1987

A Nonparametric Test for Autoregressive Conditional Heteroskedasticity (ARCH): A Markov Chain Approach

Allan W. Gregory

ISSN:0318-725X
ISBN:0-7714-0929-X

RESEARCH REPORT 8713

A NONPARAMETRIC TEST FOR AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY (ARCH): A MARKOV CHAIN APPROACH

Allan W. Gregory

Department of Economics

University of Western Ontario

London, Ontario, Canada

N6A 5C2

September, 1987

A NONPARAMETRIC TEST FOR AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY (ARCH): A MARKOV CHAIN APPROACH

Allan W. Gregory*
University of Western Ontario
London, Ontario, Canada.

revised September 1987

This research was completed while the author was a visiting professor at Queen's University. The author would like to thank the editor, two referees, David Backus, Robert Engle, James MacKinnon, Michael Sampson, Aman Ullah and Stan Zin for helpful comments and suggestions.

ABSTRACT

In this paper we propose a nonparametric test for autoregressive conditional heteroskedasticity (ARCH) based upon finite state Markov chains. A simple Monte Carlo experiment suggests that in finite samples it performs comparably to the LM test under conditional normality and is superior for the t-, lognormal, and exponential distributions. As an illustration, we apply both tests to Canadian/United States forward foreign exchange data.

Conditional Heteroskedascity; Markov Chain; Monte Carlo.

1. INTRODUCTION

Traditionally, applied researchers have approached the problem of estimating and testing economic time-series models under the assumption of constant conditional variances. Indeed interest for time-series models appeared to be confined solely to questions of conditional means. However, with the high volatility of both micro and macro time-series data over the 1970's and 1980's, attention has recently been focussed upon developing and testing various forms of heteroskedasticity (see Pagan and Hall, 1983 and references therein). One popular form of the heteroskedasticity that seems to capture many important features of actual time-series data is the autoregressive conditional heteroskedasticity (ARCH) models first introduced by Engle (1982). Since the appearance of that paper there has been an impressive amount of work investigating ARCH models in a variety of circumstances (for a survey of some useful applications of ARCH models see Engle and Bollerslev, 1986). In the applied literature, the principal tool for determining whether an ARCH effect is present is the Lagrange Multiplier (LM) test (also suggested by Engle, 1982). The test is simple to calculate and appears to work well under conditional normality in finite samples (see Engle, Hendry, and Trumble, 1985). Moreover, as Weiss (1986) (see also, Koenker, 1982) has discussed, the LM test is also appropriate (subject to some moment conditions) for non-normal distributions.

The purpose of this paper is to propose a nonparametric test for autoregressive conditional heteroskedasticity models based upon finite state stationary Markov processes. The proposed test is as simple to calculate as the LM test, does not require any moment restrictions and appears to have better finite sample properties over a wider class of probability distributions. In a simple set of Monte Carlo experiments, we show that for quite small samples the Markov chain test based on a two-state definition performs comparably to the LM tests for conditional normal distributions and outperforms it for conditional student t-, lognormal and exponential distributions.

In Section 2 we outline the ARCH model and the LM test. In Section 3 we develop a finite state Markov chain test for the ARCH effect, and in Section 4 we conduct a simple Monte

Carlo experiment investigating the finite sample properties of the Markov chain test as well as the LM test. Also in this section we illustrate both tests using Canadian/United States foreign forward exchange data. A brief conclusion follows in Section 5.

2. THE ARCH MODEL AND A LAGRANGE MULTIPLIER (LM) TEST

Consider the p-order conditionally normal ARCH regression model (see Engle, 1982):

$$\begin{aligned} &y_t \mid \Psi_{t-1} \sim N(X_t\beta, h_t) \\ &h_t = h(\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots, \epsilon_{t-p}^2, \alpha) \qquad t = 1 \ldots N \\ &\epsilon_t = y_t - X_t\beta, \end{aligned} \tag{1}$$

where y_t is the dependent variable, Ψ_{t-1} is information set available at time $t-1$, $X_t\beta$ is a linear combination of lagged endogenous and exogenous variables included in Ψ_{t-1}, β is a $(k \times 1)$ column vector of unknown parameters, ϵ_t is a conditionally normal disturbance term, and h is a variance function with arguments $\epsilon_{t-1}^2, \ldots, \epsilon_{t-p}^2$ which are associated with the unknown $(p \times 1)$ vector of coefficients α. The error term has a mean of zero and a (non-constant) variance which depends upon p lagged values of the squared disturbances (hence the name – autoregressive conditional heteroskedasticity). Often the variance function is linearized as in Engle (1982):

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_p \epsilon_{t-p}^2, \tag{2}$$

where suitable restrictions are imposed upon α to ensure both the stationarity of the unconditional process and non-negative variances.

