BEBR
FACULTY WORKING PAPER NO. 89-1554

A Class of Nonlinear Arch Models: Properties, Testing and Applications

M. L. Higgins
A. K. Bera

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 89-1554

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 1989

A Class of Nonlinear Arch Models: Properties, Testing and Applications

M. L. Higgins
University of Wisconsin-Milwaukee

A. K. Bera, Associate Professor
Department of Economics

Revised: April 1989

A CLASS OF NONLINEAR ARCH MODELS: PROPERTIES, TESTING AND APPLICATIONS*

M. L. Higgins

University of Wisconsin-Milwaukee, Milwaukee, WI 53201

and

A. K. Bera

University of Illinois, Champaign, IL 61820

A class of nonlinear ARCH model is suggested. The proposed class encompasses several functional forms for ARCH which have been put forth in the literature. For this more general ARCH model, existence of moments are discussed. A Lagrange multiplier test is developed to test Engle's ARCH specification against the wider class of models. This test provides an easily computed diagnostic check of the adequacy of an ARCH model after it has been estimated. Lastly, the theory is applied to specify a nonlinear ARCH model for the weekly U.S./Canadian dollar exchange rate.

*We benefited from the comments of the participants of the Australasian Meeting of the Econometric Society, Canberra, August 28-31, 1988, and the Department of Economics, University of Alberta, Canada where earlier versions of the paper were presented. In particular we would like to thank Clive Granger for his comments and helpful discussion. We are also indebted to Nichola Dyer, Roger Koenker, Dan Nelson and Paul Newbold for many helpful suggestions. However, we retain the responsibility for any remaining errors.

*Correspondence should be addressed to M. L. Higgins, Department of Economics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201.

1. Introduction

Since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model in an influential paper by Engle (1982), there has been considerable interest in models in which the variance of the current observation is a function of past observations. Engle formally defined the ARCH regression model for a dependent variable y_t by specifying its conditional distribution as

$$y_t | \Phi_{t-1} \sim N(x_t'\beta, h_t)$$

where

$$h_t = h(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-p}; \alpha)$$

$$\varepsilon_t = y_t - x_t'\beta,$$

Φ_t is the information set at time t, x_t is a vector of exogenous variables and lagged values of the dependent variables, and β and α are parameter vectors. Engle suggested several functional forms for $h(\cdot)$, but concentrated primarily on the following linear ARCH model

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2. \quad (1)$$

To ensure that the conditional variance is strictly positive for all realizations of ε_t, the linear ARCH model (1) requires that the parameter space be restricted to $\alpha_0 > 0$ and $\alpha_i \geq 0$, $i=1,\ldots,p$.

Engle showed that if

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

then the ARCH process is stationary and the unconditional variance of ε_t is

$$\mathrm{Var}(\varepsilon_t) = \sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}$$

or

$$\alpha_0 = (1 - \sum_{i=1}^{p} \alpha_i)\sigma^2.$$

Substituting this into (1), we have

$$h_t = (1 - \sum_{i=1}^{p} \alpha_i)\sigma^2 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2.$$

Therefore, Engle's specification of the conditional variance can be viewed as a weighted average of the "global" variance σ^2 and the "local" variances ε_{t-1}^2, ..., ε_{t-p}^2.

Since Engle's paper, many extensions and generalizations of the ARCH model have appeared [see Engle and Bollerslev (1986) for a survey of ARCH models and their applications]. We believe that this research falls into three general areas, each of which addresses one of the assumptions of Engle's original ARCH specification.

First, a majority of research has examined ARCH models in which the conditional variance h_t is a function not only of the ε_t's but also of other elements of Φ_{t-1}. Weiss (1984, 1986) suggests ARCH models in which h_t is a function of lagged values of y_t, exogenous variables, and predictions of y_t based on elements of the information set Φ_{t-1}. Engle, Granger and Kraft (1984) and Granger, Robins and Engle (1982) consider bivariate time series for which the conditional variance of each series is dependent upon lagged values of the other. Bollerslev (1985) makes h_t also a function of lagged values of itself, i.e. $h_{t-1},\ldots,h_{t-m}$, and the resulting model is called the generalized-ARCH or GARCH model.

A second area of research addresses the assumption of conditional normality. The conditional distribution of data for which ARCH models are used, is frequently skewed and leptokurtic. To account for heavy tails of the conditional distribution, Engle and Bollerslev (1986) and Bollerslev (1987) examine GARCH models in which the conditional distribution is assumed to be Student's-t rather than normal. For the same reason, Lee and Tse (1988) suggest using a Gram-Charlier type distribution. It is difficult, however, to specify a conditional distribution which allows for both skewness and kurtosis, and still maintain the tractability of the model. Proper specification of the conditional distribution of an ARCH process is a question which warrants further research.

A third extension of the ARCH model, by far the one having received the least amount of attention to date, considers functional forms for h_t other than Engle's linear ARCH specification. Engle (1982) in fact suggested several functional forms, but concentrated on the linear model for its analytic convenience and its plausibility as a data generating mechanism. Engle and Bollerslev (1986) and Geweke (1986) have suggested other functional forms. The use of alternatives to the linear ARCH model is hindered by the lack of a specification test to determine the adequacy of the linear ARCH model and a procedure for selecting an alternative functional form if the linear ARCH model is rejected.

The specification of the correct conditional variance function is important in several respects. The accuracy of forecast intervals depends upon selecting the function which correctly relates the future variances to the current information set. Also, the test for detecting the presence of ARCH is partially determined by the functional form of the ARCH process. Further, Pagan and Sabau (1987) have shown that an incorrect functional form of the ARCH process for the errors of a regression model can result in inconsistent maximum likelihood estimators of the regression parameters.

In this paper, we propose a nonlinear ARCH model which encompasses several of the models which have been proposed in the literature. Using this more general ARCH model, the specification testing of the functional form of the ARCH model is addressed. Section 2 is a survey of functional forms which have been proposed for ARCH. In Section 3, we introduce a nonlinear ARCH model and show that it encompasses models discussed in Section 2. In Section 4, a Lagrange multiplier (LM) test is developed to test the linear ARCH model against a class of nonlinear ARCH models. In Section 5, the theory is applied to specifying a functional form for an ARCH model for the Canadian weekly exchange rate.

2. Functional forms for ARCH

Although Engle (1982) focused on the convenient linear ARCH model, he acknowledged that "it is likely that other formulations of the variance model may be more appropriate for particular applications" (p. 993). He suggested two alternatives, the exponential and absolute value models

$$h_t = \exp(\alpha_0+\alpha_1\varepsilon_{t-1}^2+\ldots+\alpha_p\varepsilon_{t-p}^2) \quad (2)$$

$$h_t = \alpha_0+\alpha_1|\varepsilon_{t-1}|+\ldots+\alpha_p|\varepsilon_{t-p}|. \quad (3)$$

The exponential model (2) has the advantage that the variance is positive for all values of α; however, as stated by Engle, it has the unfortunate property that data generated by such a process will have infinite variance, making estimation and inference difficult. The absolute value model (3), like the linear model, requires restrictions on the parameter space to ensure that the variance is positive. For this model, however, the variance of the generated data will be finite for all positive parameter values, a property not shared by the linear ARCH model.

