frac{\partial\varepsilon_t}{\partial\gamma^T} + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial\gamma}\frac{\partial h_t}{\partial\gamma^T}\right]$$

and

$$\hat{\mathbf{\Omega}}_\delta = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\delta}\frac{\partial h_t}{\partial\delta^T}\right], \tag{20}$$

where the various terms are evaluated at $\lambda = \hat{\lambda}_n$.

Because the off-diagonal blocks of the information matrix, $\sum E[(\partial^2 l_t)/(\partial\lambda\partial\delta^T)]$, are 0, γ and δ can be estimated separately without an asymptotic loss in efficiency. In practice, it is sometimes necessary to use a variable-step length procedure, such as the Levenberg–Marquardt adjustment algorithm (Goldfeld, Quandt, and Trotter 1966; Thisted 1988). In the case where model (7)–(8) includes unknown drift μ (as in Li and Mcleod 1986), the series may be centered by the sample mean and then estimated by the foregoing procedure. Some simulation results in Section 5 illustrate that the estimators obtained by this method are very close to those obtained when the mean $\mu = 0$ is 0 and known.

3.2 The Nonstationary Case

For $d < 1/2$, the FARIMA(p,d,q) model is stationary and ergodic. For $d \geq 1/2$, it is well known that the FARIMA(p,d,q) models are not stationary. How to estimate the nonstationary FARIMA model and how to check for the adequacy of nonstationary FARIMA model are obviously important problems. Beran (1995) proposed an approximate MLE for nonstationary FARIMA model with a constant conditional variance. However, as far as we know, there are as yet no results for the latter problem. In the following, we generalize Beran's (1995) approximate ML estimation procedure for the nonstationary FARIMA(p,d,q)–GARCH(r,s) model. The diagnostic checking procedure can be found in the next section.

Suppose that $d = m + d_1$, where $-1/2 < d_1 < 1/2$ and m is a positive integer. Then Equation (7) can be written as

$$\phi(B)(1-B)^{d_1}(1-B)^m Y_t = \theta(B)\varepsilon_t, \tag{21}$$

where ε_t is defined by (8). Denoting $U_t = (1-B)^m Y_t$, then U_t follows the equation

$$\phi(B)(1-B)^{d_1}U_t = \theta(B)\varepsilon_t. \tag{22}$$

This means that, after mth order differencing, the resulting process is a stationary FARIMA(p,d_1,q)–GARCH(r,s) model. The estimation and diagnostic checking procedures in the stationary case can be used. However, note that $d > -1/2$, by Theorem 2.3b, as all roots of $\phi(B)$ and $\theta(B)$ are outside the unit circle, ε_t has the following representation:

$$\varepsilon_t(\gamma) = \sum_{k=0}^{\infty} a_k(\gamma)Y_{t-k}, \tag{23}$$

where $\gamma = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,d)^T$ and the $a_k(\gamma)$'s are continuously differentiable with respect to γ. On the other hand, by (22), ε_t has the following representation:

$$\varepsilon_t(\gamma^*) = \sum_{k=0}^{\infty}\tilde{a}_k(\gamma^*)U_{t-k}, \tag{24}$$

where $\gamma^* = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,d_1)^T$ and the $\tilde{a}_k(\gamma^*)$'s are continuously differentiable with respect to γ^*. Given $Y_1,\ldots,Y_n$ and assuming that $Y_t = 0$ if $t \leq 0$, ε_t can be estimated by $e_t(\gamma) = \sum_{k=0}^{t-1} a_k(\gamma)Y_{t-k}$ or $e_t(\gamma^*) = \sum_{k=0}^{t-1}\tilde{a}_k(\gamma^*)U_{t-k}$, where $t = 1,\ldots,n$. Similarly, h_t can be estimated by $h_t(\lambda) = \alpha_0 + \sum_{i=1}^{r}\alpha_i e_{t-i}^2(\gamma) + \sum_{i=1}^{s}\beta_i h_{t-i}(\lambda)$ or $h_t(\lambda^*) = \alpha_0 + \sum_{i=1}^{r}\alpha_i e_{t-i}^2(\gamma^*) + \sum_{i=1}^{s}\beta_i h_{t-i}(\lambda^*)$, where $t = 1,\ldots,n$ and $h_t(\lambda) = \sum_{t=1}^{n} e_t(\gamma)/n$ and $h_t(\lambda^*) = \sum_{t=1}^{n} e_t(\gamma^*)/n$ if $t \leq 0$. $\lambda = (\gamma^T,\delta^T)^T$, $\lambda^* = (\gamma^{*T},\delta^T)^T$, and δ is as defined in Section 3.1. Again, λ and its true value λ_0 are interior points of a compact set. Note that $\varepsilon_t(\gamma) = \varepsilon_t(\gamma^*)$; therefore, $e_t(\gamma) = e_t(\gamma^*)$ and $h_t(\lambda) = h_t(\lambda^*)$. This suggests that we can estimate λ directly by an approximate MLE, $\hat{\lambda}_n$, which maximizes (11). Let $l_t(\lambda) = -(1/2)\ln h_t(\lambda) - e_t^2(\gamma)/[2h_t(\lambda)]$. Note that $L(\lambda) = L(\lambda^*) = -(1/n)\sum_{t=1}^{n} l_t(\lambda^*)$. It is easy to show that

$$\frac{\partial l_t(\lambda)}{\partial\lambda_i} = \frac{\partial l_t(\lambda^*)}{\partial\lambda_i^*},$$

$$\frac{\partial l_t^2(\lambda)}{\partial\lambda_i\partial\lambda_j} = \frac{\partial l_t^2(\lambda^*)}{\partial\lambda_i^*\partial\lambda_j^*},$$

and

$$\frac{\partial l_t^3(\lambda)}{\partial\lambda_i\partial\lambda_j\partial\lambda_k} = \frac{\partial l_t^3(\lambda^*)}{\partial\lambda_i^*\partial\lambda_j^*\partial\lambda_k^*},$$

where λ_i and λ_i^* are the ith components of λ and λ^*.

Because $\{U_t\}$ is a stationary time series, by Theorems 3.1 and 3.2, we have the following theorems.

Theorem 3.3. Suppose that $\{Y_t\}$ and $\{\varepsilon_t\}$ are generated by (7)–(8) with $\alpha_0 > 0$, $\alpha_1,\ldots,\alpha_r$, $\beta_1,\ldots,\beta_s \geq 0$, $\sum_{i=1}^{r}\alpha_i + \sum_{i=1}^{s}\beta_i < 1$, $E(\varepsilon_t^4) < \infty$, and $d = m + d_1$ and that all roots of $\phi(B)$ and $\theta(B)$ lie outside the unit circle. Then

$$-\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}\partial^2 l_t(\lambda)/(\partial\gamma\partial\gamma^T) & \partial^2 l_t(\lambda)/(\partial\delta\partial\gamma^T)\\ \partial^2 l_t(\lambda)/(\partial\gamma\partial\delta^T) & \partial^2 l_t(\lambda)/(\partial\delta\partial\delta^T)\end{pmatrix} \xrightarrow{\text{a.s.}} \begin{pmatrix}\mathbf{\Omega}_\gamma & O\\ O & \mathbf{\Omega}_\delta\end{pmatrix}$$

as $n \to \infty$ and $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\delta$ are positive matrices, where $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\delta$ are defined in (19) and m and d_1 are defined in (21).

