

## GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY

Tim BOLLERSLEV\*

*University of California at San Diego, La Jolla, CA 92093, USA  
Institute of Economics, University of Aarhus, Denmark*

Received May 1985, final version received February 1986

A natural generalization of the ARCH (Autoregressive Conditional Heteroskedastic) process introduced in Engle (1982) to allow for past conditional variances in the current conditional variance equation is proposed. Stationarity conditions and autocorrelation structure for this new class of parametric models are derived. Maximum likelihood estimation and testing are also considered. Finally an empirical example relating to the uncertainty of the inflation rate is presented.

### 1. Introduction

While conventional time series and econometric models operate under an assumption of constant variance, the ARCH (Autoregressive Conditional Heteroskedastic) process introduced in Engle (1982) allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant.

This type of model behavior has already proven useful in modelling several different economic phenomena. In Engle (1982), Engle (1983) and Engle and Kraft (1983), models for the inflation rate are constructed recognizing that the uncertainty of inflation tends to change over time. In Coulson and Robins (1985) the estimated inflation volatility is related to some key macroeconomic variables. Models for the term structure using an estimate of the conditional variance as a proxy for the risk premium are given in Engle, Lilien and Robins (1985). The same idea is applied to the foreign exchange market in Domowitz and Hakkio (1985). In Weiss (1984) ARMA models with ARCH errors are found to be successful in modelling thirteen different U.S. macroeconomic time series. Common to most of the above applications however, is the introduction of a rather arbitrary linear declining lag structure in the condi-

\*I am grateful to David Hendry and Rob Engle for introducing me to this new idea, and to Rob Engle for many helpful discussions. I would also like to thank Sastry Pantula for suggesting the alternative parameterization, two anonymous referees for useful comments, and Kirsten Stentoft for typing the manuscript. The usual disclaimer applies.

tional variance equation to take account of the long memory typically found in empirical work, since estimating a totally free lag distribution often will lead to violation of the non-negativity constraints.

In this paper a new, more general class of processes, GARCH (Generalized Autoregressive Conditional Heteroskedastic), is introduced, allowing for a much more flexible lag structure. The extension of the ARCH process to the GARCH process bears much resemblance to the extension of the standard time series AR process to the general ARMA process and, as is argued below, permits a more parsimonious description in many situations.

The paper proceeds as follows. In section 2 the new class of processes is formally presented and conditions for their wide-sense stationarity are derived. The simple GARCH(1, 1) process is considered in some detail in section 3. It is well established, that the autocorrelation and partial autocorrelation functions are useful tools in identifying and checking time series behavior of the ARMA form in the conditional mean. Similarly the autocorrelations and partial autocorrelations for the squared process may prove helpful in identifying and checking GARCH behavior in the conditional variance equation. This is the theme in section 4. In section 5 maximum likelihood estimation of the linear regression model with GARCH errors is briefly discussed, and it is seen that the asymptotic independence between the estimates of the mean and the variance parameters carries over from the ARCH regression model. Some test results are presented in section 6. As in the ARMA analogue, cf. Godfrey (1978), a general test for the presence of GARCH is not feasible. Section 7 contains an empirical example explaining the uncertainty of the inflation rate. It is argued that a simple GARCH model provides a marginally better fit and a more plausible learning mechanism than the ARCH model with an eighth-order linear declining lag structure as in Engle and Kraft (1983).

## 2. The GARCH( $p, q$ ) process

The ARCH process introduced by Engle (1982) explicitly recognizes the difference between the unconditional and the conditional variance allowing the latter to change over time as a function of past errors. The statistical properties of this new parametric class of models has been studied further in Weiss (1982) and in a recent paper by Milhøj (1984).

In empirical applications of the ARCH model a relatively long lag in the conditional variance equation is often called for, and to avoid problems with negative variance parameter estimates a fixed lag structure is typically imposed, cf. Engle (1982), Engle (1983) and Engle and Kraft (1983). In this light it seems of immediate practical interest to extend the ARCH class of models to allow for both a longer memory and a more flexible lag structure.

Let  $\varepsilon_t$  denote a real-valued discrete-time stochastic process, and  $\psi_t$  the information set ( $\sigma$ -field) of all information through time  $t$ . The GARCH( $p, q$ )

process (Generalized Autoregressive Conditional Heteroskedasticity) is then given by<sup>1</sup>

$$\epsilon_t | \psi_{t-1} \sim N(0, h_t), \tag{1}$$

$$\begin{aligned}
 h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} \\
 &= \alpha_0 + A(L)\epsilon_t^2 + B(L)h_t,
 \end{aligned}
 \tag{2}$$

where

$$\begin{aligned}
 p &\geq 0, & q &> 0 \\
 \alpha_0 &> 0, & \alpha_i &\geq 0, \quad i = 1, \dots, q, \\
 \beta_i &\geq 0, & i &= 1, \dots, p.
 \end{aligned}$$

For  $p = 0$  the process reduces to the ARCH( $q$ ) process, and for  $p = q = 0$   $\epsilon_t$  is simply white noise. In the ARCH( $q$ ) process the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH( $p, q$ ) process allows lagged conditional variances to enter as well. This corresponds to some sort of adaptive learning mechanism.

The GARCH( $p, q$ ) regression model is obtained by letting the  $\epsilon_t$ 's be innovations in a linear regression,

$$\epsilon_t = y_t - x_t' b, \tag{3}$$

where  $y_t$  is the dependent variable,  $x_t$  a vector of explanatory variables, and  $b$  a vector of unknown parameters. This model is studied in some detail in section 5.

If all the roots of  $1 - B(z) = 0$  lie outside the unit circle, (2) can be rewritten as a distributed lag of past  $\epsilon_t^2$ 's,

$$\begin{aligned}
 h_t &= \alpha_0 (1 - B(1))^{-1} + A(L)(1 - B(L))^{-1} \epsilon_t^2 \\
 &= \alpha_0 \left( 1 - \sum_{i=1}^p \beta_i \right)^{-1} + \sum_{i=1}^{\infty} \delta_i \epsilon_{t-i}^2,
 \end{aligned}
 \tag{4}$$

which together with (1) may be seen as an infinite-dimensional ARCH( $\infty$ ) process. The  $\delta_i$ 's are found from the power series expansion of  $D(L) =$

<sup>1</sup>We follow Engle (1982) in assuming the conditional distribution to be normal, but of course other distributions could be applied as well. Instead of  $\epsilon_{t-i}^2$  in eq. (2) the absolute value of  $\epsilon_{t-i}$  may be more appropriate in some applications; cf. McCulloch (1983).

$$A(L)(1 - B(L))^{-1},$$

$$\begin{aligned} \delta_i &= \alpha_i + \sum_{j=1}^n \beta_j \delta_{i-j}, & i = 1, \dots, q, \\ &= \sum_{j=1}^n \beta_j \delta_{i-j}, & i = q + 1, \dots, \end{aligned} \tag{5}$$

where  $n = \min\{p, i - 1\}$ . It follows, that if  $B(1) < 1$ ,  $\delta_i$  will be decreasing for  $i$  greater than  $m = \max\{p, q\}$ . Thus if  $D(1) < 1$ , the GARCH( $p, q$ ) process can be approximated to any degree of accuracy by a stationary ARCH( $Q$ ) for a sufficiently large value of  $Q$ . But as in the ARMA analogue, the GARCH process might possibly be justified through a Wald's decomposition type of arguments as a more parsimonious description.