In an empirical application, Engle and Bollerslev (1986) report having estimated a variety of functional forms for a GARCH model of the U.S. dollar/Swiss franc exchange rate. They report results only for the two best models. The ARCH analogues of these GARCH are

$$h_t = \alpha_0+\alpha_1|\varepsilon_{t-1}|^\mu+\ldots+\alpha_p|\varepsilon_{t-1}|^\mu \quad (4)$$

and

$$h_t = \alpha_0+\alpha_1(2F(\varepsilon^2_{t-1}/\mu)-1)+\ldots+\alpha_p(2F(\varepsilon^2_{t-p}/\mu)-1) \quad (5)$$

where F is the cumulative normal distribution. Model (4) is a simple extension of the absolute value model (3). The model (5) is an attempt to provide a model in which the conditional variance h_t remains bounded as the values of ε^2_{t-i}, i=1,...,p, become arbitrarily large.

Concerned with the non-negativity restrictions on the parameters of the linear ARCH model, Geweke (1986) suggests the functional form

$$\log(h_t) = \alpha_0+\alpha_1\log(\varepsilon^2_{t-1})+\ldots+\alpha_p\log(\varepsilon^2_{t-p}) \quad (6)$$

which ensures that the conditional variance is positive for all values of α. Geweke also demonstrates that the log-likelihood for this ARCH model is globally concave, making maximum likelihood estimation comparatively easy. Engle and Bollerslev (1986) criticize this specification, however, because the likelihood function becomes undefined if a residual of zero is encountered.

Each of the above models has its individual benefits and limitations. Clearly any one model cannot be chosen *a priori* as the most plausible specification of an ARCH process. Their usefulness depends upon the particular empirical application.

3. A general functional form for ARCH

In this section we propose a general functional form for the ARCH model and then show that this more general model encompasses models in Section 2.

Consider the ARCH model (2) with

$$h_t = \left[\phi_0(\sigma^2)^\delta + \phi_1(\varepsilon_{t-1}^2)^\delta + \ldots . + \phi_p(\varepsilon_{t-p}^2)^\delta\right]^{1/\delta} \qquad (7)$$

where

$\sigma^2 > 0$

$\phi_i \geq 0$ for $i = 0,1,\ldots,p$

$\delta > 0$

and the ϕ_i´s are such that

$$\sum_{i=0}^{p} \phi_i = 1.$$

The conditional variance function (7) has p+3 parameters. The restriction that the ϕ_i´s sum to one reduces the dimension of the parameter space by one; hence, the model effectively has one more parameter than Engle´s model. The restriction that $\delta>0$ ensures that the conditional variance is defined for all ε_t´s. We will refer to a model with conditional variance function (7) as a nonlinear ARCH model of order p, or in short NARCH(p).

The function (7) is familiar to economists as the constant elasticity of substitution (CES) production function of Arrow *et al.* (1961). If we heuristically view the conditional variance as output and the "global" and "local" variances as inputs, then the

ARCH model has the form of a linear production function with infinite elasticity of substitution. The NARCH specification, on the other hand, is very flexible; the elasticity of substitution $1/(1-\delta)$ can take values in the range $(1,\infty)$ for $0<\delta\leq 1$. Equation (7) could also be written as

$$\frac{h_t^{\delta}-1}{\delta} = \phi_0\frac{(\sigma^2)^{\delta}-1}{\delta}+\phi_1\frac{(\varepsilon_{t-1}^2)^{\delta}-1}{\delta}+\ldots+\phi_p\frac{(\varepsilon_{t-p}^2)^{\delta}-1}{\delta} \tag{8}$$

which is a Box-Cox (1964) power transformation on both sides of the Engle (1982) specification. This transformation has been found to be quite useful in econometric functional form specification analysis. It has traditionally been used to linearize otherwise nonlinear models [see Carroll and Ruppert (1988, p.118)]. In time series analysis, Granger and Newbold (1976) and Hopwood, McKeown and Newbold (1984) have found that incorporating a power transformation into the ARIMA model is beneficial in terms of forecast accuracy. Therefore, the NARCH specification can be expected to improve the estimates of the forecast intervals. Moreover, as we will see below, the NARCH form encompasses different conditional heteroskedasticity models discussed in Section 2, and provides a framework for developing a specification test for Engle's linear ARCH model.

We now show that some of the models of Section 2 are special cases of the NARCH model (7).

<u>Proposition 1</u>: The conditional variance function (7) is equivalent to:

(i) Engle's conditional variance function when $\delta=1$

(ii) Geweke's conditional variance function as $\delta \to 0$.

<u>Proof</u>: The proof of (i) is obvious. Set $\delta=1$ in (7) and let $\alpha_0 = \phi_0(\sigma^2)$ and $\alpha_i=\phi_i$, for i=1,...k. To prove (ii), express the NARCH conditional variance function as (8) and take the limit as $\delta \to 0$ of both sides, giving

$$\log(h_t) = \phi_0 \log(\sigma^2)+\phi_1 \log(\varepsilon_{t-1}^2)+\ldots+\phi_p \log(\varepsilon_{t-p}^2)$$

which is equivalent to Geweke's model with $\alpha_0=\phi_0 \log(\sigma^2)$ and $\alpha_i=\phi_i$, for i=1,...,p.

Here we should note that Geweke's model is only a limiting case of the NARCH model. Imposing the restriction $\delta>0$ allows for models arbitrarily close to Geweke's model, and yet guarantees that the log-likelihood function is defined. Unfortunately, the absolute value model (3) is not nested within this general model. However, the NARCH model can be generalized by including an additional parameter, in the same way that the constant elasticity form is generalized to the variable elasticity of substitution form. The variance function becomes

$$h_t = \left[\phi_0(\sigma^2)^\delta + \phi_1(\varepsilon_{t-1}^2)^\delta + \ldots + \phi_p(\varepsilon_{t-p}^2)^\delta\right]^{\mu/\delta}. \qquad (9)$$

Setting both $\delta=\frac{1}{2}$ and $\mu=\frac{1}{2}$ makes (9) equivalent to the absolute value model (3). We shall not consider this generalization further in this paper; rather it is an area for future research.

The assumption that the conditional mean of ε_t is zero ensures that the unconditional mean and all autocorrelations of ε_t are also zero. The nonlinearity of the conditional variance function makes it difficult to find an explicit expression for the unconditional variance. We are, however, able to state a sufficient condition for the unconditional variance of the NARCH(1) model to be finite.

<u>Theorem 1</u>: The variance of the NARCH(1) model is finite if

$$\phi_1 \pi^{-\frac{1}{2}} 2^{\delta} \Gamma(\delta+\tfrac{1}{2}) < 1$$

where $\Gamma(\cdot)$ is the gamma function.

The proof of Theorem 1 is given in the Appendix. Extension of the theorem to the NARCH(p) model is straightforward. The region for the parameter space for which the variance is finite is shown in Figure 1.

4. A Lagrange multiplier test for NARCH

Once it has been determined that conditional heteroskedasticity is present in the data, it is natural to begin the specification search for functional form with Engle's linear model. It would, therefore, be useful to possess a test to determine whether Engle's model provides an adequate description of the data, or whether a wider class of functional forms for the conditional variance needs to be considered. In this section we derive an LM test for Engle's model against the more general class of ARCH models with conditional variance function (7). Since Geweke's model is a member of this class, the test should have good power against this alternative.