Theorem 3.4. Under the assumptions of Theorem 3.3:

a. There exists a MLE $\hat{\lambda}_n$ satisfying the equation $\partial L(\lambda)/\partial\lambda = 0$ and $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ as $n \to \infty$.
b. For such a sequence, $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{D} N(O, \mathbf{\Omega}_0^{-1})$ as $n \to \infty$, where $\mathbf{\Omega}_0 = \operatorname{diag}(\mathbf{\Omega}_{\gamma 0}, \mathbf{\Omega}_{\delta 0})$, $\mathbf{\Omega}_{\gamma 0}$ and $\mathbf{\Omega}_{\delta 0}$ are values of $\mathbf{\Omega}_\gamma$ and $\mathbf{\Omega}_\delta$ at $\lambda = \lambda_0$. Further, the information matrices $\mathbf{\Omega}_{\gamma 0}$ and $\mathbf{\Omega}_{\delta 0}$ can be estimated consistently by $\hat{\mathbf{\Omega}}_\gamma$ and $\hat{\mathbf{\Omega}}_\delta$ as in (20) with h_t and ε_t replaced by $h_t(\lambda)$ and $\varepsilon_t(\lambda)$.

In the case where $(1-B)^m Y_t$ in model (21) includes a nonzero unknown drift μ (i.e., Y_t follows the equation $\phi(B)(1-B)^{d_1}\{(1-B)^m Y_t - \mu\} = \theta(B)\varepsilon_t$), the time series $\{Y_t\}$ is not stationary, and hence $n^{-1}\sum_{t=1}^n Y_t$ does not converge to the mean μ. Following Beran (1995), we can use a simple method to overcome this. Let $U_t = (1-B)^m Y_t$ and $\hat{U} = (n-m)^{-1}\sum_{t=m+1}^n U_t$; then $\hat{U}$ is a consistent estimate of μ. The residual ε_t can be estimated by the following adjusted residual: $e_t(\lambda) = \sum_{k=0}^{t-1} a_k(\lambda^*)(U_{t-k} - \hat{U})$. All other quantities defined earlier can be defined analogously. As shown by simulation in Section 5, the estimates obtained by adjusted residuals are very close to those in the case with mean μ known and equal to 0.

Remark. If $h_t =$ a positive constant, then the foregoing estimate procedure reduces to that given by Beran (1995) for the FARIMA(p, d, q) model with a constant conditional variance.

4. STATISTICS FOR DIAGNOSTIC CHECKING

In model building, identification and diagnostic checking are two important steps. Identification of the FARIMA(p, d, q)–GARCH(r, s) model can be divided into two parts. We can first identify the order (p, d, q) of the conditional mean equation, then identify the order (r, s) of the conditional variance. The former can be done by the Akaike information criterion (AIC) procedure given by Hosking (1984), and the latter can be done by the procedure given by Bollerslev (1986). In this section we mainly consider diagnostic checking of the fitted model.

Residual autocorrelations have been found to be useful in checking the adequacy of ARMA models (see Box and Jenkin 1976 and Ljung and Box 1978). Li (1992) obtained the asymptotic distribution of residual autocorrelations for general nonlinear time series. Li and Mak (1994) proposed a formal diagnostic checking tool for nonlinear time series with conditional heteroscedasticity. Following their ideas, we derive the asymptotic standard error of residual autocorrelations and squared residual autocorrelations of the FARIMA–GARCH model and construct two chi-squared statistics, $Q(M)$ and $Q^2(M)$, for checking model adequacy. In particular, the results in this section can be applied to nonstationary FARIMA(p, d, q)–GARCH(r, s) models.

Denoting $\mu_t = \sum_{i=1}^p \phi_i(Y_{t-i} - \mu) + E[\theta(B)(1-B)^{-d}\varepsilon_t | F_{t-1}]$, $h_t = \alpha_0 + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^s \beta_i h_{t-i}$, $\varepsilon_t = \theta^{-1}(B)\phi(B)(1-B)^d(Y_t - \mu)$, and $\hat{\varepsilon}_t$ is the corresponding residual when λ is replaced by the MLE $\hat{\lambda}_n$ in Section 3.1 or Section 3.2. Similarly, we define $\hat{\mu}_t$ and $\hat{h}_t$. The lag k standardized residual autocorrelation is defined as $\hat{\rho}_k = n^{-1}\sum_{t=k+1}^n (\hat{\varepsilon}_t/\sqrt{\hat{h}_t} - \tilde{\varepsilon})(\hat{\varepsilon}_{t-k}/\sqrt{\hat{h}_{t-k}} - \tilde{\varepsilon})$ for $k = 1, 2, 3, \ldots$, where $\tilde{\varepsilon} = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t/\sqrt{\hat{h}_t}$. Let $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)^T$. We then have the following theorem.

Theorem 4.1. Under the assumptions of Theorem 3.1 or 3.3, $\sqrt{n}\hat{\rho}$ is asymptotically normal with mean 0 and covariance matrix

$$\mathbf{V}_1 = \mathbf{I}_M - Y\mathbf{\Omega}_0^{-1}Y^T, \tag{25}$$

where $Y = (Y_1, \ldots, Y_M)^T$, $Y_k = -E[(\varepsilon_{t-k}/\sqrt{h_t h_{t-k}})(\partial\mu_t/\partial\gamma)]|_{\lambda=\lambda_0}$, and $\mathbf{I}_M$ is the $M \times M$ identity matrix.

The lag k squared standardized residual autocorrelation is defined as

$$\hat{r}_k = \frac{\sum(\hat{\varepsilon}_t^2/\hat{h}_t - \tilde{\varepsilon}^{*2})(\hat{\varepsilon}_{t-k}^2/\hat{h}_{t-k} - \tilde{\varepsilon}^{*2})}{\sum(\hat{\varepsilon}_t^2/\hat{h}_t - \tilde{\varepsilon}^{*2})^2}$$

for $k = 1, 2, \ldots$, where $\tilde{\varepsilon}^{*2} = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^2/\hat{h}_t$. If the model is correct, then $\tilde{\varepsilon}^{*2}$ converges to 1 almost surely, so that $\hat{r}_k$ can be replaced by

$$\hat{r}_k = \frac{\sum(\hat{\varepsilon}_t^2/\hat{h}_t - 1)(\hat{\varepsilon}_{t-k}^2/\hat{h}_{t-k} - 1)}{\sum(\hat{\varepsilon}_t^2/\hat{h}_t - 1)}$$

for $k = 1, 2, \ldots$. Letting $\hat{r} = (\hat{r}_1, \ldots, \hat{r}_M)^T$, we have the following theorem.