From the theory on finite-dimensional ARCH( $q$ ) processes it is to be expected that  $D(1) < 1$ , or equivalently  $A(1) + B(1) < 1$ , suffices for wide-sense stationarity; cf. Milhøj (1984). This is indeed the case.

*Theorem 1.* The GARCH( $p, q$ ) process as defined in (1) and (2) is wide-sense stationary with  $E(\varepsilon_t) = 0$ ,  $\text{var}(\varepsilon_t) = \alpha_0(1 - A(1) - B(1))^{-1}$  and  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$  for  $t \neq s$  if and only if  $A(1) + B(1) < 1$ .

*Proof.* See appendix.

As pointed out by Sastry Pantula and an anonymous referee, an equivalent representation of the GARCH( $p, q$ ) process is given by

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^p \beta_j v_{t-j} + v_t, \tag{6}$$

and

$$v_t = \varepsilon_t^2 - h_t = (\eta_t^2 - 1)h_t, \tag{7}$$

where

$$\eta_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1).$$

Note, by definition  $v_t$  is serially uncorrelated with mean zero. Therefore, the GARCH( $p, q$ ) process can be interpreted as an autoregressive moving average process in  $\varepsilon_t^2$  of orders  $m = \max\{p, q\}$  and  $p$ , respectively. Although a parameterization along the lines of (6) might be more meaningful from a theoretical time series' point of view, (1) and (2) are easier to work with in practice.

### 3. The GARCH(1, 1) process

The simplest but often very useful GARCH process is of course the GARCH(1, 1) process given by (1) and

$$h_t = \alpha_0 \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \quad \alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0. \tag{8}$$

From Theorem 1,  $\alpha_1 + \beta_1 < 1$  suffices for wide-sense stationarity, and in general we have:<sup>2</sup>

*Theorem 2. For the GARCH(1, 1) process given by (1) and (6) a necessary and sufficient condition for existence of the 2mth moment is*

$$\mu(\alpha_1, \beta_1, m) = \sum_{j=0}^m \binom{m}{j} a_j \alpha_1^j \beta_1^{m-j} < 1, \tag{9}$$

where

$$a_0 = 1, \quad a_j = \prod_{i=1}^j (2i - 1), \quad j = 1, \dots. \tag{10}$$

The 2mth moment can be expressed by the recursive formula

$$\begin{aligned} E(\varepsilon_t^{2m}) &= a_m \left[ \sum_{n=0}^{m-1} a_n^{-1} E(\varepsilon_t^{2n}) \alpha_0^{m-n} \binom{m}{m-n} \mu(\alpha_1, \beta_1, n) \right] \\ &\times [1 - \mu(\alpha_1, \beta_1, m)]^{-1}. \end{aligned} \tag{11}$$

*Proof.* See appendix.

The conditions for existence of the first twelve moments are illustrated in fig. 1.

It follows by symmetry that if the 2mth moment exists,  $E(\varepsilon_t^{2m-1}) = 0$ .

For  $\beta_1 = 0$ , (9) reduces to the well-known condition for the ARCH(1) process,  $a_m \alpha_1^m < 1$ ; cf. Engle (1982). Thus if  $\alpha_1 > (a_m)^{-1/m}$  in the ARCH(1) process, the 2mth moment does not exist, whereas even if  $\sum_{i=1}^{\infty} \delta_i = \alpha_1 (1 - \beta_1)^{-1} > (a_m)^{-1/m}$  in the GARCH(1, 1) process, the 2mth moment might very well exist because of the longer memory in this process.

In the GARCH(1, 1) process the mean lag in the conditional variance equation is given by

$$\zeta = \sum_{i=1}^{\infty} i \delta_i / \sum_{i=1}^{\infty} \delta_i = (1 - \beta_1)^{-1},$$

<sup>2</sup>As in Engle (1982) we assume for simplicity that the process starts indefinitely far in the past with 2m finite initial moments.

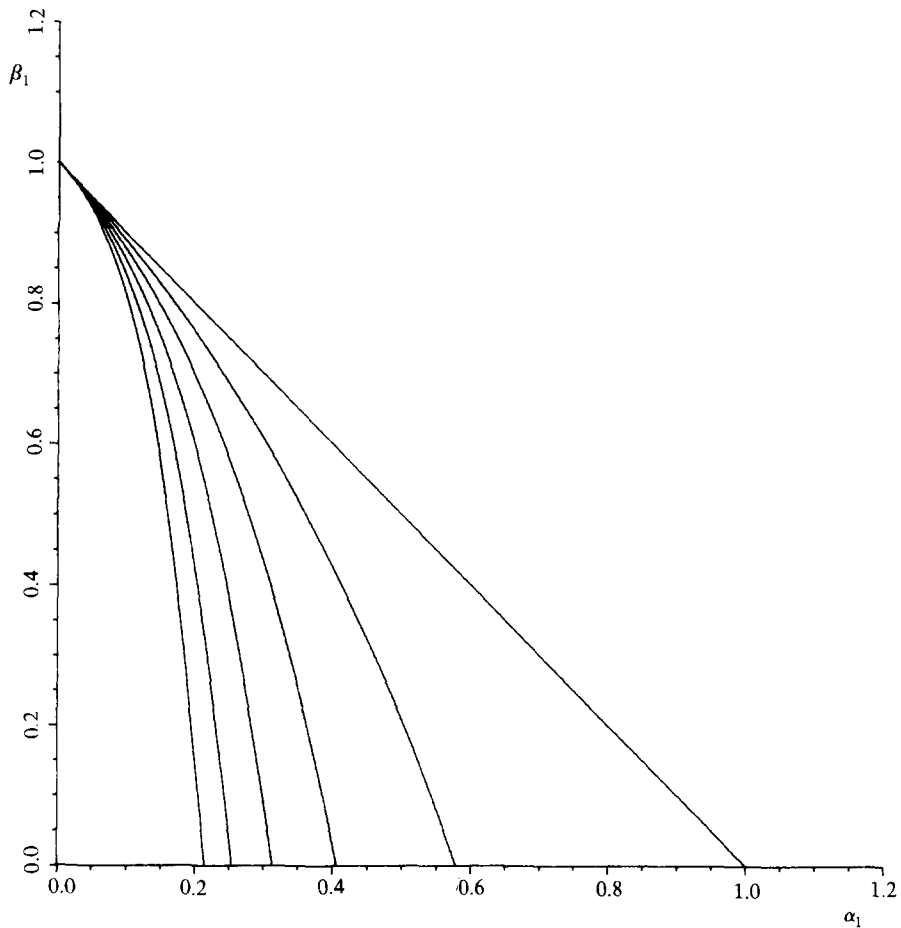


Fig. 1. Moment conditions for GARCH(1,1).

and the median lag is found to be

$$\nu = -\log 2 / \log \beta_1,$$

where  $\sum_{i=1}^{\nu} \delta_i / \sum_{i=1}^{\infty} \delta_i = \frac{1}{2}$  and the  $\delta_i$ 's are defined in (5); cf. Harvey (1982).

