Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation
Author(s): Robert F. Engle
Source: *Econometrica*, Vol. 50, No. 4 (Jul., 1982), pp. 987-1007
Published by: Econometric Society
Stable URL: http://www.jstor.org/stable/1912773
Accessed: 26-01-2016 00:11 UTC

Econometrica, Vol. 50, No. 4 (July, 1982)

AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY WITH ESTIMATES OF THE VARIANCE OF UNITED KINGDOM INFLATION[1]

BY ROBERT F. ENGLE

Traditional econometric models assume a constant one-period forecast variance. To generalize this implausible assumption, a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes are introduced in this paper. These are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance.

A regression model is then introduced with disturbances following an ARCH process. Maximum likelihood estimators are described and a simple scoring iteration formulated. Ordinary least squares maintains its optimality properties in this set-up, but maximum likelihood is more efficient. The relative efficiency is calculated and can be infinite. To test whether the disturbances follow an ARCH process, the Lagrange multiplier procedure is employed. The test is based simply on the autocorrelation of the squared OLS residuals.

This model is used to estimate the means and variances of inflation in the U.K. The ARCH effect is found to be significant and the estimated variances increase substantially during the chaotic seventies.

1. INTRODUCTION

IF A RANDOM VARIABLE y_t is drawn from the conditional density function $f(y_t \mid y_{t-1})$, the forecast of today's value based upon the past information, under standard assumptions, is simply $E(y_t \mid y_{t-1})$, which depends upon the value of the conditioning variable y_{t-1}. The variance of this one-period forecast is given by $V(y_t \mid y_{t-1})$. Such an expression recognizes that the conditional forecast variance depends upon past information and may therefore be a random variable. For conventional econometric models, however, the conditional variance does not depend upon y_{t-1}. This paper will propose a class of models where the variance does depend upon the past and will argue for their usefulness in economics. Estimation methods, tests for the presence of such models, and an empirical example will be presented.

Consider initially the first-order autoregression

$$y_t = \gamma y_{t-1} + \epsilon_t$$

where ϵ is white noise with $V(\epsilon) = \sigma^2$. The conditional mean of y_t is γy_{t-1} while the unconditional mean is zero. Clearly, the vast improvement in forecasts due to time-series models stems from the use of the conditional mean. The conditional

[1]This paper was written while the author was visiting the London School of Economics. He benefited greatly from many stimulating conversations with David Hendry and helpful suggestions by Denis Sargan and Andrew Harvey. Special thanks are due Frank Srba who carried out the computations. Further insightful comments are due to Clive Granger, Tom Rothenberg, Edmond Malinvaud, Jean-Francois Richard, Wayne Fuller, and two anonymous referees. The research was supported by NSF SOC 78-09476 and The International Centre for Economics and Related Disciplines. All errors remain the author's responsibility.

variance of y_t is σ^2 while the unconditional variance is $\sigma^2/1-\gamma^2$. For real processes one might expect better forecast intervals if additional information from the past were allowed to affect the forecast variance; a more general class of models seems desirable.

The standard approach of heteroscedasticity is to introduce an exogenous variable x_t which predicts the variance. With a known zero mean, the model might be

$$y_t = \epsilon_t x_{t-1}$$

where again $V(\epsilon) = \sigma^2$. The variance of y_t is simply $\sigma^2 x_{t-1}^2$ and, therefore, the forecast interval depends upon the evolution of an exogenous variable. This standard solution to the problem seems unsatisfactory, as it requires a specification of the causes of the changing variance, rather than recognizing that both conditional means and variances may jointly evolve over time. Perhaps because of this difficulty, heteroscedasticity corrections are rarely considered in time-series data.

A model which allows the conditional variance to depend on the past realization of the series is the bilinear model described by Granger and Andersen [**13**]. A simple case is

$$y_t = \epsilon_t y_{t-1}.$$

The conditional variance is now $\sigma^2 y_{t-1}^2$. However, the unconditional variance is either zero or infinity, which makes this an unattractive formulation, although slight generalizations avoid this problem.

A preferable model is

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

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2,$$

with $V(\epsilon_t) = 1$. This is an example of what will be called an autoregressive conditional heteroscedasticity (ARCH) model. It is not exactly a bilinear model, but is very close to one. Adding the assumption of normality, it can be more directly expressed in terms of ψ_t, the information set available at time t. Using conditional densities,

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

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2. \tag{2}$$

The variance function can be expressed more generally as

$$h_t = h(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, \alpha) \tag{3}$$

where p is the order of the ARCH process and α is a vector of unknown parameters.

The ARCH regression model is obtained by assuming that the mean of y_t is given as $x_t\beta$, a linear combination of lagged endogenous and exogenous variables included in the information set ψ_{t-1} with β a vector of unknown parameters. Formally,

$$\begin{aligned} & y_t \mid \psi_{t-1} \sim N(x_t\beta, h_t), \\ (4) \qquad & h_t = h(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p}, \alpha), \\ & \epsilon_t = y_t - x_t\beta. \end{aligned}$$

The variance function can be further generalized to include current and lagged x's as these also enter the information set. The h function then becomes

$$(5) \qquad h_t = h(\epsilon_{t-1}, \ldots, \epsilon_{t-p}, x_t, x_{t-1}, \ldots, x_{t-p}, \alpha)$$

or simply

$$h_t = h(\psi_{t-1}, \alpha).$$

This generalization will not be treated in this paper, but represents a simple extension of the results. In particular, if the h function factors into

$$h_t = h_\epsilon(\epsilon_{t-1}, \ldots, \epsilon_{t-p}, \alpha) h_x(x_t, \ldots, x_{t-p}),$$

the two types of heteroscedasticity can be dealt with sequentially by first correcting for the x component and then fitting the ARCH model on the transformed data.

The ARCH regression model in (4) has a variety of characteristics which make it attractive for econometric applications. Econometric forecasters have found that their ability to predict the future varies from one period to another. McNees [**17**, p. 52] suggests that, "the inherent uncertainty or randomness associated with different forecast periods seems to vary widely over time." He also documents that, "large and small errors tend to cluster together (in contiguous time periods)." This analysis immediately suggests the usefulness of the ARCH model where the underlying forecast variance may change over time and is predicted by past forecast errors. The results presented by McNees also show some serial correlation during the episodes of large variance.

A second example is found in monetary theory and the theory of finance. By the simplest assumptions, portfolios of financial assets are held as functions of the expected means and variances of the rates of return. Any shifts in asset demand must be associated with changes in expected means and variances of the rates of return. If the mean is assumed to follow a standard regression or time-series model, the variance is immediately constrained to be constant over time. The use of an exogenous variable to explain changes in variance is usually not appropriate.

A third interpretation is that the ARCH regression model is an approximation to a more complex regression which has non-ARCH disturbances. The ARCH specification might then be picking up the effect of variables omitted from the estimated model. The existence of an ARCH effect would be interpreted as evidence of misspecification, either by omitted variables or through structural change. If this is the case, ARCH may be a better approximation to reality than making standard assumptions about the disturbances, but trying to find the omitted variable or determine the nature of the structural change would be even better.