Theorem 4.2. Under the assumptions of Theorem 3.1 or 3.3, $\sqrt{n}\hat{r}$ is asymptotically normal with mean 0 and covariance matrix

$$\mathbf{V}_2 = \mathbf{I}_M - \frac{1}{4}X\mathbf{\Omega}_0^{-1}X^T, \tag{26}$$

where $X = (X_1, \ldots, X_M)^T$, $X_k = -E[(1/h_t)(\partial h_t/\partial\lambda)(\varepsilon_{t-k}^2/h_{t-k} - 1)]|_{\lambda=\lambda_0}$, and $\mathbf{I}_M$ is the $M \times M$ identity matrix.

The main spirit of the proofs of Theorems 4.1 and 4.2 is maintained by using Taylor expansions of $\hat{\rho}_k$ and $\hat{r}_k$ at the true value λ_0, the large-sample properties of the MLE $\hat{\lambda}_n$ given by Theorems 3.2 and 3.4, and the stationarity and ergodicity of the residual process $\{\varepsilon_t\}$. In the nonstationary case, although the observed process $\{Y_t\}$ is nonstationary, the residual process $\varepsilon_t = \theta^{-1}(B)\phi(B)(1-B)^d(Y_t - \mu) = \theta^{-1}(B)\phi(B)(1-B)^{d_1}U_t$ is still stationary and ergodic. Thus all details are similar to arguments of Li (1992) and Li and Mak (1994) and hence are omitted.

By Theorems 4.1 and 4.2, we know that

$$Q(M) = n\hat{\rho}^T\hat{\mathbf{V}}_1^{-1}\hat{\rho} \sim \chi^2(M)$$

and

$$Q^2(M) = n\hat{r}^T\hat{\mathbf{V}}_2^{-1}\hat{r} \sim \chi^2(M), \tag{27}$$

where $\hat{\mathbf{V}}_1 = \mathbf{I} - \hat{Y}\hat{\mathbf{\Omega}}_0^{-1}\hat{Y}^T$, $\hat{\mathbf{V}}_2 = \mathbf{I} - \hat{X}\hat{\mathbf{\Omega}}_0^{-1}\hat{X}^T/4$, $\hat{\mathbf{\Omega}}_0 = \operatorname{diag}(\hat{\mathbf{\Omega}}_\gamma, \hat{\mathbf{\Omega}}_\delta)$, $\hat{\mathbf{\Omega}}_\gamma$ and $\hat{\mathbf{\Omega}}_\delta$ are defined as in Theorems 3.2 and 3.4 for the stationary and nonstationary cases, $\hat{Y} =$

Table 1. *Estimated Parameter Bias and Square Root of the Mean Squared Error for the FARIMA(1, d, 0)–GARCH(1, 1) Model (500 Replications)*

d	ϕ_1	α_0	α_1	β_1		$\hat{d}$	$\hat{\phi}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
					T = 200/Known mean					
.3	.5	.4	.3	.3	BIAS	−.027	−.025	.045	.028	−.072
					$\sqrt{\text{MSE}}$	.135	.148	.160	.107	.171
					$\sqrt{\text{MAV}}$	.176	.165	.186	.123	.227
.3	−.5	.4	.3	.3	BIAS	−.031	.025	.081	.005	−.092
					$\sqrt{\text{MSE}}$	.083	.082	.174	.125	.200
					$\sqrt{\text{MAV}}$	.076	.067	.219	.119	.260
					T = 200/Unknown mean					
.3	.5	.4	.3	.3	BIAS	−.068	.029	.049	.015	−.077
					$\sqrt{\text{MSE}}$	.196	.190	.158	.113	.185
					$\sqrt{\text{MAV}}$	.162	.175	.194	.121	.239
.3	−.5	.4	.3	.3	BIAS	−.023	.018	−.094	.012	−.108
					$\sqrt{\text{MSE}}$	.067	.077	.167	.121	.196
					$\sqrt{\text{MAV}}$	.074	.064	.208	.121	.243
					T = 400/Known mean					
.3	.5	.4	.3	.3	BIAS	−.010	.002	−.035	.001	−.064
					$\sqrt{\text{MSE}}$	.128	.116	.154	.089	.167
					$\sqrt{\text{MAV}}$	.112	.103	.146	.085	.170
.3	−.5	.4	.3	.3	BIAS	.000	.008	−.078	.003	.086
					$\sqrt{\text{MSE}}$	.046	.056	.146	.090	.173
					$\sqrt{\text{MAV}}$	.052	.046	.142	.085	.167
					T = 400/Unknown mean					
.3	.5	.4	.3	.3	BIAS	−.021	.007	−.069	−.007	−.070
					$\sqrt{\text{MSE}}$	.123	.114	.144	.087	.171
					$\sqrt{\text{MAV}}$	.117	.115	.137	.085	.164
.3	−.5	.4	.3	.3	BIAS	−.013	.014	.073	.003	−.082
					$\sqrt{\text{MSE}}$	.049	.058	.140	.089	.166
					$\sqrt{\text{MAV}}$	.053	.047	.143	.086	.168

$(\hat{Y}_1, \ldots, \hat{Y}_M)^T$, $\hat{Y}_k = -n^{-1}\sum_{t=k+1}^{n}[(\varepsilon_{t-k}/\sqrt{h_t h_{t-k}})\partial\mu_t/\partial\theta]$ is the estimated value of Y_k at $\lambda = \hat{\lambda}_n$, $\hat{X} = (\hat{X}_1, \ldots, \hat{X}_M)^T$, and $\hat{X}_k = -n^{-1}\sum_{t=k+1}^{n}[(1/h_t)(\partial h_t/\partial\lambda)(\varepsilon_{t-k}^2/h_{t-k}-1)]$ is the estimated value of X_k at $\lambda = \hat{\lambda}_n$. The statistics $Q(M)$ and $Q^2(M)$ can be used for testing the joint significance of $\hat{\rho}_i$ and $\hat{r}_i$, $i = 1, \ldots, M$. Similar to the idea of Li and Mak (1994), if $\beta_1 = \cdots = \beta_s = 0$ and $M > k$, then the entries of X are approximately 0 from the $(r+1)$th row onward. The standard errors of r_i, $i = r+1, \ldots, M$ are then just $1/\sqrt{n}$ and $Q^2(r, M) = n\sum_{i=r+1}^{M}\hat{r}_i^2 \sim \chi^2(M-r)$. Hence $Q^2(r, M)$ can be used as a portmanteau statistic for testing the overall significance of $\hat{r}_i$, $i = r+1, \ldots, M$. In addition, if $\alpha_1 = \cdots = \alpha_r = \beta_1 = \cdots = \beta_s = 0$, then $X = O$, $\mathbf{V}_2 = \mathbf{I}$, and $Q^2(M) = n\sum_{i=1}^{M}\hat{r}_i^2 \sim \chi^2(M)$. In this case $Q^2(M)$ can be used to test whether or not the FARIMA model has GARCH disturbance.