If  $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ , the fourth-order moment exists and by Theorem 2

$$E(\varepsilon_t^2) = \alpha_0(1 - \alpha_1 - \beta_1)^{-1},$$

and

$$E(\epsilon_t^4) = 3\alpha_0^2(1 + \alpha_1 + \beta_1)[(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)]^{-1}.$$

The coefficient of kurtosis is therefore

$$\begin{aligned} \kappa &= \left( E(\epsilon_t^4) - 3E(\epsilon_t^2)^2 \right) E(\epsilon_t^2)^{-2} \\ &= 6\alpha_1^2(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)^{-1}, \end{aligned}$$

which is greater than zero by assumption. Hence the GARCH(1,1) process is leptokurtic (heavily tailed), a property the process shares with the ARCH( $q$ ) process; cf. Milhøj (1984).

#### 4. Autocorrelation and partial autocorrelation structure

The use of autocorrelation and partial autocorrelation functions to identify and check time series behaviour of the ARMA form in the conditional mean is well established; cf. Box and Jenkins (1976). In this section, the autocorrelation and partial autocorrelation functions for the squared process are shown to be useful in identifying and checking time series behaviour in the conditional variance equation of the GARCH form. The idea of using the squared process to check for model adequacy is not new; cf. Granger and Anderson (1978) where it is found that some of the series modelled in Box and Jenkins (1976) exhibit autocorrelated squared residuals even though the residuals themselves do not seem to be correlated over time.

Consider the general GARCH( $p, q$ ) process as specified in (1) and (2), and let us assume the process has finite fourth-order moment.<sup>3</sup> Let the covariance function for  $\epsilon_t^2$  be denoted

$$\gamma_n = \gamma_{-n} = \text{cov}(\epsilon_t^2, \epsilon_{t-n}^2). \tag{12}$$

<sup>3</sup>The general conditions for the existence of finite fourth-order moment are unknown. However, given a specific order of the model the conditions may be derived following the same line of arguments as lead to Theorem 2 for the GARCH(1,1) process. For instance the necessary and sufficient condition for the GARCH(1,2) process is found to be

$$\alpha_2 + 3\alpha_1^2 + 3\alpha_2^2 + \beta_1^2 + 2\alpha_1\beta_1 - 3\alpha_2^3 + 3\alpha_1^2\alpha_2 + 6\alpha_1\alpha_2\beta_1 + \alpha_2\beta_1^2 < 1,$$

and for the GARCH(2,1) the condition is

$$\beta_2 + 3\alpha_1^2 + \beta_1^2 + \beta_2^2 + 2\alpha_1\beta_1 - \beta_2^3 - \alpha_1^2\beta_2 + 2\alpha_1\beta_1\beta_2 + \beta_1^2\beta_2 < 1.$$

In Milhøj (1984) the condition for the ARCH( $q$ ) process is derived and expressed in terms of the inverse of a  $q \times q$  matrix,  $3\varphi'(I - \phi)^{-1}\varphi < 1$ , where  $\varphi' = (\alpha_1, \dots, \alpha_q)$ ,  $\phi_{ij} = \varphi_{i+j} + \varphi_{i-j}$ ,  $i, j = 1, \dots, q$ , and  $\varphi_k = 0$  for  $k \leq 0$  and  $k > q$ .

It follows then immediately from (6) and (7) that

$$\begin{aligned}\gamma_n &= \sum_{i=1}^q \alpha_i \gamma_{n-i} + \sum_{i=1}^p \beta_i \gamma_{n-i} \\ &= \sum_{i=1}^m \varphi_i \gamma_{n-i}, \quad n \geq p+1,\end{aligned}\tag{13}$$

where  $m = \max\{p, q\}$ ,

$$\varphi_i = \alpha_i + \beta_i, \quad i = 1, \dots, q,$$

$\alpha_i \equiv 0$  for  $i > q$  and  $\beta_i \equiv 0$  for  $i > p$ . From (13) we get the following analogue to the Yule-Walker equations:

$$\rho_n = \gamma_n \gamma_0^{-1} = \sum_{i=1}^m \varphi_i \rho_{n-i}, \quad n \geq p+1.\tag{14}$$

Thus, the first  $p$  autocorrelations for  $\varepsilon_t^2$  depend 'directly' on the parameters  $\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p$ , but given  $\rho_p, \dots, \rho_{p+1-m}$  the above difference equation uniquely determines the autocorrelations at higher lags. This is similar to the result for the autocorrelations for an ARMA( $m, p$ ) process; cf. Box and Jenkins (1976). Note also, that (14) depends on the parameters  $\alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p$  only through  $\varphi_1, \dots, \varphi_m$ .

Let  $\phi_{kk}$  denote the  $k$ th partial autocorrelation for  $\varepsilon_t^2$  found by solving the set of  $k$  equations in the  $k$  unknown  $\phi_{k1}, \dots, \phi_{kk}$ :

$$\rho_n = \sum_{i=1}^k \phi_{ki} \rho_{n-i}, \quad n = 1, \dots, k.\tag{15}$$

By (14)  $\phi_{kk}$  cuts off after lag  $q$  for an ARCH( $q$ ) process

$$\begin{aligned}\phi_{kk} &\neq 0, & k \leq q, \\ &= 0, & k > q.\end{aligned}\tag{16}$$

This is identical to the behaviour of the partial autocorrelation function for an AR( $q$ ) process. Also from (14) and well-known results in the time series literature, the partial autocorrelation function for  $\varepsilon_t^2$  for a GARCH( $p, q$ ) process is in general non-zero but dies out; see Granger and Newbold (1977).

In practice, of course, the  $\rho_n$ 's and  $\phi_{kk}$ 's will be unknown. However, the sample analogue, say  $\hat{\rho}_n$ , yields a consistent estimate for  $\rho_n$ , and  $\phi_{kk}$  is consistently estimated by the  $k$ th coefficient, say  $\hat{\phi}_{kk}$ , in a  $k$ th-order autore-



gression for  $\epsilon_t^2$ ; see Granger and Newbold (1977). These estimates together with their asymptotic variance under the null of no GARCH  $1/T$  [cf. Weiss (1984) and McLeod and Li (1983)] can be used in the preliminary identification stage, and are also useful for diagnostic checking.