We are interested in testing the hypotheses

$H_0 \; : \delta = 1$

$H_1 \; : \delta \neq 1$

in (7), i.e., that Engle's model provides an adequate description of the data generating process. The log likelihood function for a single observation, omitting a constant, is

$$l_t = -\frac{1}{2}\log(h_t) - \frac{\varepsilon_t^2}{2h_t}.$$

Let $\theta' = (\beta', \phi_1, \ldots, \phi_p, \sigma^2, \delta)$ be the complete parameter vector. The LM test for the above hypotheses is given by

$$LM = d(\tilde{\theta})' I(\tilde{\theta})^{-1} d(\tilde{\theta})$$

where

$$d(\theta) = \sum_{t=1}^{T} \frac{dl_t}{d\theta}$$

is the score vector and

$$I(\theta) = -E\left[\sum_{t=1}^{T} \frac{d^2 l_t}{d\theta\, d\theta'}\right]$$

is the information matrix and "~" denotes quantities evaluated at the maximum likelihood estimates subject to the restriction $\delta=1$. Let $v = (\sigma^2, \phi_1, \ldots, \phi_p, \delta)$ be the vector of variance parameters. The elements of $dl_t/d\theta$ are given by

$$\frac{dl_t}{d\beta} = \frac{\varepsilon_t x_t'}{h_t} + \frac{1}{2h_t}\frac{dh_t}{d\beta}\left[\frac{\varepsilon_t^2}{h_t} - 1\right]$$

and

$$\frac{dl_t}{dv} = \frac{1}{2h_t}\frac{dh_t}{dv}\left[\frac{\varepsilon_t^2}{h_t} - 1\right] \tag{10}$$

where

$$\frac{dh_t}{d\beta_i} = \left[\phi_0(\sigma^2)^\delta+\phi_1(\varepsilon_{t-1}^2)^\delta+\ldots+\phi_p(\varepsilon_{t-p}^2)^\delta\right]^{(1/\delta)-1}$$

$$\cdot\sum_{j=1}^{p}\delta\phi_j(\varepsilon_{t-j}^2)^{\delta-1}\cdot 2\varepsilon_{t-j}(-x_{t-j,i})$$

$$= h_t^{1-\delta}\sum_{j=1}^{p}\delta\phi_j(\varepsilon_{t-j}^2)^{\delta-1}\cdot 2\varepsilon_{t-j}(-x_{t-j,i}) \quad \text{for } i=1,\ldots,k \tag{11}$$

$$\frac{dh_t}{d\sigma^2} = \left[\phi_0(\sigma^2)^\delta+\phi_1(\varepsilon_{t-1}^2)^\delta+\ldots+\phi_p(\varepsilon_{t-p}^2)^\delta\right]^{(1/\delta)-1}\cdot\phi_0(\sigma^2)^{\delta-1}$$

$$= \phi_0\left[\frac{h_t}{\sigma^2}\right]^{1-\delta} \tag{12}$$

$$\frac{dh_t}{d\phi_i} = \frac{1}{\delta}\left[\phi_0(\sigma^2)^\delta+\phi_1(\varepsilon_{t-1}^2)^\delta+\ldots+\phi_p(\varepsilon_{t-p}^2)^\delta\right]^{(1/\delta)-1}$$

$$\cdot\left[-(\sigma^2)^\delta+(\varepsilon_{t-i}^2)^\delta\right]$$

$$= \frac{h_t^{1-\delta}}{\delta}\left[-(\sigma^2)^\delta+(\varepsilon_{t-i}^2)^\delta\right] \quad \text{for } i = 1,\ldots,p \tag{13}$$

$$\frac{dh_t}{d\delta} = h_t\left[\frac{\pi_t}{\phi_0(\sigma^2)^\delta+\phi_1(\varepsilon_{t-1}^2)^\delta+\ldots +\phi_p(\varepsilon_{t-p}^2)^\delta}\cdot\delta^{-1}\right.$$

$$\left. - \log\left[\phi_0(\sigma^2)^\delta+\phi_1(\varepsilon_{t-1}^2)^\delta+\ldots+\phi_p(\varepsilon_{t-p}^2)^\delta\right]\cdot\delta^{-2}\right]$$

$$= \frac{1}{\delta}\left[\pi_t\cdot h_t^{1-\delta}-h_t\log(h_t)\right] \tag{14}$$

and where

$$\pi_t = \phi_0\log(\sigma^2)(\sigma^2)^\delta + \sum_{i=1}^{p} \phi_i\log(\varepsilon_{t-i}^2)(\varepsilon_{t-i}^2)^\delta$$

The form of the LM test simplifies significantly when the information matrix is block diagonal between ß and v. The block diagonality can be shown in general; for simplicity, we show it for NARCH(1).

Theorem 2: For the NARCH(1), the information matrix is block diagonal between the regression parameters ß and the variance parameters v, that is:

$$-E\left[\sum_t \frac{d^2 l_t}{d\beta dv'}\right] = 0$$

The proof is given in the appendix. Under block diagonality the LM statistic simplifies to

$$LM = d_v(\tilde{\theta})'I_v(\tilde{\theta})^{-1}d_v(\tilde{\theta})$$

where

$$d_v(\theta) = \sum_{t=1}^{T} \frac{dl_t}{dv}$$

$$I_v(\theta) = -E\left[\sum_{t=1}^{T} \frac{d^2 l_t}{dv dv'}\right]$$

The matrix of second partial derivatives is

$$\frac{d^2 l_t}{dv dv'} = -\frac{1}{2h_t^2} \frac{dh_t}{dv} \frac{dh_t}{dv'}\left[\frac{\varepsilon_t^2}{h_t}\right] + \left[\frac{\varepsilon_t^2}{h_t} - 1\right] \frac{d}{dv'}\left[\frac{1}{2h_t} \frac{dh_t}{dv}\right] \qquad (15)$$

and the information matrix $I_v(\theta)$ is given by the negative of the expectation of the matrix (15) summed over t. This expectation can be simplified by taking iterated expectations on the information set Φ_{t-1}:

$$-E \frac{d^2 l_t}{dv dv'} = -E\left[E \frac{d^2 l_t}{dv dv'}\middle|\Phi_{t-1}\right]$$

$$= E\left[\frac{1}{2h_t^2} \frac{dh_t}{dv} \frac{dh_t}{dv'}\right]$$

and hence, $I_v(\theta)$ can be estimated by

$$\hat{I}_v(\theta) = \frac{1}{2} \sum_{t=1}^{T} \left[\frac{1}{h_t} \frac{dh_t}{dv}\right]\left[\frac{1}{h_t} \frac{dh_t}{dv'}\right].$$

The LM test for H_0 is then

$$LM = d_v(\tilde{\theta})' I_v(\tilde{\theta})^{-1} d_v(\tilde{\theta})$$

$$= \frac{1}{2}\left[\Sigma\left[\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t} - 1\right]\left[\frac{1}{\tilde{h}_t}\frac{d\tilde{h}_t}{dv}\right]\right]'\left[\Sigma\left[\frac{1}{\tilde{h}_t}\frac{d\tilde{h}_t}{dv}\right]\left[\frac{1}{\tilde{h}_t}\frac{d\tilde{h}_t}{dv'}\right]\right]^{-1}$$

$$X\left[\Sigma\left[\frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t} - 1\right]\left[\frac{1}{\tilde{h}_t}\frac{d\tilde{h}_t}{dv}\right]\right]$$