We mention two generalizations of the ARCH model which have recently been considered in the literature. The first by Bollerslev (1985 and 1986) is the GARCH model in which lagged dependent variables (i.e., h_{t-1}) are introduced into equation (2). A point noted by Bollerslev (1986) is that the LM test for a GARCH effect is appropriate for other different, but locally equivalent hypotheses (this property of LM tests has been noted in other applications, see for example Godrey, 1978). The second extension to the ARCH model in equation (1) is to have the variance function (2) directly in the regression equation (called ARCH-M models) as in Domowitz and Hakkio (1985) and Engle, Lilien and Robins (1987) or GARCH-M models as in Bollerslev, Engle and Wooldridge (1985) and McCurdy and Morgan (1987). The intention of placing a time-varying variance in the regression mean is to capture the risk premium frequently modelled in future and forward markets.

Notice that under the assumptions of equation (1), ordinary least squares (OLS) estimation of β that does not take account of the ARCH effect still produces consistent parameter estimates. This might suggest that we may estimate β by either: (i) OLS and then use a heteroskedastic-consistent estimator of the covariance-matrix along the lines of White (1980) or; (ii) maximum likelihood estimation (MLE) as in Engle (1982). Engle, Hendry and Trumble (1985) discuss a pretest estimator where either OLS or MLE is employed depending upon an outcome of some diagnostic test such as the one proposed here. For that matter, since conditional normality is not essential in (1) (see Weiss, 1986), we might also estimate the ARCH model by quasi-maximum likelihood methods as suggested in Weiss (1986) or even semi-nonparametric MLE devloped in Gallant and Nychka (1987) and Gallant and Tauchen (1987) which permit conditional dependence of the entire probability distribution and not just the second moment. While such topics are extremely interesting they are well beyond the scope of the present study.

As long as h in equation (1) is a differentiable function, and ϵ has finite second and fourth moments, a Lagrange multiplier (LM) test for the ARCH effect may be calculated without actually specifying the exact form of h (see Engle, 1982 and Pagan and Hall, 1983). Specifically, we test the null hypothesis that $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$ by: (i) estimating equation (1) by OLS and saving the residuals; (ii) regressing the squared residuals upon a constant and p lags; and (iii) calculating N times the R^2 (coefficient of determination) from this auxiliary regression. Under the null hypothesis of homoskedasticity (no ARCH effect), NR^2 is asymptotically distributed as X^2 with p degrees of freedom. For linear ARCH models of order 1 and conditional normality, Monte Carlo results in Engle, Hendry and Trumble (1985) suggest that this LM test has reasonable finite sample properties. However, the performance of the LM test in the absence of conditional normality is unknown. Strictly speaking, we should also note that this test is not exactly the score form of the LM test unless conditional normality is assumed (see Koenker, 1982)

Both the LM and Markov tests proposed below share the advantage that only the simpler restricted (non-ARCH) model need be estimated. On the other hand, Wald tests are computationally more difficult to calculate, requiring explicit formulation of the alternative hypothesis. Monte Carlo evidence in Engle, Hendry and Trumble (1985) indicate that for linear ARCH errors of order one under conditional normality, the Wald test is likely to have seriously reduced test sizes.

3. A MARKOV CHAIN TEST OF THE ARCH EFFECT

The ARCH test that we advance also requires that we estimate equation (1) by OLS and obtain the squared residuals denoted by e_t^2. However we no longer assume the existence of moments for ϵ . Instead we assume that the squared residuals follow a discrete stationary Markov process. Excellent introductions to the theory of Markov chains may be found in Cox and Miller (1965), Feller (1950), and Kemeny and Snell (1960).