### 5. Estimation of the GARCH regression model

In this section we consider maximum likelihood estimation of the GARCH regression model (1),(2),(3). Because the results are quite similar to those for the ARCH regression model, our discussion will be very schematic.

Let  $z'_t = (1, \epsilon_{t-1}^2, \dots, \epsilon_{t-q}^2, h_{t-1}, \dots, h_{t-p})$ ,  $\omega' = (\alpha_0, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$  and  $\theta \in \Theta$ , where  $\theta = (b', \omega')$  and  $\Theta$  is a compact subspace of a Euclidean space such that  $\epsilon_t$  possesses finite second moments. Denote the true parameters by  $\theta_0$ , where  $\theta_0 \in \text{int } \Theta$ . We may then rewrite the model as

$$\begin{aligned} \epsilon_t &= y_t - x'_t b, \\ \epsilon_t | \psi_{t-1} &\sim N(0, h_t), \\ h_t &= z'_t \omega. \end{aligned} \tag{17}$$

The log likelihood function for a sample of  $T$  observations is apart from some constant,<sup>4</sup>

$$\begin{aligned} L_T(\theta) &= T^{-1} \sum_{t=1}^T l_t(\theta), \\ l_t(\theta) &= -\frac{1}{2} \log h_t - \frac{1}{2} \epsilon_t^2 h_t^{-1}. \end{aligned} \tag{18}$$

Differentiating with respect to the variance parameter yields

$$\frac{\partial l_t}{\partial \omega} = \frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial \omega} \left( \frac{\epsilon_t^2}{h_t} - 1 \right), \tag{19}$$

$$\frac{\partial^2 l_t}{\partial \omega \partial \omega'} = \left( \frac{\epsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \omega'} \left[ \frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial \omega} \right] - \frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial \omega} \frac{\partial h_t}{\partial \omega'} \frac{\epsilon_t^2}{h_t}, \tag{20}$$

<sup>4</sup>For simplicity we are conditioning on the pre-sample values. This does of course not affect the asymptotic results; cf. Weiss (1982).

where

$$\frac{\partial h_t}{\partial \omega} = z_t + \sum_{i=1}^p \beta_i \frac{\partial h_{t-i}}{\partial \omega}. \quad (21)$$

The only difference from Engle (1982) is the inclusion of the recursive part in (21).<sup>5</sup> Note,  $B(1) < 1$  guarantees that (21) is stable. Since the conditional expectation of the first term in (20) is zero, the part of Fisher's information matrix corresponding to  $\omega$  is consistently estimated by the sample analogue of the last term in (20) which involves first derivatives only.

Differentiating with respect to the mean parameters yields

$$\frac{\partial l_t}{\partial b} = \varepsilon_t x_t h_t^{-1} + \frac{1}{2} h_t \frac{\partial h_t}{\partial b} \left( \frac{\varepsilon_t^2}{h_t} - 1 \right), \quad (22)$$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial b \partial b'} &= -h_t^{-1} x_t x_t' - \frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial b} \frac{\partial h_t}{\partial b'} \left( \frac{\varepsilon_t^2}{h_t} \right) \\ &\quad - 2 h_t^{-2} \varepsilon_t x_t \frac{\partial h_t}{\partial b} + \left( \frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial b'} \left[ \frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial b} \right], \end{aligned} \quad (23)$$

where

$$\frac{\partial h_t}{\partial b} = -2 \sum_{j=1}^q \alpha_j x_{t-j} \varepsilon_{t-j} + \sum_{j=1}^p \beta_j \frac{\partial h_{t-j}}{\partial b}. \quad (24)$$

Again the single difference from the ARCH( $q$ ) regression model is the inclusion of the recursive part in (24). A consistent estimate of the part of the information matrix corresponding to  $b$  is given by the sample analogue of the first two terms in (23) but with  $\varepsilon_t^2 h_t^{-1}$  in the second term replaced by its expected value of one. This estimate will also involve first derivatives only.

Finally, the elements in the off-diagonal block in the information matrix may be shown to be zero. Because of this asymptotic independence  $\omega$  can be estimated without loss of asymptotic efficiency based on a consistent estimate of  $b$ , and vice versa.

To obtain maximum likelihood estimates, and second-order efficiency, an iterative procedure is called for. For the ARCH( $q$ ) regression model the method of scoring could be expressed in terms of a simple auxiliary regression, but the recursive terms in (21) and (24) complicate this procedure. Instead the

<sup>5</sup>To start up the recursion we need pre-sample estimates for  $h_t$  and  $\varepsilon_t^2$ ,  $t \leq 0$ . A natural choice is the sample analogue  $T^{-1} \sum_{t=-1}^T \varepsilon_t^2$ .

Berndt, Hall, Hall and Hausman (1974) algorithm turns out to be convenient. Let  $\theta^{(i)}$  denote the parameter estimates after the  $i$ th iteration.  $\theta^{(i+1)}$  is then calculated from

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left( \sum_{t=1}^T \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \right)^{-1} \sum_{t=1}^T \frac{\partial l_t}{\partial \theta},$$

where  $\partial l_t / \partial \theta$  is evaluated at  $\theta^{(i)}$ , and  $\lambda_i$  is a variable step length chosen to maximize the likelihood function in the given direction. Note, the direction vector is easily calculated from a least squares regression of a  $T \times 1$  vector of ones on  $\partial l_t / \partial \theta$ . Also, the iterations for  $\omega^{(i)}$  and  $b^{(i)}$  may be carried out separately because of the block diagonality in the information matrix.

From Weiss (1982) it follows that the maximum likelihood estimate  $\hat{\theta}_T$  is strongly consistent for  $\theta_0$  and asymptotically normal with mean  $\theta_0$  and covariance matrix  $\mathcal{F}^{-1} = -E(\partial^2 l_t / \partial \theta \partial \theta')^{-1}$ . However,  $\mathcal{F} = F$ , where  $F = E((\partial l_t / \partial \theta)(\partial l_t / \partial \theta'))$ , and a consistent estimate of the asymptotic covariance matrix is therefore given by  $T^{-1}(\sum_{t=1}^T (\partial l_t / \partial \theta)(\partial l_t / \partial \theta'))^{-1}$  from the last BHHH iteration.