The test statistic will be asymptotically distributed as chi-square with one degree of freedom under the null hypothesis. If we let

$$f_t = \frac{\tilde{\varepsilon}_t^2}{\tilde{h}_t} - 1$$

and

$$z_t = \frac{1}{\tilde{h}_t}\frac{d\tilde{h}_t}{dv},$$

and let $f'=(f_1,...,f_T)$ and $Z'=(z_1,...,z_T)$, the statistic can be expressed as

$$LM = \frac{1}{2}\left[\Sigma\ f_t z_t\right]'\left[\Sigma\ z_t z_t'\right]^{-1}\left[\Sigma\ f_t z_t\right]$$

$$= \tfrac{1}{2}f'Z(Z'Z)^{-1}Z'f.$$

This statistic can be computed as ½ the SSR from the regression of f on Z. Furthermore, since $\text{plim}(f'f/T) = 2$, an asymptotically equivalent form of the statistic is

$$LM = T\cdot f'Z(Z'Z)^{-1}Z'f/f'f$$

which can be computed as T times the uncentered coefficient of determination from the regression of f on Z.

The test statistic is similar in form to Engle's original LM test for ARCH. It differs in the function h_t and the elements of dh_t/dv. In the above test, under the null, h_t is the conditional variance as estimated from Engle's model. The elements of dh_t/dv, evaluated under the null hypothesis $\delta=1$, simplify to

$$\frac{d\tilde{h}_t}{d\phi_i} = -(\tilde{\sigma}^2) + \tilde{\varepsilon}^2_{t-i} \quad \text{for } i=1,\ldots,p$$

$$\frac{d\tilde{h}_t}{d\sigma^2} = \tilde{\phi}_0$$

$$\frac{d\tilde{h}_t}{d\delta} = \tilde{\pi}_t - \tilde{h}_t \log(\tilde{h}_t)$$

where

$$\tilde{\pi}_t = \tilde{\phi}_0 \tilde{\sigma}^2 \log(\tilde{\sigma}^2) + \sum_{i=1}^{p} \tilde{\phi}_i \tilde{\varepsilon}^2_{t-i} \log(\tilde{\varepsilon}^2_{t-i})$$

The test requires estimating Engle's ARCH model and computing the conditional variance for each observation. It does not, however, represent a significant computing burden. Presumably, a researcher has already concluded that ARCH is present in the data and would naturally proceed by estimating Engle's model. Once the MLEs of Engle's model are obtained, the test can be performed on any standard regression package. The test, therefore, should be viewed as a diagnostic check of the adequacy of Engle's model after it has been estimated.

5. An application to the U.S./Canadian exchange rate

ARCH is frequently used to model the volatility of exchange rates. In this section, we compare the performance of ARCH and NARCH as models of the weekly U.S/Canadian exchange rate from January 1973 to June 1986. Let $y_t = \log(e_t/e_{t-1})$, where e_t is the spot price of a Canadian dollar in terms of the U.S. dollar. The analyzed series, y_t, is the continuously compounded percentage rate of return for holding the Canadian currency one week. The effective sample size T is 649. The series y_t was centered about its mean prior to analysis and all estimation was done using IMSL subroutine ZXMIN. This series was chosen because its conditional mean can be represented by a simple autoregressive (AR) process and because ARCH of low order was evident.

Examination of the sample autocorrelation and partial autocorrelation functions of y_t indicate that an AR(1) process is a suitable model for the conditional mean. The estimated AR model is

$$y_t = \underset{(.039)}{.339}\, y_{t-1} \qquad (16)$$

where the standard error of the estimate is shown in parentheses. Higher order AR models were also fitted, but only the coefficient of the first lagged value of the series was found to be significant. Diagnostic checks of the residuals of (16) do not indicate the presence of serial correlation.

The autocorrelations of the squared residuals from (16), however, reveal that nonlinearity is present. Using the standard error $1/T^{½} = .039$ [see Mcleod and Li (1983)], the first autocorrelation of the squared residuals, .12, is significant at the 5% level; the rest are insignificant. Engle's LM test for ARCH was performed for orders 1 through 10. The statistic for 1st order ARCH is highly significant, but the value of the statistic increases very slightly as additional ARCH terms are included. Hence, an ARCH(1) model is identified and estimated to be

$$y_t = \underset{(.054)}{.29}\ y_{t-1}$$

$$h_t = \underset{(.01)}{.116} + \underset{(.089)}{.449}\ \varepsilon^2_{t-1}$$

Log-likelihood function = -332.191

Higher order ARCH models were also fitted, but the additional ARCH parameters were insignificant.

To determine whether linear ARCH provides an adequate model, the LM test of Section 4 for NARCH was performed. The computed value of the test statistic is 6.88, which is highly significant for a chi-square with one degree of freedom. The estimated NARCH(1) model is

	β	σ^2	ϕ_1	δ
estimate	.32	.247	,255	.148
standard error	.00019	.0257	.0825	.167

Log-likelihood Function = -325.537

Initial parameter values for estimating the NARCH were obtained from the estimated ARCH model, with δ taken to be 1 as implied by the linear ARCH specification. The striking feature of the estimated NARCH model is that δ falls drastically from 1 to .148. An asymptotic t-test of the hypothesis that $\delta = 1$ is easily rejected. In addition, the likelihood ratio (LR) statistic is 2(332.191-325.537) = 13.308, which is highly significant.

Therefore, all three of the tests, LM, LR and t-test, reject the linearity imposed by the ARCH model. Table 1 presents the autocorrelations of the standardized residuals $\hat{\varepsilon}_t/\hat{h}_t^{1/2}$ from the AR, ARCH and NARCH models; Table 2 presents the autocorrelations of their squares. For a correctly specified model, these autocorrelations should be close to zero. Table 1 shows that the autocorrelations from the NARCH model are closest to zero, although there is still slight evidence of seventh order autocorrelation, perhaps due to the weekly nature of the data. In terms of the autocorrelations of the squares of the standardized residuals, as we can see from Table 2, clearly the ARCH model is an improvement over the AR model. Yet additional improvement, though modest, is apparent when going from the ARCH to the NARCH model.

To further compare the performances of ARCH and NARCH, we plot the residuals and their 95% confidence intervals $\pm 1.96(h_t)^{1/2}$ in figures 2-5. From figures 2 and 3, the ARCH and NARCH residuals are seen to be almost indistinguishable, indicating that NARCH may not provide a significant improvement over ARCH in terms of the point forecast. The estimated AR parameters for the two models differ by only .03. The confidence intervals, however, are very different. Comparing figures 4 and 5, the NARCH conditional variances seem to more accurately reflect the behavior of the series. The confidence intervals of the ARCH model frequently decline to the lower bound $\pm 1.96(\alpha_0)^{1/2} = \pm .668$ during "stable" periods, yet ±.668 is clearly too large an interval during these periods. On the other hand, during the "volatile" periods, ARCH seems to frequently overstate the conditional variance and can vary drastically from one observation to the next. The confidence intervals of the NARCH model, however, track the series well. During the stable periods, the intervals become considerably smaller than the ARCH intervals. In contrast to the ARCH intervals, the lower bound for the NARCH intervals is $\pm 1.96(\phi_0^{1/\delta}\sigma^2)^{1/2} = \pm .360$. As should be expected from a good conditional variance model, we can predict the series with higher confidence during less volatile periods. During the volatile periods, the NARCH intervals widen as required, yet the transition is smooth and not subject to the erratic variation which occurs for the ARCH intervals.