Empirical work using time-series data frequently adopts *ad hoc* methods to measure (and allow) shifts in the variance over time. For example, Klein **[15]** obtains estimates of variance by constructing the five-period moving variance about the ten-period moving mean of annual inflation rates. Others, such as Khan **[14]**, resort to the notion of "variability" rather than variance, and use the absolute value of the first difference of the inflation rate. Engle **[10]** compares these with the ARCH estimates for U.S. data.

2. THE LIKELIHOOD FUNCTION

Suppose y_t is generated by an ARCH process described in equations (1) and (3). The properties of this process can easily be determined by repeated application of the relation $Ex = E(Ex \mid \psi))$. The mean of y_t is zero and all autocovariances are zero. The unconditional variance is given by $\sigma_t^2 = Ey_t^2 = Eh_t$. For many functions h and values of α, the variance is independent of t. Under such conditions, y_t is covariance stationary; a set of sufficient conditions for this is derived below.

Although the process defined by (1) and (3) has all observations conditionally normally distributed, the vector of y is not jointly normally distributed. The joint density is the product of all the conditional densities and, therefore, the log likelihood is the sum of the conditional normal log likelihoods corresponding to (1) and (3). Let l be the average log likelihood and l_t be the log likelihood of the tth observation and T the sample size. Then

$$l = \frac{1}{T}\sum_{t=1}^{T} l_t, \qquad l_t = -\tfrac{1}{2}\log h_t - \tfrac{1}{2} y_t^2/h_t, \tag{6}$$

apart from some constants in the likelihood.

To estimate the unknown parameters α, this likelihood function can be maximized. The first-order conditions are

$$\frac{\partial l_t}{\partial \alpha} = \frac{1}{2h_t}\frac{\partial h_t}{\partial \alpha}\left(\frac{y_t^2}{h_t} - 1\right) \tag{7}$$

and the Hessian is

(8) $$\frac{\partial^2 l_t}{\partial\alpha\partial\alpha'} = -\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\alpha}\frac{\partial h_t}{\partial\alpha'}\left(\frac{y_t^2}{h_t}\right) + \left[\frac{y_t^2}{h_t} - 1\right]\frac{\partial}{\partial\alpha'}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial\alpha}\right].$$

The conditional expectation of the second term, given ψ_{t-m-1}, is zero, and of the last factor in the first, is just one. Hence, the information matrix, which is simply the negative expectation of the Hessian averaged over all observations, becomes

(9) $$\mathcal{I}_{\alpha\alpha} = \sum_t \frac{1}{2T} E\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial\alpha'}\frac{\partial h_t}{\partial\alpha'}\right]$$

which is consistently estimated by

(10) $$\hat{\mathcal{I}}_{\alpha\alpha} = \frac{1}{T}\sum_t\left[\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\alpha}\frac{\partial h_t}{\partial\alpha'}\right].$$

If the h function is pth order linear (in the squares), so that it can be written as

(11) $$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_p y_{t-p}^2,$$

then the information matrix and gradient have a particularly simple form. Let $z_t = (1, y_{t-1}^2, \ldots, y_{t-p}^2)$ and $\alpha' = (\alpha_0, \alpha_1, \cdots, \alpha_p)$ so that (11) can be rewritten as

(12) $$h_t = z_t\alpha.$$

The gradient then becomes simply

(13) $$\frac{\partial l}{\partial\alpha} = \frac{1}{2h_t} z_t\left(\frac{y_t^2}{h_t} - 1\right)$$

and the estimate of the information matrix

(14) $$\hat{\mathcal{I}}_{\alpha\alpha} = \frac{1}{2T}\sum_t (z_t' z_t / h_t^2).$$

3. DISTRIBUTION OF THE FIRST-ORDER LINEAR ARCH PROCESS

The simplest and often very useful ARCH model is the first-order linear model given by (1) and (2). A large observation for y will lead to a large variance for the next period's distribution, but the memory is confined to one period. If $\alpha_1 = 0$, of course y will be Gaussian white noise and if it is a positive number, successive observations will be dependent through higher-order moments. As shown below, if α_1 is too large, the variance of the process will be infinite.

To determine the conditions for the process to be stationary and to find the marginal distribution of the y's, a recursive argument is required. The odd moments are immediately seen to be zero by symmetry and the even moments are computed using the following theorem. In all cases it is assumed that the process begins indefinitely far in the past with $2r$ finite initial moments.

THEOREM 1: *For integer r, the $2r$th moment of a first-order linear ARCH process with $\alpha_0 > 0$, $\alpha_1 \geq 0$, exists if, and only if,*

$$\alpha_1^r \prod_{j=1}^{r} (2j-1) < 1.$$

A constructive expression for the moments is given in the proof.

PROOF: See Appendix.

The theorem is easily used to find the second and fourth moments of a first-order process. Letting $w_t = (y_t^4, y_t^2)'$,

$$E(w_t \mid \psi_{t-1}) = \begin{pmatrix} 3\alpha_0^2 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} 3\alpha_1^2 & 6\alpha_0\alpha_1 \\ 0 & \alpha_1 \end{pmatrix} w_{t-1}.$$

The condition for the variance to be finite is simply that $\alpha_1 < 1$, while to have a finite fourth moment it is also required that $3\alpha_1^2 < 1$. If these conditions are met, the moments can be computed from (A4) as

$$E(w_t) = \begin{bmatrix} \left[\dfrac{3\alpha_0^2}{(1-\alpha_1)^2}\right]\left[\dfrac{1-\alpha_1^2}{1-3\alpha_1^2}\right] \\ \dfrac{\alpha_0}{1-\alpha_1} \end{bmatrix}. \tag{15}$$

The lower element is the unconditional variance, while the upper product gives the fourth moment. The first expression in square brackets is three times the squared variance. For $\alpha_1 \neq 0$, the second term is strictly greater than one implying a fourth moment greater than that of a normal random variable.

The first-order ARCH process generates data with fatter tails than the normal density. Many statistical procedures have been designed to be robust to large errors, but to the author's knowledge, none of this literature has made use of the fact that temporal clustering of outliers can be used to predict their occurrence and minimize their effects. This is exactly the approach taken by the ARCH model.

4. GENERAL ARCH PROCESSES

The conditions for a first-order linear ARCH process to have a finite variance and, therefore, to be covariance stationary can directly be generalized for pth-order processes.

THEOREM 2: *The pth-order linear ARCH processes, with* $\alpha_0 > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$, *is covariance stationary if, and only if, the associated characteristic equation has all roots outside the unit circle. The stationary variance is given by* $E(y_t^2) = \alpha_0/(1 - \sum_{j=1}^{p} \alpha_j)$.

PROOF: See Appendix.