Replacing (1) with the weaker set of assumptions

$$\begin{aligned} E(\varepsilon_t | \psi_{t-1}) &= 0, \\ E(\varepsilon_t^2 h_t^{-1} | \psi_{t-1}) &= 1, \\ E(\varepsilon_t^4 h_t^{-2} | \psi_{t-1}) &\leq M < \infty, \end{aligned} \tag{25}$$

$\hat{\theta}_T$  is still strongly consistent for  $\theta_0$  and asymptotically normal with mean  $\theta_0$  but with covariance matrix  $\mathcal{F}^{-1} F \mathcal{F}^{-1}$ ; see Weiss (1982) and White (1982). Of course, if the true conditional distribution is normal,  $F = \mathcal{F}$  and therefore  $\mathcal{F}^{-1} F \mathcal{F}^{-1} = \mathcal{F}^{-1}$ .

### 6. Testing for GARCH

Because of the complication involved in estimating a GARCH process, it seems of interest to have a formal test for the presence of GARCH instead of just relying on the more informal tools developed in section 4.

Consider the GARCH( $p, q$ ) regression model (17). As in Engle and Kraft (1983) let us partition the conditional variance equation

$$h_t = z_t' \omega = z_{1t}' \omega_1 + z_{2t}' \omega_2. \tag{26}$$

The Lagrange multiplier test statistic for  $H_0: \omega_2 = 0$  is then given by<sup>6</sup>

$$\xi_{LM}^* = \frac{1}{2} f_0' Z_0 (Z_0' Z_0)^{-1} Z_0' f_0, \tag{27}$$

where

$$f_0 = (\epsilon_1^2 h_1^{-1} - 1, \dots, \epsilon_T^2 h_T^{-1} - 1)',$$

$$Z_0 = \left( h_1 \frac{\partial h_1}{\partial \omega}, \dots, h_T \frac{\partial h_T}{\partial \omega} \right)', \tag{28}$$

and both are evaluated under  $H_0$ . When  $H_0$  is true,  $\xi_{LM}^*$  is asymptotically chi-square with  $r$ , the number of elements in  $\omega_2$ , degrees of freedom. This test differs slightly from the standard results, Breusch and Pagan (1978), in that  $\partial h_t / \partial \omega$  does not simplify when the conditional variance equation contains lagged conditional variances; cf. eq. (21).

It is well known that by normality an asymptotically equivalent test statistic is

$$\xi_{LM} = T \cdot R^2,$$

where  $R^2$  is the squared multiple correlation coefficient between  $f_0$  and  $Z_0$ . From section 5 this corresponds to  $T \cdot R^2$  from the OLS regression in the first BHHH iteration for the general model starting at the maximum likelihood estimates under  $H_0$ .

The alternative as represented by  $z_2$ , needs some consideration. Straightforward calculations show that under the null of white noise,  $Z_0' Z_0$  is singular if both  $p > 0$  and  $q > 0$ , and therefore a general test for GARCH( $p, q$ ) is not feasible. In fact if the null is an ARCH( $q$ ) process,  $Z_0' Z_0$  is singular for GARCH( $r_1, q + r_2$ ) alternatives, where  $r_1 > 0$  and  $r_2 > 0$ . It is also interesting to note that for an ARCH( $q$ ) null, the LM test for GARCH( $r, q$ ) and ARCH( $q + r$ ) alternatives coincide. This is similar to the results in Godfrey (1978), where it is shown that the LM tests for AR( $p$ ) and MA( $q$ ) errors in a linear regression model coincide and that the test procedures break down when a full ARMA( $p, q$ ) model is considered. These test results are, of course, not peculiar to the LM test, but concern the Likelihood Ratio and the Wald tests as well. A formal proof of the above statements can be constructed along the same lines as in Godfrey (1978, 1981).

<sup>6</sup>Note, because of the block diagonality in the information matrix in the GARCH regression model, the same test applies in this more general context.

## 7. Empirical example

The uncertainty of inflation is an unobservable economic variable of major importance, and within the ARCH framework several different models have already been constructed; see Engle (1982), Engle (1983) and Engle and Kraft (1983). We will here concentrate on the model in Engle and Kraft (1983) where the rate of growth in the implicit GNP deflator in the U.S. is explained in terms of its own past.

Let  $\pi_t = 100 \cdot \ln(GD_t/GD_{t-1})$  where  $GD_t$  is the implicit price deflator for GNP.<sup>7</sup> Standard univariate time series methods lead to identification of the following model for  $\pi_t$ :

$$\begin{aligned} \pi_t &= \begin{matrix} 0.240 & + & 0.552\pi_{t-1} & + & 0.177\pi_{t-2} & + & 0.232\pi_{t-3} & - & 0.209\pi_{t-4} & + & \varepsilon_t, \\ (0.080) & & (0.083) & & (0.089) & & (0.090) & & (0.080) \end{matrix} \\ h_t &= \begin{matrix} 0.282 \\ (0.034) \end{matrix} . \end{aligned} \quad (29)$$

The model is estimated on quarterly data from 1948.2 to 1983.4, i.e., a total of 143 observations, using ordinary least squares, with OLS standard errors in parentheses.<sup>8</sup> The model is stationary, and none of the first ten autocorrelations or partial autocorrelations for  $\varepsilon_t$  are significant at the 5% level. However, looking at the autocorrelations for  $\varepsilon_t^2$  it turns out that the 1st, 3rd, 7th, 9th and 10th all exceed two asymptotic standard errors. Similar results hold for the partial autocorrelations for  $\varepsilon_t^2$ . The LM test for ARCH(1), ARCH(4) and ARCH(8) are also highly significant at any reasonable level.

This leads Engle and Kraft (1983) to suggest the following specification:

$$\begin{aligned} \pi_t &= \begin{matrix} 0.138 & + & 0.423\pi_{t-1} & + & 0.222\pi_{t-2} & + & 0.377\pi_{t-3} & - & 0.175\pi_{t-4} & + & \varepsilon_t, \\ (0.059) & & (0.081) & & (0.108) & & (0.078) & & (0.104) \end{matrix} \\ h_t &= \begin{matrix} 0.058 \\ (0.033) \end{matrix} + \begin{matrix} 0.802 \\ (0.265) \end{matrix} \sum_{i=1}^8 (9-i)/36\varepsilon_{t-i}^2. \end{aligned} \quad (30)$$

The estimates are maximum likelihood with heteroskedastic consistent standard errors in parentheses; see section 5. The choice of the eighth-order linear declining lag structure is rather ad hoc, but motivated by the long memory in the conditional variance equation. The order of the lag polynomial may be viewed as an additional parameter in the conditional variance equation. From

<sup>7</sup>The values of  $GD_t$  are taken from the Citibank Economic Database and U.S. Department of Commerce, *Survey of Current Business*, Vol. 64, no. 9, September 1984.

<sup>8</sup>Note, in Engle and Kraft (1983) the estimation period is 1948.2 to 1980.3 accounting for the small differences in the estimation results.