Summarizing the above results, based on the statistical tests it appears that a nonlinear model improves the functional form specification of the conditional variances. Ideally we would also like the standardized residuals $\hat{\varepsilon}_t/\hat{h}_t^{1/2}$ to behave like white noise and for our model there is "modest" improvement as indicated by their autocorrelation function. From the plot of the forecast intervals of the residuals, we observe a more plausible evolution of the conditional variance over time.

Appendix

Proof of Theorem 1: We require the following lemma.

Lemma: Assuming that a NARCH(1) process began infinitely far in the past with all initial moments finite, then

$$E(|\varepsilon_t|^{2\delta}) < \infty$$

if

$$\phi_1 \pi^{-\frac{1}{2}} 2^{\delta} \Gamma(\delta+\tfrac{1}{2}) < 1$$

where $\Gamma(\cdot)$ is the gamma function.

Proof of Lemma: The process (7) can be expressed as the stochastic difference equation

$$\varepsilon_t = z_t (h_t)^{1/2}$$

where

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

and is independent of past ε_t's. Therefore

$$E|\varepsilon_t|^{2\delta} = E|z_t|^{2\delta}\left[\phi_0(\sigma^2)^{\delta} + \phi_1 E|\varepsilon_{t-1}|^{2\delta}\right]$$

$$= \phi_0 \mu_* (\sigma^2)^{\delta} + \phi_1 \mu_* E|\varepsilon_{t-1}|^{2\delta}$$

where $\mu_* = E|z_t|^{2\delta}$ which is assured to exist by the normality of z_t. This is a first order linear difference equation in $E|\varepsilon_t|^{2\delta}$.

If initial moments are finite, $E|\varepsilon_t|^{2\delta}$ will converge to a finite value if

$$\phi_1\mu_* = \phi_1 E|z_t|^{2\delta} = \phi_1\pi^{-\frac{1}{2}}2^{\delta}\Gamma(\delta+\tfrac{1}{2}) < 1$$

which completes the proof of the lemma.

The condition for the lemma is also sufficient for the finiteness of the variance. For if $\delta>1$, then

$$E(\varepsilon_t^2) = E\left[E(\varepsilon_t^2 \mid \Phi_{t-1})\right]$$

$$= E\left[\phi_0(\sigma^2)^{\delta}+\phi_1(\varepsilon_{t-1}^2)^{\delta}\right]^{1/\delta} \qquad (A.1)$$

$$\leq \left[\phi_0(\sigma^2)^{\delta}+\phi_1 E|\varepsilon_{t-1}|^{2\delta}\right]^{1/\delta}$$

where the inequality follows from Jensen's inequality since (A.1) is a concave function of $|\varepsilon_{t-1}|^{2\delta}$. The lemma now assures that the expectation is finite, and hence the variance is finite. For $\delta\leq 1$, we apply Minkowski's inequality [see White (1984, p.34)], to establish

$$E(\varepsilon_t^2) = E\left[E(\varepsilon_t^2 \mid \Phi_{t-1})\right]$$

$$= E\left[\phi_0(\sigma^2)^{\delta}+\phi_1(\varepsilon_{t-1}^2)^{\delta}\right]^{1/\delta}$$

$$\leq \left[\phi_0(\sigma^2)^{\delta}+\phi_1\left[E(\varepsilon_{t-1}^2)\right]^{\delta}\right]^{1/\delta} \qquad (A.2)$$

The expression in (A.2) is monotonically increasing in $E(\varepsilon_{t-1}^2)$. By repeatedly applying Minkowski's inequality

$$E(\varepsilon_t^2) \le \left[\phi_0(\sigma^2)^\delta+\phi_1\phi_0(\sigma^2)^\delta+\phi_1\left[E(\varepsilon_{t-1}^2)\right]^\delta\right]^{1/\delta}$$

$$\vdots$$

$$\le \left[\phi_0(\sigma^2)^\delta\cdot\sum_{i=0}^{k-1}\phi_1^i+\phi_1^k\left[E(\varepsilon_{t-1}^2)\right]^\delta\right]^{1/\delta}.$$

As $k\to\infty$, the variance will be finite if $\phi_1<1$. The restrictions on the parameter space of (7) guarantee $\phi_1<1$, which completes the proof.

Proof of Theorem 2: From (10)

$$\frac{d^2 l_t}{dv_j d\beta_i} = -\frac{\varepsilon_t x_{t,i}}{h_t^2}\frac{dh_t}{dv_j} - \frac{1}{2h_t^2}\frac{dh_t}{dv_j}\frac{dh_t}{d\beta_i}\left[\frac{\varepsilon_t^2}{h_t}-1\right]$$

$$+\frac{1}{2h_t}\frac{d^2h_t}{dv_jd\beta_i}\left[\frac{\varepsilon_t^2}{h_t}-1\right] - \frac{\varepsilon_t^2}{2h_t^3}\frac{dh_t}{dv_j}\frac{dh_t}{d\beta_i} \qquad (A.3)$$

The elements of the information matrix between ß and v are given by the negative of the expectation of the right hand side of (A.3) summed over t. The expectation simplifies by first taking iterated expectations on the information set at time t-1.

$$\sum_t -E\left[\frac{d^2l_t}{dv_jd\beta_i}\right] = \sum_t -E\left[E\left[\frac{d^2l_t}{dv_jd\beta_i}\right]\middle|\Phi_{t-1}\right]$$

$$= \sum_t E\left[\frac{1}{2h_t^2}\frac{dh_t}{dv_j}\frac{dh_t}{d\beta_i}\right]$$

If the expectation of the term in square brackets is zero, then the theorem is proved. From (7) and (12) to (14), it is evident that h_t and dh_t/dv_i are symmetric functions of ε_{t-1}, while from (11), $dh_t/d\beta_i$ is an antisymmetric function of ε_{t-1}. Hence, the whole expression is antisymmetric in ε_{t-1}, and since ε_{t-1} has a symmetric distribution around zero, the expectation is zero.

References

Arrow, K. J., H. B. Chenery, B. S. Minhas and R. M. Solow, 1961, Capital-labor substitution and economic efficiency, Review of Economics and Statistics 43, 225-250.

Bollerslev, T., 1986, A generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307-327.

Bollerslev, T., 1987, A conditionally heteroskedastic time series model for speculative prices and rates of return, Review of Economics and Statistics 69, 542-547.

Box, G.E.P. and D.R. Cox, 1964, An analysis of transformations, Journal of Royal Statistical Society B, 26, 211-243.

Carroll, R.J. and D. Ruppert, 1988, Transformation and weighting in regression (Chapman and Hall, New York).

Engle, R. F., 1982, Autoregressive conditional hetero-scedasticity with estimates of the variance of U.K. inflation, Econometrica 50, 987-1008.

Engle, R. F., C. W. J. Granger and D. Kraft, 1986, Combining competing forecasts of inflation using a bivariate ARCH model, Journal of Economic Dynamics and Control 8, 151-165.

Engle, R. F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, Econometric Reviews 5, 1-87.