Let e_t^2 be a random variable with $1,2,\ldots,s$ possible finite states or outcomes. If we let N and s grow arbitrarily large, then the fact that we are using the discrete least squares residual (rather than ϵ_t) is unimportant. Hence we may frame the test according to e_t. Nevertheless, from a practical point of view we must convert continuous random variables into discrete ones. Since N is finite (and usually fairly small for macroeconomic data) the number of states will necessarily be quite small. In the Monte Carlo exercise to follow we choose a number of different states. The simplest and as we shall see the one that yields the best test results is a two state definition obtained according to whether the squared residual at time t is below (state 1) or above (state 2) the sample mean ($\bar{\sigma}^2 = \sum_{t=1}^{N} N^{-1} e_t^2$). Notice that the sample mean here is the MLE of the variance under the null hypothesis of no ARCH effect and thus provides a very natural definition for the two states. However, this is obviously not the only definition possible and we discuss the choice more fully below.

We first analyze the simple first-order Markov process and then extend the argument to estimating and testing for higher order processes. The variable e_t^2 is said to be a Markov chain if:

$$Pr[e_t^2 = j \mid e_{t-1}^2 = i, e_{t-2}^2, e_{t-3}^2, \dots] = Pr[e_t^2 = j \mid e_{t-1}^2 = i] \equiv p_{ijt}. \tag{3}$$

Thus the probability distribution of e_t^2 conditional upon its entire past is identical to that conditional upon e_{t-1}^2 alone (referred to as the Markov property). We define $p_{jt} \equiv Pr[e_t^2 = j]$, $p_t \equiv [p_{jt}]$, $j = 1, \dots, s$ a $(1 \times s)$ vector and $\iota = [1]$ a $(s \times 1)$ vector. We also define the $(s \times s)$ transition matrix $Q_t = [p_{ijt}]$ with $i,j = 1, \dots, s$. Clearly $p_t \iota = 1$ and $Q_t \iota = \iota$. The unconditional probability distribution p_t changes according to:

$$p_t = p_{t-1} Q_t. \tag{4}$$

We assume that the transition matrix Q_t is independent of time so that e_t^2 has stationary transition probabilities (e_t^2 is called a homogeneous Markov chain). Equation (4) for this case may be written as:

$$p_t = p_{t-1}\, Q, \tag{5}$$

where $Q = [p_{ij}]$ and p_{ij} is the probability of going to state j from state i with $i,j = 1, \dots, s$. By iteration of equation (5) and given p_0, the starting distribution of e_t^2, it follows that:

$$p_t = p_0 Q^t. \tag{6}$$

Notice that given the initial probabilities p_0 and the matrix of transition probabilities Q, we can determine the state occupation probabilities at any time using equation (6). Assuming that

Q is irreducible and primitive, the limits of p_t and Q^t are (see Cox and Miller, 1965, pp. 118-125):

$$\lim_{t\to\infty} p_t = p \quad (p\cdot\iota = 1) \qquad \text{and}$$
$$\lim_{t\to\infty} Q^t = \iota p, \tag{7}$$

where p is the unique left-hand eigenvector of Q associated with the eigenvalue 1 and is called the equilibrium distribution.

To estimate the Markov chain we consider the likelihood function in terms of the transition probabilities p_{ij}. Assume that there are $N+1$ available observations: $e_0^2, e_1^2, \ldots, e_N^2$. Define $\alpha_i(t) = 1$ if $e_t^2 = i$ and $\alpha_i(t) = 0$ if $e_t^2 \neq i$. Then according to equation (3), the probability distribution of $e_1^2, \ldots, e_N^2$ conditional upon the first observation is:

$$\begin{aligned} &Pr[e_1^2, \ldots e_N^2 \mid e_0^2] \\ &\quad = Pr[e_1^2 \mid e_0^2]\, Pr[e_2^2 \mid e_1^2, e_0^2] \ldots Pr[e_N^2 \mid e_{N-1}^2 \ldots e_1^2, e_0^2] \\ &\quad = \mathop{\pi}_{t=1}^{N} \{\mathop{\pi}_{i} \mathop{\pi}_{j} p_{ij}^{\alpha_i(t-1)\alpha_j(t)}\} \\ &\quad = \mathop{\pi}_{i} \mathop{\pi}_{j} p_{ij}^{\beta_{ij}}, \end{aligned} \tag{8}$$

where $\beta_{ij} = \sum_{t=1}^{N} \alpha_i(t-1)\alpha_j(t)$ and is equal to the total number of transitions from i to j over the sample.