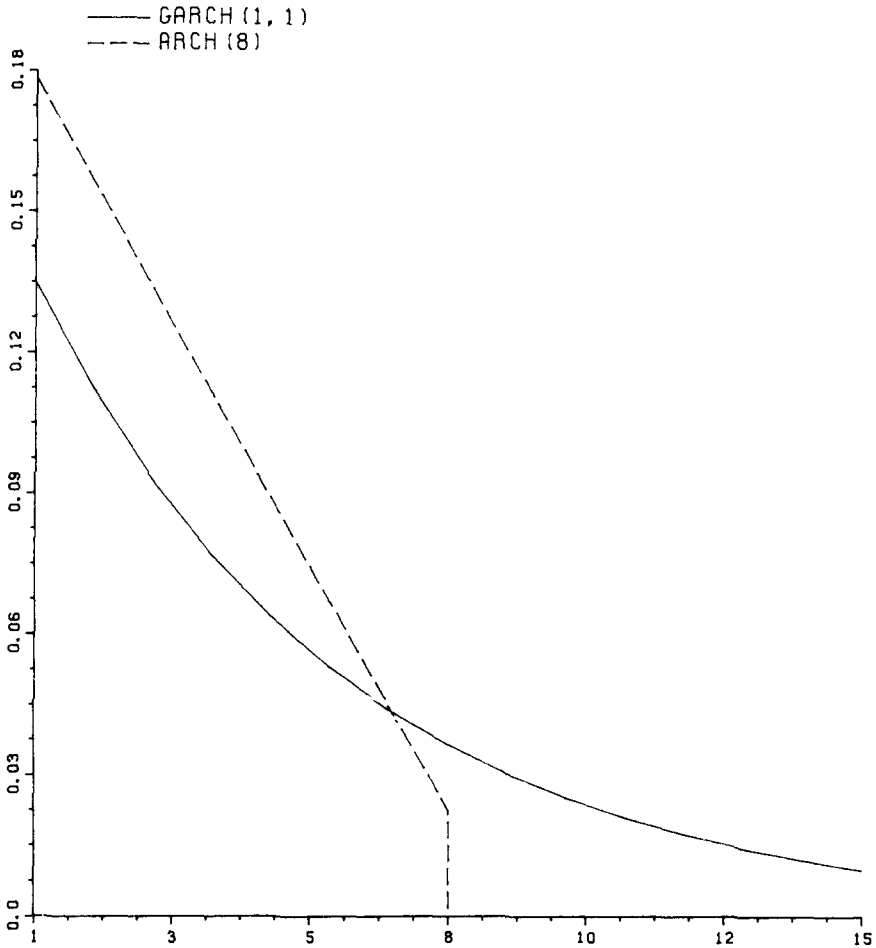


Fig. 2. Lag distribution.

Milhøj (1984) the condition for existence of the fourth-order moment of  $\varepsilon_t$  is just met.<sup>9</sup> None of the first ten autocorrelations or partial autocorrelations for  $\varepsilon_t h_t^{-1/2}$  or  $\varepsilon_t^2 h_t^{-1}$  exceed  $2/\sqrt{143}$ . The LM test statistics for the linear restriction takes the value 8.87, corresponding to the 0.74 fractile in the  $\chi_7^2$  distribution. However, the LM test statistic for the inclusion of  $h_{t-1}$  in the conditional variance equation is 4.57, which is significant at the 5% level.

<sup>9</sup>The condition as discussed in footnote 3 takes the value 0.989 for 0.802.

Let us therefore consider the alternative specification

$$\begin{aligned} \pi_t &= \frac{0.141}{(0.060)} + \frac{0.433}{(0.081)}\pi_{t-1} + \frac{0.229}{(0.110)}\pi_{t-2} + \frac{0.349}{(0.077)}\pi_{t-3} - \frac{0.162}{(0.104)}\pi_{t-4} + \varepsilon_t, \\ h_t &= \frac{0.007}{(0.006)} + \frac{0.135}{(0.070)}\varepsilon_{t-1}^2 + \frac{0.829}{(0.068)}h_{t-1}. \end{aligned} \tag{31}$$

From Theorem 2 the fourth-order moment of  $\varepsilon_t$  exists. Again none of the first ten autocorrelations or partial autocorrelations for  $\varepsilon_t h_t^{-1/2}$  or  $\varepsilon_t^2 h_t^{-1}$  exceed two asymptotic standard errors. The LM test statistic for the inclusion of the eighth-order linear declining lag structure is 2.33, corresponding to the 0.87 fractile in the  $\chi_1^2$  distribution. The LM test statistic for GARCH(1,2), or locally equivalent GARCH(2,1), equals 3.80 and therefore is not significant at the 5% level. Also the LM test statistics for inclusion of  $\varepsilon_{t-2}^2, \dots, \varepsilon_{t-5}^2$  takes the value 5.58 which is equal to the 0.77 fractile in the  $\chi_4^2$  distribution.

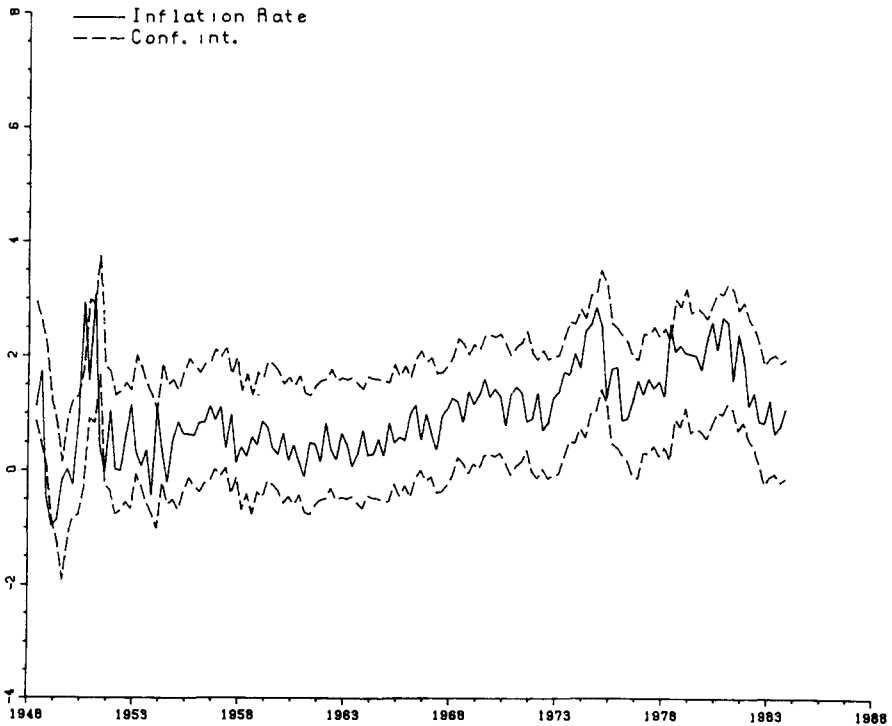


Fig. 3. 95% confidence intervals for OLS.