5. SIMULATION RESULTS FOR THE ML ESTIMATION

First, we consider the FARIMA$(1, d, 0)$–GARCH(1, 1) models with parameter sets $(d, \phi_1, \alpha_0, \alpha_1, \beta) = (.3, .5, .4, .3, .3)$ and $(.3, -.5, .4, .3, .3)$. For each parameter set, the mean μ known and μ estimated by the sample mean cases are compared. For each case, the number of replications is 500 with lengths of realization 200 and 400. In the simulation, we use the Levenberg–Marquardt adjustment algorithm, which has better convergence results. The simulation results are summarized in Table 1. In the table, the true parameter values used in the data-generating process are given in the first four columns. The estimated biases are given on the same row, and the corresponding empirical root mean squared errors ($\sqrt{\text{MSE}}$) of the estimates are given in the row below. The root mean asymptotic variance ($\sqrt{\text{MAV}}$) calculated by (19) are given under the row of empirical root mean squared errors. From Table 1, we see that the biases are generally small, and the empirical root mean squared errors and the root mean asymptotic variances are very close in all cases. When the series is centered by the sample mean, both biases and root mean squared errors are not substantially changed for the sample sizes considered. For estimated parameters in the conditional variance equation, the effect of the estimated μ on biases and root mean squared errors is also negligible. As sample size increases, all biases, empirical root mean squared errors, and root mean asymptotic variances decrease. The root mean asymptotic variances and the empirical root mean squared errors also become much closer at $n = 400$. These are consistent with our expectations. We also considered the FARIMA$(0, d, 1)$–GARCH(1, 1) model and found very similar results. (These results are available on request.)

Next we simulate nonstationary FARIMA$(0, d, 0)$–GARCH(1, 1) models with parameter sets $(d, \alpha_0, \alpha_1, \beta_1) = (.7, .3, .3, .3)$, $(1.0, .2, .2, .2)$, $(1.2, .25, .25, .25)$, $(1.4, .3, .3, .3)$ and $(2.2, .2, .2, .2)$. For each parameter set, we consider two

Table 2. Estimated Parameter Bias and Square Root of the Mean Squared Error for Non-Stationary FARIMA(0, d, 0)–GARCH(1, 1) Model (500 Replications)

					T = 200				T = 400			
d	α_0	α_1	β_1		$\hat{d}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{d}$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
.7	.3	.3	.3	BIAS	.006	.036	−.018	−.048	.006	.025	−.006	−.021
				$\sqrt{\text{MSE}}$	.058	.129	.107	.204	.041	.107	.083	.169
				$\sqrt{\text{MAV}}$	.057	.144	.121	.220	.040	.105	.083	.165
1.0	.2	.2	.2	BIAS	.002	−.008	.027	−.021	.001	−.002	−.004	−.017
				$\sqrt{\text{MSE}}$	.058	.076	.090	.245	.047	.070	.070	.227
				$\sqrt{\text{MAV}}$	.058	.111	.109	.346	.041	.085	.075	.267
1.2	.25	.25	.25	BIAS	.002	.010	.018	−.014	.002	.013	−.006	−.011
				$\sqrt{\text{MSE}}$	.060	.100	.103	.227	.040	.087	.079	.200
				$\sqrt{\text{MAV}}$	.058	.129	.114	.276	.041	.097	.079	.211
1.4	.3	.3	.3	BIAS	.006	.072	.010	−.078	−.008	.052	−.011	−.047
				$\sqrt{\text{MSE}}$	.061	.147	.118	.225	.041	.115	.085	.176
				$\sqrt{\text{MAV}}$	.057	.161	.119	.236	.040	.114	.083	.175
2.2	.2	.2	.2	BIAS	.001	.007	−.032	−.014	.001	.010	−.000	−.011
				$\sqrt{\text{MSE}}$	.062	.084	.094	.261	.047	.072	.073	.230
				$\sqrt{\text{MAV}}$	.058	.118	.120	.351	.041	.094	.074	.288

different sample sizes: 200 and 400. The number of replications is 500 in all cases. The results are summarized in Table 2. The results demonstrate that the estimation method in Section 3.2 is applicable to the nonstationary situation. For each parameter set and sample size, the estimated parameter bias is negligible, and the empirical root mean squared error is close to that given by the asymptotic theory. We have also considered the case where $\mu = 1$ and is unknown. We found that the adjusted residual has negligible effect on estimated parameters. These results are similar to those in Table 1 and thus are not reported here.

6. SIZE AND POWER OF THE GOODNESS-OF-FIT TESTS

We performed some simulation experiments to examine the empirical size and power of the goodness of fit statistics $Q(M)$ and $Q(M^2)$. Due to space limitations we report results only on the nonstationary case. Results for the stationary case are similar and are available on request.

Three pairs of data are generated by the model $(1 - B)^d X_t = \varepsilon_t$ with conditional variance $h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \alpha_3\varepsilon_{t-3}^2 + \beta_1 h_{t-1}$. Three pairs of parameter sets are used: (.7, .3, .3, .3, 0, 0) and (.7, .3, .3, .3, .2, .05), (1.4, .3, .3, .3, 0, 0) and (1.4, .3, .3, .3, .2, .05), and (2.2, .2, .2, .2, 0, 0) and (2.2, .2, .2, .2, .2, .05). Results for these three pairs of parameter sets are reported in Table 3 corresponding to model 1, model 2, and model 3. For each

Table 3. The Empirical Sizes and Power of Q(M) and $Q^2(M)$ M = 6, 500 Replications

		Size		Power	
		$Q(M)$	$Q^2(M)$	$Q(M)$	$Q^2(M)$
Model 1	n = 400	.055	.024	.768	.781
	n = 500	.045	.022	.786	.798
Model 2	n = 400	.064	.072	.775	.767
	n = 500	.044	.054	.788	.791
Model 3	n = 400	.075	.077	.764	.760
	n = 500	.060	.064	.779	.769

parameter set, two different sample sizes (i.e., 400 and 500) are considered. In all cases, the number of replications is 500. The FARIMA$(0, d, 0)$–GARCH(1, 1) model is fitted to all data. The results are summarized in Table 3, with the entries equal to the proportion of rejections based on the upper fifth percentile of the chi-squared distribution with 6 degrees of freedom. Both statistics are slightly sensitive, but the average sizes seem rather acceptable. The powers of $Q^2(M)$ and $Q(M)$ averaging over .76 are also reasonably high.