Neftci (1984) has suggested parameterizing the first observation by using its limiting value and thus maximizing the likelihood over all $N+1$ observations. Since under this strategy the optimizing problem is non-linear, we have a tradeoff between efficiency and computational simplicity. Of course, for large N, dropping the first observation is unimportant. While there are undoubtedly situations it which retaining the first observation is important, in the present case of developing a test statistic, we believe it is useful to keep the calculations as simple as possible and therefore do not investigate its effect. The log likelihood, conditional upon the first observation, is (see Anderson and Goodman, 1957):

$$L = \sum_{i=1}^{s} \sum_{j=1}^{s} \beta_{ij} \ln p_{ij} . \tag{9}$$

Letting $C_i = \sum_j \beta_{ij}$, which is the number of times the system started in state i, it follows that the maximum likelihood estimate of p_{ij} is:

$$\hat{p}_{ij} = \beta_{ij} / C_i , \tag{10}$$

and the associated value of the log likelihood at this estimate is:

$$L_u = \sum_{ij} \beta_{ij} \ln [\beta_{ij} / C_i] . \tag{11}$$

There are three likelihood ratio tests which provide some information regarding the nature of the heteroskedasticity. The first test is a test of the independence of the observations against a first order. The second is a test of a first order process against a second order (extensions to higher orders follow directly). The third is a test of the stationarity of the process generating the e_t^2 and investigates the stability of the Markov chain. The validity of standard likelihood theory for statistical inference in Markov chains is shown in Billingsley (1961). A presentation and discussion of the tests applied here appears in Anderson and Goodman (1957) and Chatfield (1975).

(i) Testing Independence of e_t^2 and e_{t+k}^2

The question we address here is whether there is independence between observations. Under homoskedasticity, the lagged squared residual e_{t-1}^2 should provide no information in predicting current squared residual e_t^2 . We test the hypothesis that e_t^2 and e_{t+k}^2 are independently distributed ($H_0 : p_{ij} = p_j$, the conditional probability of being in state j given that state i has occurred is equal to the unconditional probability of being in state j). If the restriction $p_{ij} = p_j$ is imposed, the maximum likelihood estimate of p_{ij} is $\hat{p}_{ij} = C_j / N$ and the corresponding value of the log likelihood is:

$$L_I = \sum_{i=1}^{s} C_i \ln [C_i / N] . \tag{12}$$

Under the null hypothesis of temporal independence:

$-2(L_I - L_u) \overset{a}{\sim} X_r^2$, where $r = (s-1)^2$.

(ii) Testing a First-Order Versus a Second-Order Markov Chain

Although higher order chains could be considered, in practice we have found that with stationary economic time series, higher orders are generally unnecessary. For a stationary (homogeneous) second-order Markov chain, we denote the probability of being in state k at time t given that $e_{t-2}^2 = i$ and $e_{t-1}^2 = j$ by $p_{ijk}\,(i,j,k = 1, \dots, s)$ for $t = 2, \dots, N$. Thus the first-order Markov chain is a special case of the second-order chain, since p_{ijk} does not depend on i. As Anderson and Goodman (1957) discuss, the second-order chain may be represented by a more complicated first-order chain. That is, we represent the pair of successive states i and j as a composite state (i,j). Therefore the probability of the composite state (j,k) at time t given the composite state (i,j) at time $t-1$ is p_{ijk} (we continue to assume a homogeneous Markov process). Alternatively, the probability of state (h,k) $h \neq j$, given (i,j) has occurred is zero. Hence defining composite states gives rise to a chain with s^2 states and a transition matrix in which some entries are zero. The equilibrium distribution p then pertains to the unconditional (long-run) probability of the composite state (i,j).

If we let $\beta_{ijk} = \sum_{t=2}^{N} \alpha_i(t-2)\,\alpha_j(t-1)\,\alpha_k(t)$ be the total number of transitions from i to j to k over the sample, then the log likelihood, conditional upon the first two observations (e_0^2 and e_1^2) is:

$$L_s = \sum_{i=1}^{s} \sum_{j=1}^{s} \sum_{k=1}^{s} \beta_{ijk} \ln (p_{ijk}). \tag{13}$$