Geweke, J., 1986, Comment, Econometric Reviews 5, 57-61.

Granger, C.W.J. and P. Newbold, 1976, Forecasting transformed series, Journal of Royal Statistical Society B, 38, 189-203.

Granger, C. W. J., R. P. Robins and R. F. Engle, 1984, Wholesale and retail prices: bivariate time series modeling with forecastable error variances, in: David A. Belsley and Edwin Kuh, eds., Model Reliability (MIT Press, Cambridge), 1-17.

Hopwood, W.S., J.C. McKeown and P. Newbold, 1984, Time series forecasting models involving power transformations, Journal of Forecasting, 3, 57-61.

Lee, T. K. Y. and Y. K. Tse, 1988, Term structure of interest rates in the Asian dollar market: ARCH-M Modelling with Autocorrelated Non-normal Errors, mimeograph.

Pagan, A. R. and H. Sabau, 1987, On the inconsistency of the MLE in certain heteroskedastic regression models, mimeograph.

Weiss, Andrew A., 1984, ARMA models with ARCH errors, Journal of Time Series Analysis 5, 129-143.

Weiss, Andrew A., 1986, Asymptotic theory for ARCH models: estimation and testing, Econometric Theory 2, 107-131.

White, Halbert, 1984, Asymptotic theory for econometricians (Academic Press, Orlando).

Table 1

Autocorrelations[a] of standardized residuals ($\hat{\varepsilon}_t/\hat{h}_t^{1/2}$)

Lag	AR	ARCH	NARCH
1	.01	.07	.04
2	-.01	.01	.00
3	-.02	.01	.01
4	-.06	-.04	-.04
5	-.03	-.04	-.03
6	-.02	-.01	-.01
7	-.08	-.08	-.07
8	.01	.01	.01
9	-.01	.00	-.01
10	-.03	-.05	-.05

[a] Approximate standard error for these autocorrelations is $1/T^{1/2}=0.039$.

Table 2

Autocorrelations[a] of squared standardized residuals $(\hat{\varepsilon}_t/\hat{h}_t^{1/2})^2$

Lag	AR	ARCH	NARCH
1	.12	-.02	.00
2	.03	.01	.01
3	.04	.05	.03
4	.00	.01	.00
5	.00	.00	-.01
6	-.02	-.02	-.03
7	.02	.04	.02
8	-.01	-.01	-.01
9	.05	.01	.01
10	.00	-.02	-.02

[a] Approximate standard error for these autocorrelations is $1/T^{1/2}=0.039$.