7. AN ILLUSTRATIVE EXAMPLE

The dataset that we analyze in this section is the daily Hong Kong Hang Seng index. There are altogether 495 observations from January 3, 1983, to December 31, 1984. The log index is denoted by r_t $(t = 1, \ldots, 495)$. Note that in 1983, the Chinese government and the British government were discussing the sovereignty issue of Hong Kong. The stock market in Hong Kong suffered several dramatic shocks. The effect of these shocks appears to continue into the following year. Therefore, it is reasonable to expect a long-memory model to apply in this period. Again because of the shocks, there are a few outliers in the return series. To avoid the effect of these outliers, those r_t values larger than 3σ (where σ is the standard derivation of the r_t series) are replaced by 3σ. The r_t series are estimated by the FARIMA$(1, d, 0)$ model with constant variance, the FARIMA$(1, d, 0)$–ARCH(1) model, the FARIMA$(1, d, 0)$–ARCH(2) model, and the FARIMA$(1, d, 0)$–GARCH(1, 1) model. The approximate ML estimation given in Section 3 is used. For each model, the known mean (with restriction $\mu = 0$) and unknown mean cases are considered. The usual likelihood ratio test suggests that the mean is insignificant. The modeling results for the mean $\mu = 0$ case are as follows:

Model 1: $(1 - B)^{1.2677}(1 + .2465B)r_t = \varepsilon_t$, $h_t = .1594 \times 10^{-3}$, log-likelihood value = 4,080, where the estimated standard errors of d, ϕ_1, and α_0 are .0567, .0567, and $.1017 \times 10^{-4}$, $Q(25) = 48.27$ and $Q^2(25) = 48.27$.

Model 2: $(1-B)^{1.2666}(1+.2390B)r_t = \varepsilon_t$, $h_t = .1607 \times 10^{-3} + .0160\varepsilon_{t-1}^2$, log-likelihood value $= 4{,}080$, where the estimated standard errors of d, ϕ_1, α_0, and α_1 are .0255, .0291, $.7925 \times 10^{-5}$, and $.1270 \times 10^{-4}$, $Q(25) = 48.66$ and $Q^2(25) = 48.75$.

Model 3: $(1-B)^{1.3171}(1+.2760B)r_t = \varepsilon_t$, $h_t = .1351 \times 10^{-3} + .0306\varepsilon_{t-1}^2 + .1455\varepsilon_{t-2}^2$, log-likelihood value $= 4{,}090$, where the estimated standard errors of d, ϕ_1, α_0, α_1, and α_2 are .0336, .0340, $.9474 \times 10^{-5}$, $.6612 \times 10^{-4}$, and $.6612 \times 10^{-4}$, $Q(25) = 51.54$ and $Q^2(25) = 51.72$.

Model 4: $(1-B)^{1.2781}(1+.2373B)r_t = \varepsilon_t$, $h_t = .4143 \times 10^{-4} + .0826\varepsilon_{t-1}^2 + .6630h_{t-1}$, log-likelihood value $= 4{,}095$, where the estimated standard errors of d, ϕ_1, α_0, α_1, and β_1 are .0401, .0478, $.2298 \times 10^{-5}$, $.3324 \times 10^{-5}$, and $.3324 \times 10^{-5}$, $Q(25) = 21.71$ and $Q^2(25) = 21.72$.

From these fitted models, we see that model 4—the FARIMA$(1, d, 0)$–GARCH(1, 1) model—has the largest log-likelihood value. For models 1–3, the values of the $Q(M)$ and $Q^2(M)$, $M = 25$, statistics are larger than the corresponding critical values at significance level .05, and hence these models are clearly rejected. Other values of M give similar results. For model 4, the values of all test statistics are smaller than the corresponding critical values at significance level .05, and hence model 4 seems adequate for the r_t series. However, all of the estimates for d seem to suggest a value of d around 1.28. This implies that the returns of the index, defined as the first-order differences of r_t, will have a value of d around .28. Thus long memory may be present in the return series. Note that here the order of differencing is automatically taken care of by the estimation procedure. The individual residual autocorrelation and squared residual autocorrelation also suggest that model 4 is adequate for fitting the r_t series. These are not reported here.

8. CONCLUSIONS

In this article we have discussed a class of FARIMA models that include nonstationary integrated cases and allow the innovations to follow the GARCH processes. Some properties of the proposed model and estimation and diagnostic procedures were presented. The application of the proposed model to the Hong Kong Hang Seng index (1983–1984) seems to illustrate that the long-memory phenomenon may exist simultaneously in the stock return and in the volatility of the stock return. We believe that this phenomenon may exist in many other time series. The results of this article may be useful in modeling time series that exhibit long memory in both the conditional mean and the conditional variance.

APPENDIX: PROOFS OF THEOREMS

Proof of Theorem 2.1

Because ε_t is conditionally normal, by (2) we can write ε_t as

$$\varepsilon_t^2 = \alpha_0 Z_t^2 + \sum_{i=1}^{r} \alpha_i Z_t^2 \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i Z_t^2 h_{t-i}, \qquad \text{(A.1)}$$

where Z_t is a standard normal random variable independent of F_{t-1}.

We rewrite (A.1) in vector form as

$$\tilde{\varepsilon}_t = A_t \tilde{\varepsilon}_{t-1} + \xi_t, \qquad \text{(A.2)}$$

where $\tilde{\varepsilon}_t = (\varepsilon_t^2, \ldots, \varepsilon_{t-r+1}^2, h_t, \ldots, h_{t-s+1})^T$, ξ_t is defined as in (4), and A_t is defined by (6). Consider that

$$S_{n,t} = \xi_t + \sum_{j=1}^{n} \left(\prod_{i=0}^{j-1} A_{t-i} \right) \xi_{t-j}, \qquad \text{(A.3)}$$

where $n = 1, 2, \ldots$. Let $(s_{n,t})_k$ be the kth element of $(\prod_{i=0}^{j-1} A_{t-i})\xi_{t-j}$. Then, because $\{Z_t\}$ is a sequence of iid variables and each element of A_t and ξ_t is nonnegative,

$$\begin{aligned} E|(s_{n,t})_k| &= \boldsymbol{\eta}_k^T E\left[\left(\prod_{i=0}^{j-1} A_{t-i} \right) \xi_{t-j} \right] \\ &= \boldsymbol{\eta}_k^T \left[\prod_{i=0}^{j-1} E(A_{t-i}) \right] E(\xi_{t-j}) = \boldsymbol{\eta}_k^T A^j \mathbf{c}_1, \end{aligned} \qquad \text{(A.4)}$$

where $\boldsymbol{\eta}_k = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ with 1 appears at the kth position, $\mathbf{c}_1 = E\xi_t$ is a constant vector, and

$$\mathbf{A} = \left(\begin{array}{ccc|ccc} \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ & \mathbf{I}_{(r-1)\times(r-1)} & O_{(r-1)\times 1} & & O_{(r-1)\times s} & \\ \hline \alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_s \\ & O_{(s-1)\times r} & & & \mathbf{I}_{(s-1)\times(s-1)} & O_{(s-1)\times 1} \end{array} \right). \qquad \text{(A.5)}$$