It is also interesting to note that the sample coefficient of kurtosis for  $\varepsilon_t h_t^{-1/2}$  from model (31) equals 3.81, which differs from the 'normal' value of 3.00 by slightly less than two asymptotic standard errors,  $2\sqrt{24/T} \approx 0.82$ . For models (29) and (30) the coefficient of kurtosis equals 6.90 and 4.07, respectively. The sample coefficient of skewness for  $\varepsilon_t h_t^{-1/2}$  from each of the three models,  $-0.13$ ,  $0.18$  and  $0.11$ , are all within one asymptotic standard error,  $\sqrt{6/T} \approx 0.20$ .

The mean and median lag in the conditional variance equation in (31) are estimated to be 5.848 and 3.696, respectively [cf. section 3], whereas in (30) the mean lag is forced to  $3\frac{1}{3}$  and the median lag to  $2\frac{1}{2}$ . Furthermore, the lag structure in the GARCH(1,1) model can be rationalized by some sort of adaptive learning mechanism. See also fig. 2, where the two different lag shapes are illustrated. In this light it seems that not only does the GARCH(1,1) model provide a slightly better fit than the ARCH(8) model in Engle and Kraft (1983), but it also exhibits a more reasonable lag structure.

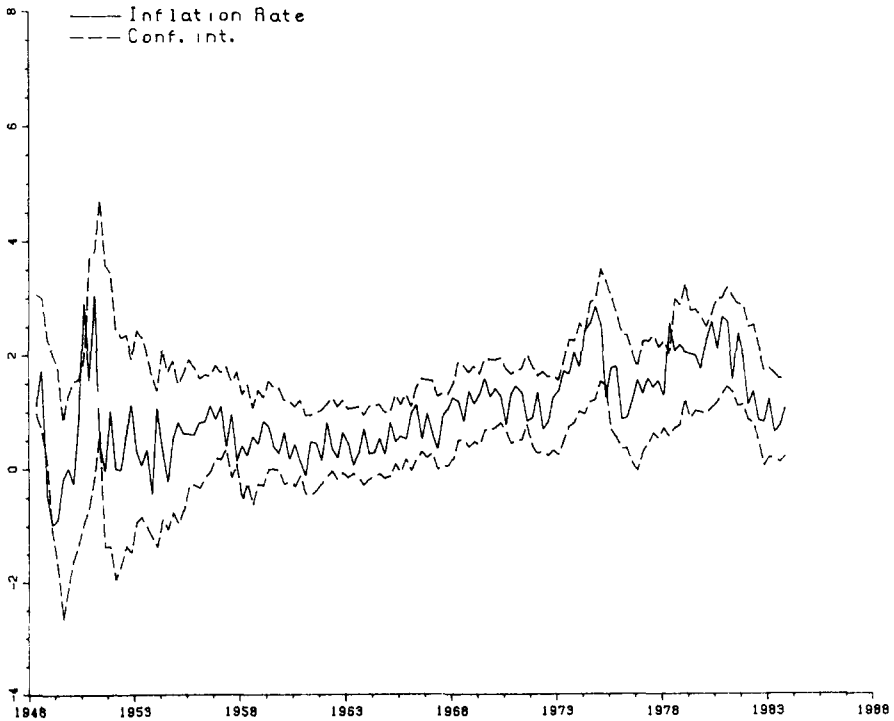


Fig. 4. 95% confidence intervals for GARCH(1,1).



In figs. 3 and 4 the actual inflation rate,  $\pi_t$ , is graphed together with 95% asymptotic confidence intervals for the one-step-ahead forecast errors for the two models (29) and (31). From the late forties until the mid-fifties the inflation rate was very volatile and hard to predict. This is reflected in the wide confidence intervals for the GARCH model. The sixties and early seventies, however, were characterized by a stable and predictable inflation rate, and the OLS confidence interval seems much too wide. Starting with the second oil crises in 1974 there is a slight increase in the uncertainty of the inflation rate, although it does not compare in magnitude to the uncertainty at the beginning of the sample period.

**Appendix**

*A.1. Proof of Theorem 1*

The basic idea of the proof follows that of Theorem 1 in Milhøj (1984). By definition

$$\varepsilon_t = \eta_t h_t^{1/2}, \quad \eta_t \stackrel{iid}{\sim} N(0, 1). \tag{A.1}$$

Subsequent substitution yields

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^q \alpha_i \eta_{t-i}^2 h_{t-i} + \sum_{i=1}^p \beta_i h_{t-i} \\ &= \alpha_0 + \sum_{j=1}^q \alpha_j \eta_{t-j}^2 \left( \alpha_0 + \sum_{i=1}^q \alpha_i \eta_{t-i-j}^2 h_{t-i-j} + \sum_{i=1}^p \beta_i h_{t-i-j} \right) \\ &\quad + \sum_{j=1}^p \beta_j \left( \alpha_0 + \sum_{i=1}^q \alpha_i \eta_{t-i-j}^2 h_{t-i-j} + \sum_{i=1}^p \beta_i h_{t-i-j} \right) \\ &\quad \vdots \\ &= \alpha_0 \sum_{k=0}^{\infty} M(t, k), \end{aligned} \tag{A.2}$$

where  $M(t, k)$  involves all the terms of the form

$$\prod_{i=1}^q \alpha_i^{a_i} \prod_{j=1}^p \beta_j^{b_j} \prod_{l=1}^n \eta_{t-l}^2,$$

for

$$\sum_{i=1}^q a_i + \sum_{j=1}^p b_j = k, \quad \sum_{i=1}^q a_i = n,$$

and

$$1 \leq S_1 < S_2 < \dots < S_n \leq \max\{kq, (k-1)q + p\}.$$

Thus,

$$M(t, 0) = 1,$$

$$M(t, 1) = \sum_{i=1}^q \alpha_i \eta_{t-i}^2 + \sum_{i=1}^p \beta_i,$$

$$\begin{aligned} M(t, 2) &= \sum_{j=1}^q \alpha_j \eta_{t-j}^2 \left( \sum_{i=1}^q \alpha_i \eta_{t-i-j}^2 + \sum_{i=1}^p \beta_i \right) \\ &\quad + \sum_{j=1}^p \beta_j \left( \sum_{i=1}^q \alpha_i \eta_{t-i-j}^2 + \sum_{i=1}^p \beta_i \right), \end{aligned}$$

and in general

$$M(t, k+1) = \sum_{i=1}^q \alpha_i \eta_{t-i}^2 M(t-i, k) + \sum_{i=1}^p \beta_i M(t-i, k). \quad (\text{A.3})$$