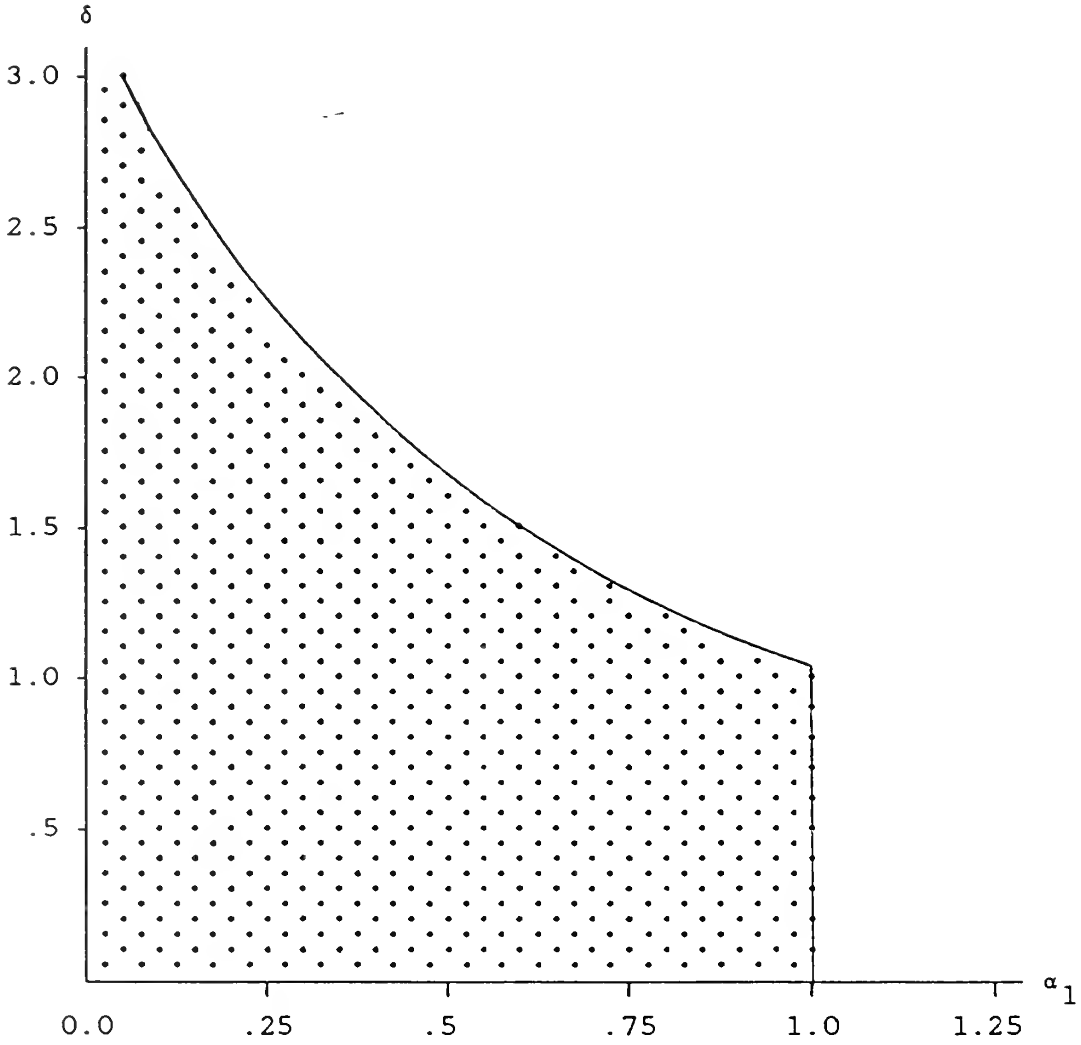

Figure 1. Region for finite NARCH(1) variance

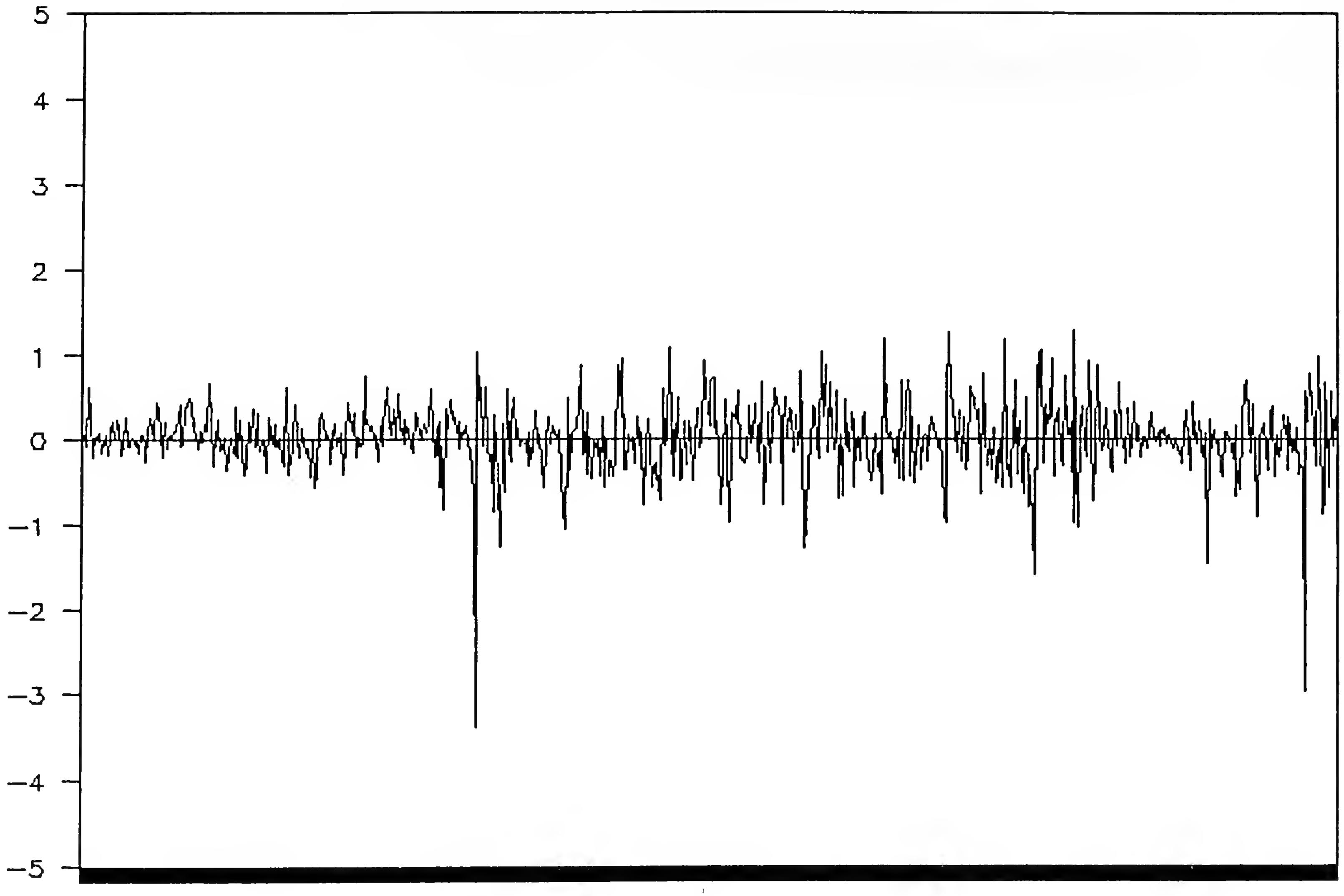

Figure 2. ARCH residuals

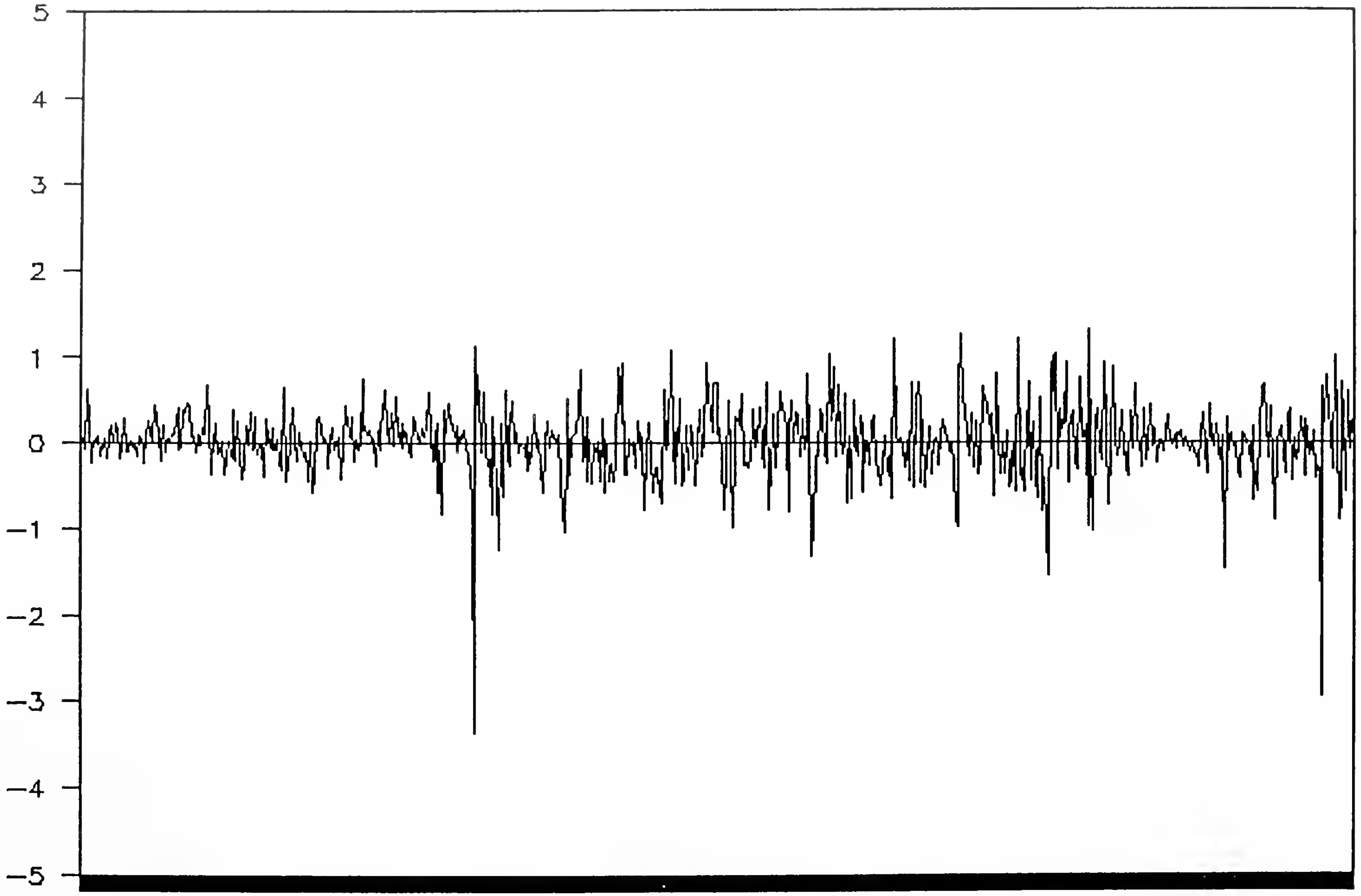