By direct calculation, we know that the characteristic polynomial of $\mathbf{A}$ is

$$f(\lambda) = \lambda^{r+s} - \lambda^s \sum_{i=1}^{r} \alpha_i \lambda^{r-i} - \lambda^r \sum_{i=1}^{s} \beta_i \lambda^{s-i}.$$

Now it follows from the inequality $|a - b| \geq |(|a| - |b|)|$ that if $|\lambda| \geq 1$, then

$$|f(\lambda)| \geq |\lambda|^{r+s} \left| 1 - \sum_{i=1}^{r} \frac{\alpha_i}{|\lambda|^i} - \sum_{i=1}^{s} \frac{\beta_i}{|\lambda|^i} \right| \geq 1 - \sum_{i=1}^{r} \alpha_i - \sum_{i=1}^{s} \beta_i.$$

Hence $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$ implies that the eigenvalues of $\mathbf{A}$ lie inside the unit circle; that is, the spectral radius ρ of $\mathbf{A}$ is less than 1. The right side of (A.4) is less than $c\rho^j$ for some constant c. Thus $S_{n,t}$ converges almost surely as $n \to \infty$. Denoting the limit of $S_{n,t}$ as $\tilde{\varepsilon}_t$, then, by the foregoing argument, the first-order moment of $\tilde{\varepsilon}_t$ exists and

$$\tilde{\varepsilon}_t = \xi_t + \sum_{j=1}^{\infty} \left(\prod_{i=0}^{j-1} A_{t-i} \right) \xi_{t-j}. \qquad \text{(A.6)}$$

It is easy to verify that $\tilde{\varepsilon}_t$ satisfies (A.2). Hence there exists a F_t-measurable second-order stationary solution $\varepsilon_t = Z_t\sqrt{h_t} = Z_t(\boldsymbol{\eta}_{r+1}^T\tilde{\varepsilon}_t)^{1/2}$ to (2), and this solution has the representation given by (4), where η_k is defined by (A.4).

For such a F_t-measurable second-order stationary solution $\{\varepsilon_t\}$, consider $Y_{n,t} = \sum_{k=0}^n a_k\varepsilon_{t-k}$, where $a_k = (k+d-1)!/(k!(d-1)!)$. By Stirling's formula, $\lim_{k\to\infty} k^{d-1}/a_k = c$, where c is a positive constant. It is easy to show that $\sum_{k=0}^\infty a_k^2 < \infty$. So $\{Y_{n,t}\}$ converges in mean squares, and hence the second-order moment of Y_t defined by (5) is finite. By the binomial expansion of $(1-z)^{-d}$, we know that $\{Y_t\}$ satisfies (1) and has the representation (5).

To show that the solution $\{\varepsilon_t, Y_t\}$ of (1)–(2) is unique, let $\{\varepsilon_t', Y_t'\}$ be another F_t-measurable second-order stationary solution of (1)–(2). Similar to (A.2), we have $\tilde{\varepsilon}_t' = A_t\tilde{\varepsilon}_{t-1}' + \xi_t$, where $\tilde{\varepsilon}_t' = (\varepsilon_t'^2, \ldots, \varepsilon_{t-r+1}'^2, h_{t-1}', \ldots h_{t-s+1}')^T$, $h_t' = \alpha_0 + \sum_{i=1}^r \alpha_i\varepsilon_{t-i}'^2 + \sum_{i=1}^s \beta_i h_{t-i}'$, and ξ_t is defined by (4). Let $U_t = \tilde{\varepsilon}_t - \tilde{\varepsilon}_t'$; then U_t is first-order stationary and, by (A.2), $U_t = A_tU_{t-1} = (\prod_{i=0}^n A_{t-i})U_{t-n-1}$.

Denoting the kth component of U_t as $u_{k,t}$, then, because each element of A_t for all t is nonnegative,

$$|u_{k,t}| = \left|\boldsymbol{\eta}_k^T\left(\prod_{i=0}^n A_{t-i}\right)U_{t-n-1}\right| \le \boldsymbol{\eta}_k^T\left(\prod_{i=0}^n A_{t-i}\right)|U_{t-n-1}|, \tag{A.7}$$

where $\boldsymbol{\eta}_k$ is defined as in (A.4) and $|U_t|$ is defined as $(|u_{1,t}|, \ldots, |u_{r+s,t}|)^T$. Because U_t is first-order stationary and F_t measurable, by (A.7) we have

$$\begin{aligned} E|u_{k,t}| &\le \boldsymbol{\eta}_k^T E\left(\prod_{i=0}^n A_{t-i}\right)E(|U_{t-n-1}|) \\ &= \boldsymbol{\eta}_k^T\mathbf{A}^n\mathbf{c} \to 0 \quad \text{as} \quad n\to\infty, \end{aligned} \tag{A.8}$$

where $\mathbf{c} = E|U_t|$ is a constant vector and $\mathbf{A}$ is defined by (A.5). So $u_{i,t} = 0$ a.s.; that is, $\tilde{\varepsilon}_t^2 = \tilde{\varepsilon}_t'^2$ a.s. Thus $h_t = h_t'$ a.s., and hence $\varepsilon_t = \varepsilon_t' = Z_t\sqrt{h_t}$ a.s. when Z_t's are given; that is, ε_t satisfying (2) is unique. Similarly, we can show that $\tilde{Y}_t = Y_t$ a.s. Therefore, $\{\varepsilon_t, Y_t\}$ is the unique F_t-measurable second-order stationary solution of the model (1)–(2). By (4)–(5), $\{\varepsilon_t, Y_t\}$ are measurable functions of the iid random variables $\{Z_t\}$ and hence are strictly stationary and ergodic.

Proof of Theorem 2.2

a. By the lemma in Section 2.1, we know that $E(\varepsilon_t^4) \le \infty$. Note that

$$\begin{aligned} EY_t^4 &= \sum_{k_1,k_2=0}^\infty a_{k_1}^2 a_{k_2}^2 E(\varepsilon_{t-k_1}^2\varepsilon_{t-k_2}^2) \\ &\le \sum_{k_1,k_2=0}^\infty a_{k_1}^2 a_{k_2}^2\sqrt{E\varepsilon_{t-k_1}^4 E\varepsilon_{t-k_2}^4} \\ &= \left(\sum_{k=0}^\infty a_k^2\right)^2 E\varepsilon_t^4, \end{aligned}$$

where $a_k = (k+d-1)!/k!(d-1)!$. Because $\sum a_k^2 < \infty$, we know that the right side is finite.

b. Again by the lemma in Section 2.1, $E(\varepsilon_t^8) < \infty$. Similar to (1),

$$\begin{aligned} EY_t^8 &= \sum_{k_1,k_2,k_3,k_4=0}^\infty a_{k_1}^2 a_{k_2}^2 a_{k_3}^2 a_{k_4}^2 E(\varepsilon_{t-k_1}^2\varepsilon_{t-k_2}^2\varepsilon_{t-k_3}^2\varepsilon_{t-k_4}^2) \\ &\le \sum_{k_1,k_2,k_3,k_4=0}^\infty a_{k_1}^2 a_{k_2}^2 a_{k_3}^2 a_{k_4}^2 \\ &\quad \times (E\varepsilon_{t-k_1}^8 E\varepsilon_{t-k_2}^8 E\varepsilon_{t-k_3}^8 E\varepsilon_{t-k_4}^8)^{1/4} \\ &= \left(\sum_{k=0}^\infty a_k^2\right)^4 E\varepsilon_t^8 < \infty, \end{aligned}$$

where a_k is defined as in part a.