Although the pth-order linear model is a convenient specification, it is likely that other formulations of the variance model may be more appropriate for particular applications. Two simple alternatives are the exponential and absolute value forms:

$$h_t = \exp(\alpha_0 + \alpha_1 y_{t-1}^2), \tag{16}$$

$$h_t = \alpha_0 + \alpha_1 |y_{t-1}|. \tag{17}$$

These provide an interesting contrast. The exponential form has the advantage that the variance is positive for all values of alpha, but it is not difficult to show that data generated from such a model have infinite variance for any value of $\alpha_1 \neq 0$. The implications of this deserve further study. The absolute value form requires both parameters to be positive, but can be shown to have finite variance for any parameter values.

In order to find estimation results which are more general than the linear model, general conditions on the variance model will be formulated and shown to be implied for the linear process.

Let ξ_t be a $p \times 1$ random vector drawn from the sample space Ξ, which has elements $\xi_t' = (\xi_{t-1}, \ldots, \xi_{t-p})$. For any ξ_t, let ξ_t^* be identical, except that the mth element has been multiplied by -1, where m lies between 1 and p.

DEFINITION: The ARCH process defined by (1) and (3) is *symmetric* if

(a) $\quad h(\xi_t) = h(\xi_t^*) \qquad$ for any m and $\xi_t \epsilon \Xi$,

(b) $\quad \partial h(\xi_t)/\partial \alpha_i = \partial h(\xi_t^*)/\partial \alpha_i \qquad$ for any m, i and $\xi_t \epsilon \Xi$,

(c) $\quad \partial h(\xi_t)/\partial \xi_{t-m} = -\partial h(\xi_t^*)/\partial \xi_{t-m} \qquad$ for any m and $\xi_t \epsilon \Xi$.

All the functions described have been symmetric. This condition is the main distinction between mean and variance models.

Another characterization of general ARCH models is in terms of regularity conditions.

DEFINITION: The ARCH model defined by (1) and (3) is *regular* if

(a) $\quad \min h(\xi_t) \geq \delta \qquad$ for some $\delta > 0$ and $\xi_t \epsilon \Xi$,

(b) $\quad E(|\partial h(\xi_t)/\partial \alpha_i| |\partial h(\xi_t)/\partial \xi_{t-m}| \,|\psi_{t-m-1}) \qquad$ exists for all i, m, t.

The first portion of the definition is very important and easy to check, as it requires the variance always to be positive. This eliminates, for example, the log-log autoregression. The second portion is difficult to check in some cases, yet should generally be true if the process is stationary with bounded derivatives, since conditional expectations are finite if unconditional ones are. Condition (b) is a sufficient condition for the existence of some expectations of the Hessian used in Theorem 4. Presumably weaker conditions could be found.

THEOREM 3: *The pth-order linear ARCH model satisfies the regularity conditions, if* $\alpha_0 > 0$ *and* $\alpha_1, \ldots, \alpha_p \geq 0$.

PROOF: See Appendix.

In the estimation portion of the paper, a very substantial simplification results if the ARCH process is symmetric and regular.

5. ARCH REGRESSION MODELS

If the ARCH random variables discussed thus far have a non-zero mean, which can be expressed as a linear combination of exogenous and lagged dependent variables, then a regression framework is appropriate, and the model can be written as in (4) or (5). An alternative interpretation for the model is that the disturbances in a linear regression follow an ARCH process.

In the pth-order linear case, the specification and likelihood are given by

$$\begin{aligned} & y_t \mid \psi_{t-1} \sim N(x_t\beta, h_t), \\ & h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_p\epsilon_{t-p}^2, \\ (18)\qquad & \epsilon_t = y_t - x_t\beta, \\ & l = \frac{1}{T}\sum_{t=1}^{T} l_t, \\ & l_t = -\tfrac{1}{2}\log h_t - \tfrac{1}{2}\epsilon_t^2/h_t, \end{aligned}$$

where x_t may include lagged dependent and exogenous variables and an irrelevant constant has been omitted from the likelihood. This likelihood function can be maximized with respect to the unknown parameters α and β. Attractive methods for computing such an estimate and its properties are discussed below.

Under the assumptions in (18), the ordinary least squares estimator of β is still consistent as x and ϵ are uncorrelated through the definition of the regression as a conditional expectation. If the x's can be treated as fixed constants then the least squares standard errors will be correct; however, if there are lagged dependent variables in x_t, the standard errors as conventionally computed will not be consistent, since the squares of the disturbances will be correlated with squares of the x's. This is an extension of White's **[18]** argument on heteroscedasticity and it suggests that using his alternative form for the covariance matrix would give a consistent estimate of the least-squares standard errors.

If the regressors include no lagged dependent variables and the process is stationary, then letting y and x be the $T\times 1$ and $T\times K$ vector and matrix of dependent and independent variables, respectively,

$$E(y\mid x)=x\beta,\qquad \operatorname{Var}(y\mid x)=\sigma^2 I, \tag{19}$$

and the Gauss–Markov assumptions are statisfied. Ordinary least squares is the best linear unbiased estimator for the model in (18) and the variance estimates are unbiased and consistent. However, maximum likelihood is different and consequently asymptotically superior; ordinary least squares does not achieve the Cramer–Rao bound. The maximum-likelihood estimator is nonlinear and is more efficient than OLS by an amount calculated in Section 6.

The maximum likelihood estimator is found by solving the first order conditions. The derivative with respect to β is

$$\frac{\partial l_t}{\partial\beta}=\frac{\epsilon_t x_t'}{h_t}+\frac{1}{2h_t}\frac{\partial h_t}{\partial\beta}\left(\frac{\epsilon_t^2}{h_t}-1\right). \tag{20}$$

The first term is the familiar first-order condition for an exogenous heteroscedastic correction; the second term results because h_t is also a function of the β's, as in Amemiya **[1]**. Substituting the linear variance function gives

$$\frac{\partial l}{\partial\beta}=\frac{1}{T}\sum\left[\frac{\epsilon_t x_t'}{h_t}-\frac{1}{h_t}\left(\frac{\epsilon_t^2}{h_t}-1\right)\sum_j\alpha_j\epsilon_{t-j}x_{t-j}'\right], \tag{21}$$

which can be rewritten approximately by collecting terms in x and ϵ as

$$\begin{aligned}\frac{\partial l}{\partial\beta}&=\frac{1}{T}\sum_t x_t'\epsilon_t\left[h_t^{-1}-\sum_{j=1}^{p}\alpha_j h_{t+j}^{-2}\left(\epsilon_{t+j}^2-h_{t+j}\right)\right]\\&\equiv\frac{1}{T}\sum_t x_t'\epsilon_t s_t.\end{aligned} \tag{22}$$

The Hessian is

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\beta\partial\beta'}=&-\frac{x_t'x_t}{h_t}-\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\beta}\frac{\partial h_t}{\partial\beta'}\left(\frac{\epsilon_t^2}{h_t}\right)\\&-\frac{2\epsilon_t x_t'}{h_t^2}\frac{\partial h_t}{\partial\beta}+\left(\frac{\epsilon_t^2}{h_t}-1\right)\frac{\partial}{\partial\beta'}\left[\frac{1}{2h_t}\frac{\partial h_t}{\partial\beta}\right].\end{aligned}$$