Figure 3. NARCH residuals

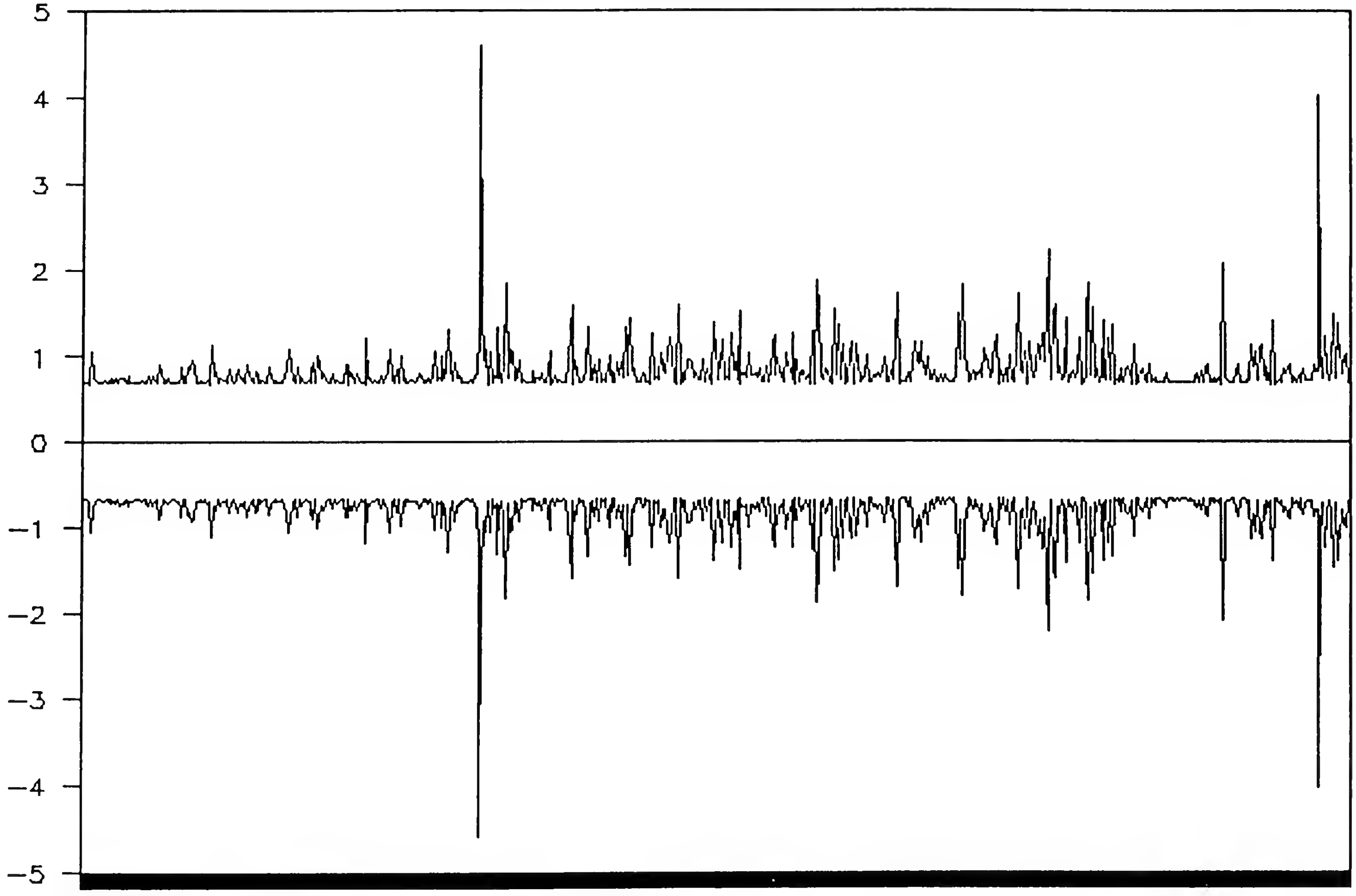

Figure 4. 95% confidence intervals for ARCH residuals

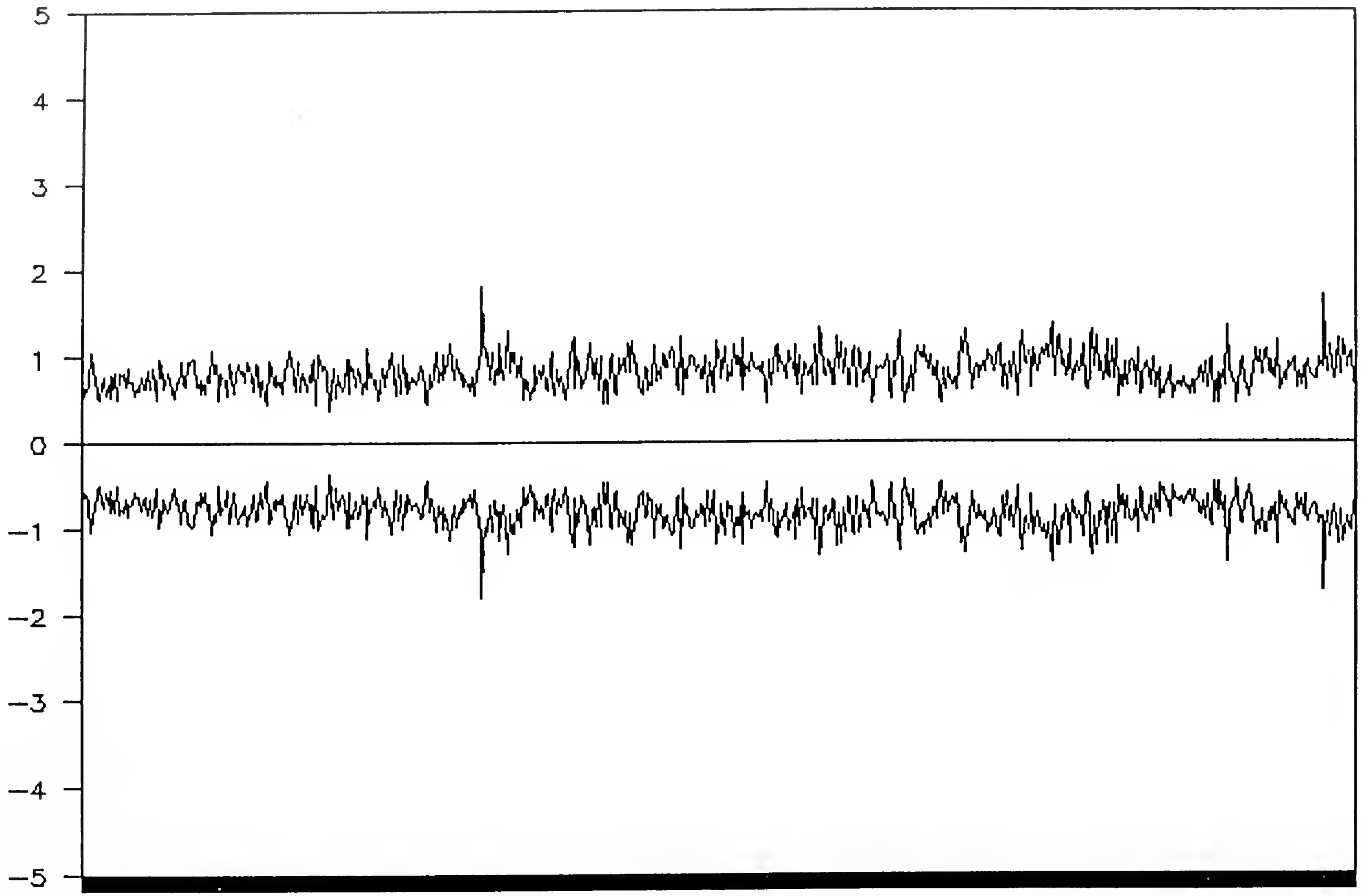

Figure 5. 95% confidence intervals for NARCH residuals