Proof of Theorem 2.3

a. Let $\varphi_1(z) = \phi^{-1}(z)\theta(z)(1-z)^{-d}$. Because the power series expansions of $\phi^{-1}(z)$ and $(1-z)^{-d}$ under the assumptions converge for $|z| \le 1$, Y_t exists with representation (9). By Theorem 2.1, $\{\varepsilon_t\}$ is second-order stationary. Similar to the proof of Theorem 2.2, we can show that $\{Y_t\}$ is also second-order stationary. By representation (4), ε_t is a measurable function of iid random variable Z_t's and hence so is $\{Y_t\}$. Therefore, $\{Y_t\}$ is strictly stationary and ergodic.
b. Let $\varphi_2(z) = \phi(z)\theta^{-1}(z)(1-z)^d$. Similar to φ_1, the power series expansion of φ_2 converges as $|z| \le 1$, and we know that (10) holds.
c. Let $U_t = (1-B)^{-d}\varepsilon_t$; then $Y_t - \mu = \phi^{-1}(B)\theta(B)U_t$. By Theorem 2.2a, $E(\varepsilon_t^4) < \infty$. Similar to the proof of Theorem 2.2a, we can show that $E(U_t^4) < \infty$ and then that $E(Y_t^4) < \infty$. Similarly, we have $E(Y_t^8) < \infty$ under the given assumptions.

Proof of Theorem 3.1

First, note that

$$E\left(\frac{\partial\varepsilon_t}{\partial d}\right)^2 = \sum_{k=1}^\infty \frac{1}{k^2}E\varepsilon_t^2 = \frac{1}{6}\pi^2 E\varepsilon_t^2 < \infty,$$

$$\frac{\partial^2\varepsilon_t}{\partial d^2} = \ln^2(1-B)\varepsilon_t = \left(-\sum_{k=1}^\infty \frac{1}{k}B^k\right)^2\varepsilon_t,$$

and $E[(\partial^2\varepsilon_t)/(\partial d^2)|F_{t-1}] = 0$. Under the given assumptions, we can show that $E[(\partial^2 l_t)/(\partial\gamma\delta\gamma^T)]$, $E[(\partial^2 l_t)/(\partial\gamma\partial\delta^T)]$, and $E[(\partial^2 l_t)/(\partial\delta\partial\delta^T)]$ are finite. Further, because ε_t has a symmetric distribution, it is not difficult to prove that $E[(\partial^2 l_t)/(\partial\gamma\partial\delta^T)] = 0$. By the strict stationarity and ergodicity of $\{Y_t\}$ and $\{\varepsilon_t\}$ and the ergodic theorem, (18) holds. Similar to the proof of Weiss's lemmas 3.1–3.3 (Weiss 1986), we can show that $\boldsymbol{\Omega}_\gamma$ and $\boldsymbol{\Omega}_\delta$ are positive definite matrices.

Proof of Theorem 3.2.(a)

Following Weiss (1986), we need to check the conditions provided by Basawa, Feign, and Heyde (1976): (1) $n^{-1}\sum\partial l_t(\lambda_0)/\partial\lambda \xrightarrow{p} 0$; (2) there exists a nonrandom matrix $\mathbf{M}(\lambda_0) > 0$ such that for all $\varepsilon > 0$, $P\{-n^{-1}\sum\partial^2 l_t(\lambda_0)/\partial\lambda\partial\lambda^T \ge \mathbf{M}(\lambda_0)\} > 1-\varepsilon$ for all $n > n_1(\varepsilon)$; and (3) there exists a constant $M < \infty$ such that $E|(\partial^3 l_t(\lambda))(\partial\lambda_i\partial\lambda_j\partial\lambda_k)| < M$ for all $\lambda \in \Theta$, where λ_i is the ith component of λ.

These three conditions ensure that a root of the equation $\partial l_t(\lambda)/\partial\lambda$ exists such that $\hat{\lambda}_n \xrightarrow{p} \lambda_0$. The following show that conditions (1)–(3) are satisfied:

a. From (12)–(15), we know that $E[\partial l_t/\partial\lambda]_{\lambda=\lambda_0} = 0$. By the ergodic theorem, condition (1) is satisfied.

b. By Theorem 3.1, the matrix $\boldsymbol{\Omega}_0$ is positive definite, and hence for any constant vector $\mathbf{c} \neq 0$, $n^{-1}\sum \mathbf{c}^T(\partial^2 l_t(\lambda_0))\lambda(\partial\lambda\partial\lambda^T)c \xrightarrow{\text{a.s.}} -\mathbf{c}^T\boldsymbol{\Omega}_0\mathbf{c} < 0$. For any given $\mathbf{c}$, let $0 < \Delta(\mathbf{c}) < \mathbf{c}^T\boldsymbol{\Omega}_0\mathbf{c}/2$. Then for all $\varepsilon > 0$, there exists $n_1 = n_1(\varepsilon)$ such that

$$P\left\{\left|n^{-1}\sum \mathbf{c}^T\,\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T}\,\mathbf{c} + \mathbf{c}^T\boldsymbol{\Omega}_0\mathbf{c}\right| < \Delta\right\} > 1-\varepsilon$$

for all $n > n_1$; that is,

$$P\left\{-n^{-1}\sum \mathbf{c}^T\,\frac{\partial^2 l_t(\lambda_0)}{\partial\lambda\partial\lambda^T}\,\mathbf{c} > \mathbf{c}^T\boldsymbol{\Omega}_0\mathbf{c}/2\right\} > 1-\varepsilon$$

for all $n > n_1$. So condition (3) is satisfied.

c. Note that

$$\frac{\partial^2\varepsilon_t}{\partial d^3} = \ln^3(1-B)\varepsilon_t = \left(-\sum_{k=1}^{\infty}\frac{1}{k}\,B^k\right)^3\varepsilon_t$$

and

$$E\left(\frac{\partial^3\varepsilon_t}{\partial d^3}\right)^2 = E\left(\sum_{k_1,k_2,k_3=1}^{\infty}\frac{1}{k_1k_2k_3}\,\varepsilon_{t-k_1-k_2-k_3}\right)^2$$
$$= \sum_{k_1,k_2,k_3=1}^{\infty}\frac{1}{k_1^2k_2^2k_3^2}\,E\varepsilon_{t-k_1-k_2-k_3}^2$$
$$= \left(\frac{\pi}{6}\right)^3 E\varepsilon_t^2 < \infty. \qquad \text{(A.9)}$$

By differentiating $\partial^2 l_t/\partial\lambda\partial\lambda^T$ and using (A.9), we can conclude that condition (3) holds.