Taking conditional expectations of the Hessian, the last two terms vanish because h_t is entirely a function of the past. Similarly, ϵ_t^2/h_t becomes one, since it is the only current value in the second term. Notice that these results hold regardless of whether x_t includes lagged-dependent variables. The information matrix is the average over all t of the expected value of the conditional expectation and is, therefore, given by

$$\text{(23)} \qquad \mathcal{I}_{\beta\beta} = \frac{1}{T}\sum_t E\left[E\left(\frac{\partial^2 l_t}{\partial\beta\partial\beta'} \mid \psi_{t-1}\right)\right]$$

$$= \frac{1}{T}\sum_t E\left[\frac{x_t' x_t}{h_t} + \frac{1}{2h_t^2}\frac{\partial h_t}{\partial\beta}\frac{\partial h_t}{\partial\beta'}\right].$$

For the pth order linear ARCH regression this is consistently estimated by

$$\text{(24)} \qquad \hat{\mathcal{I}}_{\beta\beta} = \frac{1}{T}\sum\left[\frac{x_t' x_t}{h_t} + 2\sum_j \alpha_j^2 \frac{\epsilon_{t-j}^2}{h_t^2} x_{t-j}' x_{t-j}\right].$$

By gathering terms in $x_t' x_t$, (24) can be rewritten, except for end effects, as

$$\text{(25)} \qquad \hat{\mathcal{I}}_{\beta\beta} = \frac{1}{T}\sum_t x_t' x_t \left[h_t^{-1} + 2\epsilon_t^2 \sum_{j=1}^{p} \alpha_j^2 h_{t+j}^{-2}\right]$$

$$\equiv \frac{1}{T}\sum_t x_t' x_t r_t^2 .$$

In a similar fashion, the off-diagonal blocks of the information matrix can be expressed as:

$$\text{(26)} \qquad \mathcal{I}_{\alpha\beta} = \frac{1}{T}\sum_t E\left(\frac{1}{2h_t^2}\frac{\partial h_t}{\partial\alpha}\frac{\partial h_t}{\partial\beta'}\right).$$

The important result to be shown in Theorem 4 below is that this off-diagonal block is zero. The implications are far-reaching in that estimation of α and β can be undertaken separately without asymptotic loss of efficiency and their variances can be calculated separately.

THEOREM 4: *If an ARCH regression model is symmetric and regular, then* $\mathcal{I}_{\alpha\beta} = 0$.

PROOF: See Appendix.

6. ESTIMATION OF THE ARCH REGRESSION MODEL

Because of the block diagonality of the information matrix, the estimation of α and β can be considered separately without loss of asymptotic efficiency.

Furthermore, either can be estimated with full efficiency based only on a consistent estimate of the other. See, for example, Cox and Hinkley [**6**, p. 308]. The procedure recommended here is to initially estimate β by ordinary least squares, and obtain the residuals. From these residuals, an efficient estimate of α can be constructed, and based upon these $\hat{\alpha}$ estimates, efficient estimates of β are found. The iterations are calculated using the scoring algorithm. Each step for a parameter vector ϕ produces estimates ϕ^{i+1} based on ϕ^i according to

$$\phi^{i+1} = \phi^i + \left[\hat{\mathscr{I}}^i_{\phi\phi}\right]^{-1} \frac{1}{T} \sum_t \frac{\partial l_t^i}{\partial \phi}, \tag{27}$$

where $\hat{\mathscr{I}}^i$ and $\partial l_t^i/\partial\phi$ are evaluated at ϕ^i. The advantage of this algorithm is partly that it requires only first derivatives of the likelihood function in this case and partly that it uses the statistical properties of the problem to tailor the algorithm to this application.

For the pth-order linear model, the scoring step for α can be rewritten by substituting (12), (13), and (14) into (27) and interpreting y_t^2 as the residuals e_t^2. The iteration is simply

$$\alpha^{i+1} = \alpha^i + (\tilde{z}'\tilde{z})^{-1}\tilde{z}'f^i \tag{28}$$

where

$$\tilde{z}_t = \left(1, e_{t-1}^2, \ldots, e_{t-p}^2\right)/h_t^i,$$

$$\tilde{z}' = (\tilde{z}_1', \ldots, \tilde{z}_T'),$$

$$f_t^i = \left(e_t^2 - h_t^i\right)/h_t^i,$$

$$f^{i\prime} = \left(f_1^i, \ldots, f_T^i\right).$$

In these expressions, e_t is the residual from iteration i, h_t^i is the estimated conditional variance, and α^i is the estimate of the vector of unknown parameters from iteration i. Each step is, therefore, easily constructed from a least-squares regression on transformed variables. The variance-covariance matrix of the parameters is consistently estimated by the inverse of the estimate of the information matrix divided by T, which is simply $2(\tilde{z}'\tilde{z})^{-1}$. This differs slightly from $\hat{\sigma}^2(\tilde{z}'\tilde{z})^{-1}$ computed by the auxiliary regression. Asymptotically, $\hat{\sigma}^2 = 2$, if the distributional assumptions are correct, but it is not clear which formulation is better in practice.

The parameters in α must satisfy some nonnegativity conditions and some stationarity conditions. These could be imposed via penalty functions or the parameters could be estimated and checked for conformity. The latter approach is used here, although a perhaps useful reformulation of the model might employ squares to impose the nonnegativity constraints directly:

$$h_t = \alpha_0^2 + \alpha_1^2\epsilon_{t-1}^2 + \cdots + \alpha_p^2\epsilon_{t-p}^2. \tag{29}$$

Convergence for such an iteration can be formulated in many ways. Following Belsley [3], a simple criterion is the gradient around the inverse Hessian. For a parameter vector, ϕ, this is

$$\theta = \frac{\partial l'}{\partial \phi}\left(\frac{\partial^2 l}{\partial \phi \partial \phi'}\right)^{-1}\frac{\partial l}{\partial \phi}. \tag{30}$$

Using θ as the convergence criterion is attractive, as it provides a natural normalization and as it is interpretable as the remainder term in a Taylor-series expansion about the estimated maximum. In any case, substituting the gradient and estimated information matrix in (30), $\theta = R^2$ of the auxiliary regression.

For a given estimate of α, a scoring step can be computed to improve the estimate of beta. The scoring algorithm for β is

$$\beta^{i+1} = \beta^i + \left[\hat{\mathscr{I}}_{\beta\beta}\right]^{-1}\frac{\partial l^i}{\partial \beta}. \tag{31}$$

Defining $\tilde{x}_t = x_t r_t$ and $\tilde{e}_t = e_t s_t / r_t$ with $\tilde{x}$ and $\tilde{e}$ as the corresponding matrix and vector, (31) can be rewritten using (22) and (24) and e_t for the estimate of ϵ_t on the ith iteration, as

$$\beta^{i+1} = \beta^i + (\tilde{x}'\tilde{x})^{-1}\tilde{x}'\tilde{e}. \tag{32}$$

Thus, an ordinary least-squares program can again perform the scoring iteration, and $(\tilde{x}'\tilde{x})^{-1}$ from this calculation will be the final variance-covariance matrix of the maximum likelihood estimates of β.