Since  $\eta_t^2$  is i.i.d., the moments of  $M(t, k)$  do not depend on  $t$ , and in particular

$$E(M(t, k)) = E(M(s, k)) \quad \text{for all } k, t, s, \quad (\text{A.4})$$

From (A.3) and (A.4) we get

$$\begin{aligned} E(M(t, k+1)) &= \left( \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right) E(M(t, k)) \\ &\quad \vdots \\ &= \left( \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right)^{k+1} E(M(t, 0)) \\ &= \left( \sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i \right)^{k+1}. \end{aligned} \quad (\text{A.5})$$

Finally by (A.1), (A.2) and (A.5),

$$\begin{aligned}
 E(\varepsilon_t^2) &= \alpha_0 E\left(\sum_{k=0}^{\infty} M(t, k)\right) \\
 &= \alpha_0 \sum_{k=0}^{\infty} E(M(t, k)) \\
 &= \alpha_0 \left(1 - \sum_{i=1}^q \alpha_i - \sum_{i=1}^p \beta_i\right)^{-1},
 \end{aligned}
 \tag{A.6}$$

if and only if

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1,$$

and  $\varepsilon_t^2$  converges almost surely.

$E(\varepsilon_t) = 0$  and  $\text{cov}(\varepsilon_t, \varepsilon_s) = 0$  for  $t \neq s$  follows immediately by symmetry.

### A.2. Proof of Theorem 2

By normality

$$E(\varepsilon_t^{2m}) = a_m E(h_t^m),
 \tag{A.7}$$

where  $a_m$  is defined in (10). The binomial formula yields

$$\begin{aligned}
 h_t^m &= (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1})^m \\
 &= \sum_{n=0}^m \binom{m}{n} \alpha_0^{m-n} \sum_{j=0}^n \binom{n}{j} \alpha_1^j \beta_1^{n-j} \varepsilon_{t-1}^{2j} h_{t-1}^{n-j}.
 \end{aligned}$$

Because

$$E(\varepsilon_{t-1}^{2j} h_{t-1}^{n-j} | \psi_{t-2}) = a_j h_{t-1}^n,$$

we have

$$E(h_t^m | \psi_{t-2}) = \sum_{n=0}^m h_{t-1}^n \binom{m}{n} \alpha_0^{m-n} \sum_{j=0}^n \binom{n}{j} a_j \alpha_1^j \beta_1^{n-j}.
 \tag{A.8}$$

Let  $w_t = (h_t^m, h_t^{m-1}, \dots, h_t)'$ , then by (A.8)

$$E(w_t | \psi_{t-2}) = d + Cw_{t-1},
 \tag{A.9}$$

where  $C$  is an  $m \times m$  upper triangular matrix with diagonal elements.

$$\mu(\alpha_1, \beta_1, i) = \sum_{j=0}^i \binom{i}{j} a_j \alpha_1^j \beta_1^{i-j}, \quad i = 1, \dots, m. \quad (\text{A.10})$$

Substituting in (A.9) yields

$$E(w_t | \psi_{t-k-1}) = (I + C + C^2 + \dots + C^{k-1})d + C^k w_{t-k}.$$

Since the process is assumed to start indefinitely far in the past with finite  $2m$  moments, the limit as  $k$  goes to infinity exists and does not depend on  $t$  if and only if all the eigenvalues of  $C$  lie inside the unit circle,

$$\lim_{k \rightarrow \infty} E(w_t | \psi_{t-k-1}) = (I - C)^{-1}d = E(w_t).$$

Because  $C$  is upper triangular, the eigenvalues are equal to the diagonal elements as given in (A.10). Tedious, but rather straightforward calculations show that  $\mu(\alpha_1, \beta_1, i) < 1$  implies  $\mu(\alpha_1, \beta_1, i-1) < 1$  for  $\alpha_1 + \beta_1 \leq 1$ , and  $\mu(\alpha, \beta_1, m) < 1$  suffices for the  $2m$ th moment to exist; cf. fig. 1.

Finally (11) follows from (A.7) and (A.8) by rearranging terms.

## References

- Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman, 1974, Estimation inference in nonlinear structural models, *Annals of Economic and Social Measurement*, no. 4, 653–665.
- Box, G.E.P. and J.M. Jenkins, 1976, *Time series analysis: Forecasting and control* (Holden-Day, San Francisco, CA).
- Breusch, T.S. and A.R. Pagan, 1978, A simple test for heteroskedastic city and random coefficient variation, *Econometrica* 46, 1287–1294.
- Coulson, N.E. and R.P. Robins, 1985, Aggregate economic activity and the variance of inflation: Another look, *Economics Letters* 17, 71–75.
- Domowitz, I. and C.S. Hakkio, 1985, Conditional variance and the risk premium in the foreign exchange market, *Journal of International Economics* 19, 47–66.
- Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.
- Engle, R.F., 1983, Estimates of the variance of U.S. inflation based on the ARCH model, *Journal of Money Credit and Banking* 15, 286–301.
- Engle, R.F. and D. Kraft, 1983, Multiperiod forecast error variances of inflation estimated from ARCH models, in: A. Zellner, ed., *Applied time series analysis of economic data* (Bureau of the Census, Washington, DC) 293–302.
- Engle, R.F., D. Lilien and R. Robins, 1985, Estimation of time varying risk premiums in the term structure, Discussion paper 85-17 (University of California, San Diego, CA).
- Garbade, K., 1977, Two methods for examining the stability of regression coefficients, *Journal of the American Statistical Association* 72, 54–63.
- Godfrey, L.J., 1978, Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica* 46, 1293–1302.
- Godfrey, L.J., 1981, On the invariance of the Lagrange multiplier test with respect to certain changes in the alternative hypothesis, *Econometrica* 49, 1443–1455.

- Granger, C.W.J. and A.P. Andersen, 1978, An introduction to bilinear time series models (Vandenhoeck and Ruprecht, Göttingen).
- Granger, C.W.J. and P. Newbold, 1977, Forecasting economic time series (Academic Press, New York).
- Harvey, A.C., 1982, The econometric analysis of time series (Philip Allen, Oxford).
- McCulloch, J.H., 1983, Interest-risk sensitive deposit insurance premia: Adaptive conditional heteroscedastic estimates, Discussion paper, June (Ohio State University, Columbus, OH).
- McLeod, A.J. and W.K. Li, 1990, Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis* 4, 269–273.
- Milhøj, A., 1984, The moment structure of ARCH processes, Research report 94 (Institute of Statistics, University of Copenhagen, Copenhagen).
- Weiss, A.A., 1982, Asymptotic theory for ARCH models: Stability, estimation and testing, Discussion paper 82-36 (University of California, San Diego, CA).
- Weiss, A.A., 1984, ARMA models with ARCH errors, *Journal of Time Series Analysis* 5, 129–143.
- White, H., 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.