Proof of Theorem 3.2.b

We again need to check the conditions given by Basawa et al. (1976); that is, (1) $n^{-1/2}\sum[\partial l_t(\lambda_0)/\partial\lambda] \xrightarrow{D} N(O,\mathbf{B}_0)$ for a nonrandom $\mathbf{B}_0 > 0$; (2) $n^{-1}\sum[\partial^2 l_t(\lambda_0)/\partial\lambda\partial\lambda^T] \xrightarrow{p} -\mathbf{A}_0$ for a nonrandom $\mathbf{A}_0 > 0$; and (3) the condition (3) in the proof of Theorem 3.2a.

First, by direct calculation it is easy to know that $E[(\partial l_t/\partial\lambda)(\partial l_t/\partial\lambda^T)] = \text{diag}(\boldsymbol{\Omega}_\gamma,\boldsymbol{\Omega}_\delta)$. By the ergodic theorem,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial l_t}{\partial\lambda}\frac{\partial l_t}{\partial\lambda^T}\right)\Big|_{\lambda=\lambda_0} \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}_0.$$

Let $S_n = \sum_{t=1}^{n}\boldsymbol{\eta}_0^T(\partial l_t/\partial\lambda)|_{\lambda=\lambda_0}$, where $\boldsymbol{\eta}_0$ is an arbitrary constant vector with $\boldsymbol{\eta}_0^T\boldsymbol{\eta}_0 \neq 0$. Then it is easy to show that S_n is a martingale with $(1/n)ES_n^2 = \boldsymbol{\eta}_0^T\boldsymbol{\Omega}_0\boldsymbol{\eta}_0 > 0$. From the strict stationarity and ergodicity of $\{Y_t\}$ and $\{\varepsilon_t\}$, $[(1/n)ES_n^2]^{-1}[(1/n)E(S_n^2|F_{n-1})] \xrightarrow{\text{a.s.}} 1$. Using the central limit theorem of Stout (1974), we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t}{\partial\lambda}\Big|_{\lambda=\lambda_0} \xrightarrow{D} N(O,\boldsymbol{\Omega}_0).$$

That is, condition (1) holds. Next by Theorem 3.1, condition (2) holds. Finally by the proof of Theorem 3.2.a, condition (3) holds. Thus we complete the proof.

[Received November 1995. Revised July 1996.]

REFERENCES

Agiakloglou, C., and Newbold, P. (1993), "Lagrange Multiplier Tests for Fractional Difference," *Journal of Time Series Analysis*, 15, 253–262.

Baillie, R. T. (1995), "Long-Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

Baillie, R. T., Chung, C. F., and Tiles, M. A. (1995), "Analyzing Inflation by the Fractionally Integrated ARFIMA–GARCH Model," *Journal of Applied Econometrics*, 11, 23–40.

Basawa, I. V., Feign, P. D., and Heyde, C. C. (1976), "Asymptotic Properties of Maximum Likelihood Estimators for Stochastic Processes," *Sankhya*, Ser. A, 38, 259–270.

Beran, J. (1995), "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short- and Long-Memory Autoregressive Integrated Moving Average Models," *Journal of the Royal Statistical Society*, Ser. B, 57, 659–672.

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

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH Modeling in Finance," *Journal of Econometrics*, 52, 5–59.

Bollerslev, T., Engle, R. F., and Woodridge, J. M. (1988), "A Capital Asset Pricing Model With Time-Varying Covariance," *Journal of Political Economy*, 96, 116–131.

Box, G. E. P., and Jenkin, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.

Chung, C. F., and Baillie, R. T. (1994), "Small-Sample Bias in Conditional Sum-of-squares Estimators of Fractionally Integrated ARMA Models," in *New Developments in Time Series Econometrics*, ed. J.-M. Dufour, New York: Springer-Verlag.

Diebold, F., and Rudebusch, G. D. (1989), "Long Memory and Persistence in Aggregate Output," *Journal of Monetary Economics*, 24, 189–209.

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

Fox, R., and Taqqu, M. S. (1986), "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *The Annals of Statistics*, 14, 517–532.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis*, 4, 229–238.

Goldfeld, S. M., Quandt, R. E., and Trotter, H. F. (1966), "Maximization by Quadratic Hill-Climbing," *Econometrica*, 34, 541–551.

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

Hassler, U., and Wolters, J. (1994), "On the Power of Unit Root Tests Against Fractional Alternatives," *Economics Letters*, 45, 1–5.

Hosking, J. R. M. (1981), "Fractional Differencing," *Biometrika*, 68, 165–176.

——— (1984), "Modeling Persistence in Hydrological Time Series Using Fractional Differencing," *Water Resources Research*, 20, 1898–1908.

Li, W. K. (1981), "Topics in Time Series Modeling," unpublished Ph.D. thesis, University of Western Ontario.

——— (1992), "On the Asymptotic Standard Errors of Residual Autocorrelations in Nonlinear Time Series Modeling," *Biometrika*, 79, 435–437.

Li, W. K., and Mak, T. K. (1994), "On the Squared Residual Autocorrelations in Nonlinear Time Series With Conditional Heteroskedasticity," *Journal of Time Series Analysis*, 15, 627–636.

Li, W. K., and McLeod, A. I. (1986), "Fractional Time Series Modeling," *Biometrika*, 73, 217–221.

Li, C. W., and Li, W. K. (1996), "On a Double-Threshold Autoregressive With Heteroskedasticity Time Series Model," *Journal of Applied Econometrics*, 11, 253–274.

Ling, S. (1995), "On the Probabilistic Properties of a Double-Threshold ARMA Conditional Heteroskedasticity Model," Technical report, University of Hong Kong, Dept. of Statistics.

Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 65, 297–303.

McLeod, A. I., and Hipel, K. W. (1978), "Preservation of the Rescaled Adjusted Range, I. A Reassessment of the Hurst Phenomenon," *Water Resources Research*, 14, 491–508.

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

Pantula, S. G. (1989), "Estimation of Autoregressive Models With ARCH Errors," *Sankhya*, Ser. B, 50, 119–138.

Sowell, F. (1992), "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models," *Journal of Econometrics*, 53, 165–188.

Stout, W. F. (1974), *Almost Sure Convergence*, New York: Academic Press.

Thisted, R. A. (1991), *Elements of Statistical Computing: Numerical Computation*, London: Chapman and Hall.

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

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

Yajima, Y. (1988), "On Estimation of Regression Models With Long-Memory Stationary Errors," *The Annals of Statistics*, 18, 791–807.