Under the conditions of Crowder's [7] theorem for martingales, it can be established that the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ are asymptotically normally distributed with limiting distribution

$$\begin{aligned} &\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{D} N\left(0, \mathscr{I}_{\alpha\alpha}^{-1}\right), \\ &\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{D} N\left(0, \mathscr{I}_{\beta\beta}^{-1}\right). \end{aligned} \tag{33}$$

7. GAINS IN EFFICIENCY FROM MAXIMUM LIKELIHOOD ESTIMATION

The gain in efficiency from using the maximum-likelihood estimation rather than OLS has been asserted above. In this section, the gains are calculated for a special case. Consider the linear stationary ARCH model with $p = 1$ and all x_t exogenous. This is the case where the Gauss–Markov theorem applies and OLS has a variance matrix $\sigma^2(x'x)^{-1} = E\epsilon_t^2(\sum_t x_t' x_t)^{-1}$. The stationary variance is $\sigma^2 = \alpha_0/(1 - \alpha_1)$.

The information matrix for this case becomes, from (25),

$$E\left[\sum_t x_t' x_t\left(h_t^{-1} + 2\epsilon_t^2\alpha_1^2/h_{t+1}^2\right)\right].$$

With x exogenous, the expectation is only necessary over the scale factor. Because the disturbance process is stationary, the variance-covariance matrix is proportional to that for OLS and the relative efficiency depends only upon the scale factors. The relative efficiency of MLE to OLS is, therefore,

$$R = E\left(h_t^{-1} + 2\epsilon_t^2\alpha_1^2/h_{t+1}^2\right)\sigma^2.$$

Now substitute $h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2$, $\sigma^2 = \alpha_0/1 - \alpha_1$, and $\gamma = \alpha_1/1 - \alpha_1$. Recognizing that ϵ_{t-1}^2 and ϵ_t^2 have the same density, define for each

$$u = \epsilon\sqrt{(1 - \alpha_1)/\alpha_0}\ .$$

The expression for the relative efficiency becomes

$$R = E\left(\frac{1+\gamma}{1+\gamma u^2}\right) + 2\gamma^2 E\frac{u^2}{\left(1+\gamma u^2\right)^2}\ , \tag{34}$$

where u has variance one and mean zero. From Jensen's inequality, the expected value of a reciprocal exceeds the reciprocal of the expected value and, therefore, the first term is greater than unity. The second is positive, so there is a gain in efficiency whenever $\gamma \neq 0$. Eu^{-2} is infinite because u^2 is conditionally chi squared with one degree of freedom. Thus, the limit of the relative efficiency goes to infinity with γ:

$$\lim_{\gamma\to\infty} R \to \infty.$$

For α_1 close to unity, the gain in efficiency from using a maximum likelihood estimator may be very large.

8. TESTING FOR ARCH DISTURBANCES

In the linear regression model, with or without lagged-dependent variables, OLS is the appropriate procedure if the disturbances are not conditionally heteroscedastic. Because the ARCH model requires iterative procedures, it may be desirable to test whether it is appropriate before going to the effort to estimate it. The Lagrange multiplier test procedure is ideal for this as in many similar cases. See, for example, Breusch and Pagan [**4**, **5**], Godfrey [**12**], and Engle [**9**].

Under the null hypothesis, $\alpha_1 = \alpha_2 \cdots = \alpha_p = 0$. The test is based upon the score under the null and the information matrix under the null. Consider the ARCH model with $h_t = h(z_t\alpha)$, where h is some differentiable function which, therefore, includes both the linear and exponential cases as well as lots of others and $z_t = (1, e_{t-1}^2, \ldots, e_{t-p}^2)$ where e_t are the ordinary least squares residuals. Under the null, h_t is a constant denoted h^0. Writing $\partial h_t/\partial\alpha = h'z_t'$, where h' is the scalar derivative of h, the score and information can be written as

$$\left.\frac{\partial l}{\partial \alpha}\right|_0 = \frac{h'}{2h^0}\sum_t z_t'\left(\frac{e_t^2}{h^0}-1\right) = \frac{h^{0\prime}}{2h^0} z' f^0,$$

$$\mathcal{I}_{\alpha\alpha}^0 = \frac{1}{2}\left(\frac{h^{0\prime}}{h^0}\right)^2 E z' z,$$

and, therefore, the LM test statistic can be consistently estimated by

$$\text{(35)} \qquad \xi^* = \tfrac{1}{2} f^{0\prime} z (z'z)^{-1} z' f^0$$

where $z' = (z_1', \ldots, z_T')$, f^0 is the column vector of

$$\left(\frac{e_t^2}{h^0}-1\right).$$

This is the form used by Breusch and Pagan **[4]** and Godfrey **[12]** for testing for heteroscedasticity. As they point out, all reference to the h function has disappeared and, thus, the test is the same for any h which is a function only of $z_t\alpha$.

In this problem, the expectation required in the information matrix could be evaluated quite simply under the null; this could have superior finite sample performance. A second simplification, which is appropriate for this model as well as the heteroscedasticity model, is to note that plim $f^{0\prime}f^0/T = 2$ because normality has already been assumed. Thus, an asymptotically equivalent statistic would be

$$\text{(36)} \qquad \xi = T f^{0\prime} z (z'z)^{-1} z' f^0 / f^{0\prime} f^0 = TR^2$$

where R^2 is the squared multiple correlation between f^0 and z. Since adding a constant and multiplying by a scalar will not change the R^2 of a regression, this is also the R^2 of the regression of e_t^2 on an intercept and p lagged values of e_t^2. The statistic will be asymptotically distributed as chi square with p degrees of freedom when the null hypothesis is true.

The test procedure is to run the OLS regression and save the residuals. Regress the squared residuals on a constant and p lags and test TR^2 as a χ_p^2. This will be an asymptotically locally most powerful test, a characterization it shares with likelihood ratio and Wald tests. The same test has been proposed by Granger and Anderson **[13]** to test for higher moments in bilinear time series.

9. ESTIMATION OF THE VARIANCE OF INFLATION

Economic theory frequently suggests that economic agents respond not only to the mean, but also to higher moments of economic random variables. In financial theory, the variance as well as the mean of the rate of return are determinants of portfolio decisions. In macroeconomics, Lucas **[16]**, for example, argues that the variance of inflation is a determinant of the response to various shocks. Furthermore, the variance of inflation may be of independent interest as it is the unanticipated component which is responsible for the bulk of the welfare loss due to inflation. Friedman **[11]** also argues that, as high inflation will generally be associated with high variability of inflation, the statistical relationship between inflation and unemployment should have a positive slope, not a negative one as in the traditional Phillips curve.

Measuring the variance of inflation over time has presented problems to various researchers. Khan **[14]** has used the absolute value of the first difference of inflation while Klein **[15]** has used a moving variance around a moving mean. Each of these approaches makes very simple assumptions about the mean of the distribution, which are inconsistent with conventional econometric approaches. The ARCH method allows a conventional regression specification for the mean function, with a variance which is permitted to change stochastically over the sample period. For a comparison of several measures for U.S. data, see Engle **[10]**.

A conventional price equation was estimated using British data from 1958-II through 1977-II. It was assumed that price inflation followed wage increases; thus the model is a restricted transfer function.

Letting $\dot{p}$ be the first difference of the log of the quarterly consumer price index and w be the log of the quarterly index of manual wage rates, the model chosen after some experimentation was

$$\dot{p} = \beta_1 \dot{p}_{-1} + \beta_2 \dot{p}_{-4} + \beta_3 \dot{p}_{-5} + \beta_4 (p - w)_{-1} + \beta_5. \tag{37}$$

The model has typical seasonal behavior with the first, fourth, and fifth lags of the first difference. The lagged value of the real wage is the error correction mechanism of Davidson, et al. **[8]**, which restricts the lag weights to give a constant real wage in the long run. As this is a reduced form, the current wage rate cannot enter.

The least squares estimates of this model are given in Table I. The fit is quite good, with less than 1 per cent standard error of forecast, and all t statistics greater than 3. Notice that $\dot{p}_{-4}$ and $\dot{p}_{-5}$ have equal and opposite signs, suggesting that it is the acceleration of inflation one year ago which explains much of the short-run behavior in prices.

TABLE I

ORDINARY LEAST SQUARES (36)[a]

Variable	$\dot{p}_{-1}$	$\dot{p}_{-4}$	$\dot{p}_{-5}$	$(p-w)_{-1}$	Const.	$\alpha_0\ (\times 10^{-6})$	α_1
Coeff.	0.334	0.408	−0.404	−0.0559	0.0257	89	0
St. Err.	0.103	0.110	0.114	0.0136	0.00572		
t Stat.	3.25	3.72	3.55	4.12	4.49		

[a] Dependent variable $p = \log(P) - \log(P_{-1})$ where P is quarterly U.K. consumer price index. $w = \log(W)$ where W is the U.K. index of manual wage rates. Sample period 1958-II to 1977-II.

To establish the reliability of the model by conventional criteria, it was tested for serial correlation and for coefficient restrictions. Godfrey's **[12]** Lagrange multiplier test, for serial correlation up to sixth order, yields a chi-squared statistic with 6 degrees of freedom of 4.53, which is not significant, and the square of Durbin's h is 0.57. Only the 9th autocorrelation of the least squares residuals exceeds two asymptotic standard errors and, thus, the hypothesis of white noise disturbances can be accepted. The model was compared with an unrestricted regression, including all lagged p and w from one quarter through six. The asymptotic F statistic was 2.04, which is not significant at the 5 per cent level. When (37) was tested for the exclusion of w_{-1} through w_{-6}, the statistic was 2.34, which is barely significant at the 5 per cent but not the 2.5 per cent level. The only variable which enters significantly in either of these regressions is w_{-6} and it seems unattractive to include this alone.

The Lagrange multiplier test for a first-order linear ARCH effect for the model in (37) was not significant. However, testing for a fourth-order linear ARCH process, the chi-squared statistic with 4 degrees of freedom was 15.2, which is highly significant. Assuming that agents discount past residuals, a linearly declining set of weights was formulated to give the model

$$h_t = \alpha_0 + \alpha_1(0.4\epsilon_{t-1}^2 + 0.3\epsilon_{t-2}^2 + 0.2\epsilon_{t-3}^2 + 0.1\epsilon_{t-4}^2) \tag{38}$$

which is used in the balance of the paper. A two-parameter variance function was chosen because it was suspected that the nonnegativity and stationarity constraints on the α's would be hard to satisfy in an unrestricted model. The chi-squared test for $\alpha_1 = 0$ in (38) was 6.1, which has one degree of freedom.

One step of the scoring algorithm was employed to estimate model (37) and (38). The scoring step on α was performed first and then, using the new efficient $\hat{\alpha}$, the algorithm obtains in one step, efficient estimates of β. These are given in Table II. The procedure was also iterated to convergence by doing three steps on α, followed by three steps on β, followed by three more steps on α, and so forth. Convergence, within 0.1 per cent of the final value, occurred after two sets of α and β steps. These results are given in Table III.

The maximum likelihood estimates differ from the least squares effects primarily in decreasing the sizes of the short-run dynamic coefficients and increasing

TABLE II
MAXIMUM LIKELIHOOD ESTIMATES OF ARCH MODEL (36) (37)
ONE-STEP SCORING ESTIMATES[a]

Variable	$\dot{p}_{-1}$	$\dot{p}_{-4}$	$\dot{p}_{-5}$	$(p-w)_{-1}$	Const.	$\alpha_0\ (\times 10^{-6})$	α_1
Coeff.	0.210	0.270	− 0.334	− 0.0697	0.0321	19	0.846
St. Err.	0.110	0.094	0.109	0.0117	0.00498	14	0.243
t Stat.	1.90	2.86	3.06	5.98	6.44	1.32	3.49

[a] Dependent variable $p = \log(P) - \log(P_{-1})$ where P is quarterly U.K. consumer price index. $w = \log(W)$ where W is the U.K. index of manual wage rates. Sample period 1958-II to 1977-II.

TABLE III

MAXIMUM LIKELIHOOD ESTIMATES OF ARCH MODEL (36) (37)
ITERATED ESTIMATES[a]

Variables	$\dot{p}_{-1}$	$\dot{p}_{-4}$	$\dot{p}_{-5}$	$(p-w)_{-1}$	Const.	$\alpha_0\ (\times 10^{-6})$	α_1
Coeff.	0.162	0.264	−0.325	−0.0707	0.0328	14	0.955
St. Err.	0.108	0.0892	0.0987	0.0115	0.00491	8.5	0.298
t Stat.	1.50	2.96	3.29	6.17	6.67	1.56	3.20

[a] Dependent variable $\dot{p} = \log(P) - \log(P_{-1})$ where P is quarterly U.K. consumer price index. $w = \log(W)$ where W is the U.K. index of manual wage rates. Sample period 1958-II to 1977-II.

the coefficient on the long run, as incorporated in the error correction mechanism. The acceleration term is not so clearly implied as in the least squares estimates. These seem reasonable results, since much of the inflationary dynamics are estimated by a period of very severe inflation in the middle seventies. This, however, is also the period of the largest forecast errors and, hence, the maximum likelihood estimator will discount these observations. By the end of the sample period, inflationary levels were rather modest and one might expect that the maximum likelihood estimates would provide a better forecasting equation.

The standard errors for ordinary least squares are generally greater than for maximum likelihood. The least squares standard errors are 15 per cent to 25 per cent greater, with one exception where the standard error actually falls by 5 per cent to 7 per cent. As mentioned earlier, however, the least squares estimates are biased when there are lagged dependent variables. The Wald test for $\alpha_1 = 0$ is also significant.

The final estimates of h_t are the one-step-ahead forecast variances. For the one-step scoring estimator, these vary from 23×10^{-6} to 481×10^{-6}. That is, the forecast standard deviation ranges from 0.5 per cent to 2.2 per cent, which is more than a factor of 4. The average of the h_t, since 1974, is 230×10^{-6}, as compared with 42×10^{-6} during the last four years of the sixties. Thus, the standard deviation of inflation increased from 0.6 per cent to 1.5 per cent over a few years, as the economy moved from the rather predictable sixties into the chaotic seventies.

In order to determine whether the confidence intervals arising from the ARCH model were superior to the least squares model, the outliers were examined. The expected number of residuals exceeding two (conditional) standard deviations is 3.5. For ordinary least squares, there were 5 while ARCH produced 3. For least squares these occurred in '74-I, '75-I, '75-II, '75-IV, and '76-II; they all occur within three years of each other and, in fact, three of them are in the same year. For the ARCH model, they are much more spread out and only one of the least squares points remains an outlier, although the others are still large. Examining the observations exceeding one standard deviation shows similar effects. In the seventies, there were 13 OLS and 12 ARCH residuals outside one sigma, which are both above the expected value of 9. In the sixties, there were 6 for OLS, 10 for ARCH and an expected number of 12. Thus, the number of outliers for ordinary least squares is reasonable; however, the timing of their occurrence is far from random. The ARCH model comes closer to truly random residuals after standardizing for their conditional distributions.

This example illustrates the usefulness of the ARCH model for improving the performance of a least squares model and for obtaining more realistic forecast variances.

University of California, San Diego

Manuscript received July, 1979; final revision received July, 1981.

APPENDIX

PROOF OF THEOREM 1: Let

$$w_t' = (y_t^{2r}, y_t^{2(r-1)}, \ldots, y_t^2). \tag{A2}$$

First, it is shown that there is an upper triangular $r \times r$ matrix A and $r \times 1$ vector b such that

$$E(w_t \mid \psi_{t-1}) = b + Aw_{t-1}. \tag{A2}$$

For any zero-mean normal random variable u, with variance σ^2,

$$E(u^{2r}) = \sigma^{2r} \prod_{j=1}^{r} (2j-1).$$

Because the conditional distribution of y is normal

$$\begin{aligned} E(y_t^{2m} \mid \psi_{t-1}) &= h_t^{2m} \prod_{j-1}^{m} (2j-1) \\ &= (\alpha_1 y_{t-1}^2 + \alpha_0)^m \prod_{j=1}^{m} (2j-1). \end{aligned} \tag{A3}$$

Expanding this expression establishes that the moment is a linear combination of w_{t-1}. Furthermore, only powers of y less than or equal to $2m$ are required; therefore, A in (A2) is upper triangular.

Now

$$E(w_t \mid \psi_{t-2}) = b + A(b + Aw_{t-2})$$

or in general

$$E(w_t \mid \psi_{t-k}) = (\mathcal{I} + A + A^2 + \cdots + A^{k-1})b + A^k w_{t-k}.$$

Because the series starts indefinitely far in the past with $2r$ finite moments, the limit as k goes to infinity exists if, and only if, all the eigenvalues of A lie within the unit circle.

The limit can be written as

$$\lim_{k\to\infty} E(w_t \mid \psi_{t-k}) = (I - A)^{-1}b,$$

which does not depend upon the conditioning variables and does not depend upon t. Hence, this is an expression for the stationary moments of the unconditional distribution of y.

$$E(w_t) = (I - A)^{-1}b. \tag{A4}$$

It remains only to establish that the condition in the theorem is necessary and sufficient to have all eigenvalues lie within the unit circle. As the matrix has already been shown to be upper triangular, the diagonal elements are the eigenvalues. From (A3), it is seen that the diagonal elements are simply

$$\alpha_1^m \prod_{j=1}^{m} (2j-1) = \prod_{j=1}^{m} \alpha_1(2j-1) \equiv \theta_m$$

for $m = 1, \ldots, r$. If θ_r exceeds or equals unity, the eigenvalues do not lie in the unit circle. It must also be shown that if $\theta_r < 1$, then $\theta_m < 1$ for all $m < r$. Notice that θ_m is a product of m factors which are monotonically increasing. If the mth factor is greater than one, then θ_{m-1} will necessarily be smaller than θ_m. If the mth factor is less than one, all the other factors must also be less than one and, therefore, θ_{m-1} must also have all factors less than one and have a value less than one. This establishes that a necessary and sufficient condition for all diagonal elements to be less than one is that $\theta_r < 1$, which is the statement in the theorem. *Q.E.D.*

PROOF OF THEOREM 2: Let

$$w_t' = \left(y_t^2, y_{t-1}^2, \ldots, y_{t-p}^2\right).$$

Then in terms of the companion matrix A,

$$E(w_t \mid \psi_{t-1}) = b + Aw_{t-1} \tag{A5}$$

where $b' = (\alpha_0, 0, \ldots, 0)$ and

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_p & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Taking successive expectations

$$E(w_t \mid \psi_{t-k}) = (I + A + A^2 + \cdots + A^{k-1})b + A^k w_{t-k}.$$

Because the series starts indefinitely far in the past with finite variance, if, and only if, all eigenvalues lie within the unit circle, the limit exists and is given by

$$\lim_{k \to \infty} E(w_t \mid \psi_{t-k}) = (I - A)^{-1} b. \tag{A6}$$

As this does not depend upon initial conditions or on t, this vector is the common variance for all t. As is well known in time series analysis, this condition is equivalent to the condition that all the roots of the characteristic equation, formed from the α's, lie outside the unit circle. See Anderson [**2**, p. 177]. Finally, the limit of the first element can be rewritten as

$$Ey_t^2 = \alpha_0 \Big/ \left(1 - \sum_{j=1}^{p} \alpha_j\right). \tag{A7}$$

Q.E.D.

PROOF OF THEOREM 3: Clearly, under the conditions, $h(\xi_t) \geq \alpha_0 > 0$, establishing part (a). Let

$$\begin{aligned} \phi_{i,m,t} &= E(|\partial h(\xi_t)/\partial \alpha_i| \partial h(\xi_t)/\partial \xi_{t-m} \mid \psi_{t-m-1}) \\ &= 2\alpha_m E\left(|\xi_{t-i}|^2 |\xi_{t-m}| \psi_{t-m-1}\right). \end{aligned}$$

Now there are three cases; $i > m$, $i = m$, and $i < m$. If $i > m$, then $\xi_{t-i} \in \psi_{t-m-1}$ and the conditional expectation of $|\xi_{t-m}|$ is finite, because the conditional density is normal. If $i = m$, then the expectation becomes $E(|\xi_{t-m}|^3 \mid \psi_{t-m-1})$. Again, because the conditional density is normal, all moments exist including the expectation of the third power of the absolute value. If $i < m$, the expectation is taken in two parts, first with respect to $t - i - 1$:

$$\begin{aligned}\phi_{i,m,t} &= 2\alpha_m E\left\{|\xi_{t-m}| E(\xi^2_{t-i} \mid \psi_{t-i-1}) \mid \psi_{t-m-1}\right\}\\ &= 2\alpha_m E\left\{|\xi_{t-m}|\alpha_0 + \sum_{j=1}^{p} \alpha_j \xi^2_{t-i-j}\Big) \mid \psi_{t-m-1}\right\}\\ &= 2\alpha_m \alpha_0 E\{\xi_{t-m} \mid \psi_{t-m-1}\} + \sum_{j=1}^{p} \alpha_j \phi_{i+j,m,t}.\end{aligned}$$

In the final expression, the initial index on ϕ is larger and, therefore, may fall into either of the preceding cases, which, therefore, establishes the existence of the term. If there remain terms with $i + j < m$, the recursion can be repeated. As all lags are finite, an expression for $\phi_{i,m,t}$ can be written as a constant times the third absolute moment of ξ_{t-m} conditional on ψ_{t-m-1}, plus another constant times the first absolute moment. As these are both conditionally normal, and as the constants must be finite as they have a finite number of terms, the second part of the regularity condition has been established. *Q.E.D.*

To establish Theorem 4, a careful symmetry argument is required, beginning with the following lemma.

LEMMA: *Let u and v be any two random variables. $E(g(u,v) \mid v)$ will be an anti-symmetric function of v if g is anti-symmetric in v, the conditional density of $u \mid v$ is symmetric in v, and the expectation exists.*

PROOF:

$$\begin{aligned}E(g(u,-v) \mid -v) &= -E(g(u,v) \mid -v) && \text{because } g \text{ is anti-symmetric in } v\\ &= -E(g(u,v) \mid v) && \text{because the conditional density is symmetric.}\end{aligned}$$

Q.E.D.

PROOF OF THEOREM 4: The i, j element of $l_{\alpha\beta}$ is given by

$$\begin{aligned}(l_{\alpha\beta})_{ij} &= \frac{1}{2T}\sum_t E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \beta_j}\right)\\ &= -\frac{1}{2T}\sum_t \sum_{m=1}^{p} E\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} x_{jt-m}\right] \quad \text{by the chain rule.}\end{aligned}$$

If the expectation of the term in square brackets, conditional on ψ_{t-m-1}, is zero for all i, j, t, m, then the theorem is proven.

$$E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} x_{jt-m} \mid \psi_{t-m-1}\right) = x_{jt-m} E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} \mid \psi_{t-m-1}\right)$$

because x_{jt-m} is either exogenous or it is a lagged dependent variable, in which case it is included in ψ_{t-m-1}.

$$\begin{aligned}\left|E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \epsilon_{t-m}} \mid \psi_{t-m-1}\right)\right| &\le E\left(\frac{1}{h_t^2}\left|\frac{\partial h_t}{\partial \alpha_i}\right|\left|\frac{\partial h_t}{\partial \epsilon_{t-m}}\right| \mid \psi_{t-m-1}\right)\\ &\le \frac{1}{\delta^2} E\left(\left|\frac{\partial h_t}{\partial \alpha_i}\right|\left|\frac{\partial h_t}{\partial \epsilon_{t-m}}\right| \mid \psi_{t-m-1}\right)\end{aligned}$$

by part (a) of the regularity conditions and this integral is finite by part (b) of the condition. Hence, each term is finite. Now take the expectation in two steps, first with respect to ψ_{t-m}. This must therefore also be finite.

$$E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial \alpha_t}\frac{\partial h_t}{\partial \epsilon_{t-m}} \mid \psi_{t-m}\right) \equiv g(\epsilon_{t-m}).$$

By the symmetry assumption, h_t^{-1} is symmetric in ϵ_{t-m}, $\partial h_t/\partial \epsilon_{t-m}$ is anti-symmetric. Therefore, the whole expression is anti-symmetric in ϵ_{t-m}, which is part of the conditioning set ψ_{t-m}. Because h is symmetric, the conditional density must be symmetric in ϵ_{t-m} and the lemma can be invoked to show that $g(\epsilon_{t-m})$ is anti-symmetric.

Finally, taking expectations of g conditional on ψ_{t-m-1} gives zero, because the density of ϵ_{t-m} conditional on the past is a symmetric (normal) density and the theorem is established. *Q.E.D.*

REFERENCES

[1] AMEMIYA, T.: "Regression Analysis when the Variance of the Dependent Variable is Proportional to the Square of Its Expectation," *Journal of the American Statistical Association*, 68(1973), 928–934.

[2] ANDERSON, T. W.: *The Statistical Analysis of Time Series*. New York: John Wiley and Sons, 1958.

[3] BELSLEY, DAVID: "On the Efficient Computation of Non-Linear Full-Information Maximum Likelihood Estimator," paper presented to the European Meetings of the Econometric Society, Athens,. 1979.

[4] BREUSCH, T. S., AND A. R. PAGAN: "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 46(1978), 1287–1294.

[5] ———: "The Lagrange Multiplier Test and Its Applications to Model Specification," *Review of Economic Studies*, 47(1980), 239–254.

[6] COX, D. R., AND D. V. HINKLEY: *Theoretical Statistics*. London: Chapman and Hall, 1974.

[7] CROWDER, M. J.: "Maximum Likelihood Estimation for Dependent Observations," *Journal of the Royal Statistical Society*, Series B, 38(1976), 45–53.

[8] DAVIDSON, J. E. H., D. F. HENDRY, F. SRBA, AND S. YEO: "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *The Economic Journal*, 88(1978), 661–691.

[9] ENGLE, R. F.: "A General Approach to the Construction of Model Diagnostics Based upon the Lagrange Multiplier Principle," University of California, San Diego Discussion Paper 79-43, 1979.

[10] ———: "Estimates of the Variance of U.S. Inflation Based on the ARCH Model," University of California, San Diego Discussion Paper 80-14, 1980.

[11] FRIEDMAN, MILTON: "Nobel Lecture: Inflation and Unemployment," *Journal of Political Economy*, 85(1977), 451–472.

[12] GODFREY, L. G.: "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46(1978), 1293–1302.

[13] GRANGER, C. W. J., AND A. ANDERSEN: *An Introduction to Bilinear Time-Series Models*. Göttingen: Vandenhoeck and Ruprecht, 1978.

[14] KHAN, M. S.: "The Variability of Expectations in Hyperinflations," *Journal of Political Economy*, 85(1977), 817–827.

[15] KLEIN, B.: "The Demand for Quality-Adjusted Cash Balances: Price Uncertainty in the U.S. Demand for Money Function," *Journal of Political Economy*, 85(1977), 692–715.

[16] LUCAS, R. E., JR.: "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, 63(1973), 326–334.

[17] MCNEES, S. S.: "The Forecasting Record for the 1970's," *New England Economic Review*, September/October 1979, 33–53.

[18] WHITE, H.: "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," *Econometrica*, 48(1980), 817–838.