If we let $C_{ij} = \sum_k \beta_{ijk}$ be the number of times the system started in state i and j then the maximum likelihood estimate of p_{ijk} is: $\hat{p}_{ijk} = \beta_{ijk} / C_{ij}$. Substitute $\hat{p}_{ijk}$ into (13) and denote this maximized value of the log likelihood as L_s. Therefore, if we also maximize the log of the likelihood for the first-order Markov chain (9) over $(N-1)$ observations, then a test of the null hypothesis that the chain is first order against a second order is:

$-2(L_u - L_s) \overset{a}{\sim} X_r^2$, where $r = s(s-1)^2$. Analogous to the LM test for ARCH of order two, we may test the independence hypothesis against a second order Markov process $(H_0 : p_{ijk} = p_k)$. This test is: $-2(L_I - L_s) \overset{a}{\sim} X_r^2$, where $r = (s+1)(s-1)^2$.

(iii) Stability Tests of the Markov Chain

To test the hypothesis that the Markov is stable over time we carry out a simple Chow-type test (see Anderson and Goodman, 1957). For the first-order process, we maximize the log likelihood (9) over the entire sample (N observations). Then we divide the sample into as many parts as desired (say; m divisions) and maximize the likelihood for each subsample which we denote by $L_1, L_2, \ldots, L_m$. All N observations must be used in subsample estimation and this requires some care when defining subsample intervals. Under the null hypothesis that each of the subsample processes is from the same first-order Markov chain:

$$-2(L_u - L_1 - L_2 - \ldots - L_m) \overset{a}{\sim} X_r^2, \tag{14}$$

where $r = (m-1)s(s-1)$. The stability test for a second order Markov process uses L_s instead of L_u and there are $(m-1)s^2(s-1)$ degrees of freedom.

Once we have the discrete squared OLS residuals, a natural way to proceed is: (i) to estimate and to test the stability of the second order Markov process; (ii) then if the second order process is deemed stable (at some appropriate significance level) test a first-order against a second; and (iii) if the data supports the restriction to a first order process, stability tests could then be applied, followed by a test of independence against the first order process. Of course there are many other potential testing avenues which could lead to possibly different conclusions. As a word of caution, we note that since the tests are not always independent, adherence to strict rejection regions should probably be avoided.

As discussed earlier, to operationalize this procedure, we must decide upon a rule to turn the continuous variable e_t^2 into a discrete one. That is, we must define the number of states and the cell width for each state. Although some 'reasonable' guidelines may often be suggested from the problem itself (say, on the basis of sign change of the variable as in Gregory and Sampson, 1987 or unconditional variances as in Tauchen, 1986) the decision is nevertheless arbitrary.

From a viewpoint of gaining a complete characterization of how the current squared residual depends upon its past, it would be desirable to have many states defined for narrow cell widths for each state. In this way, we could fully investigate whether magnitudes of lagged squared residuals are important in determining the probability of what state will likely occur. Unfortunately, a large number of states substantially increases the dimensions of the transition matrix Q and necessitates estimating a large parameter space relative to the number of observations. A further problem with a large number of state definitions is that many of the transitions would never actually be observed over the time period. In this case the corresponding estimate of the transition probability would be zero which would create the (false) impression that such transitions could never occur. Hence there is some virtue in defining a fairly wide cell width and consequently limiting the number of possible states. This certainly raises some worry about the ability of the tests of independence to detect influences that are very sensitive to the magnitudes of the squared residual.

For the present application of testing for ARCH errors, a natural dividing point is the sample mean of e_t^2 . This is a consistent estimate of the variance under the null hypothesis of no ARCH effect. Using the sample mean as the boundary condition gives rise to a simple two state classification: above and below the mean. Our Monte Carlo results below clearly demonstrate that this provides quite reasonable test results (both in terms of test size and power). However, if finer partitioning is desired then the choice of state definitions should reflect the compactness of the squared residuals around the sample mean. Without any further information it is probably best to opt for symmetric definitions on the basis of some volatility measure like sample standard deviation. We illustrate these kinds of state definitions in the Monte Carlo experiment.

In this paper we present three possible rules for choosing the states. In total, we consider two, three, and four state definitions. The simplest and as it turns out the best in the Monte Carlo experiment is to define state occupancy according to whether the observation e_t^2 is below or above the sample mean $\bar{\sigma}^2 = \sum_{t=1}^{N} N^{-1} e_t^2$. This yields two states: low variance (L) and high variance (H). For three states we use the sample mean plus and minus one quarter of the unconditional standard deviation of the squared residuals as boundary conditions. This gives rise to three states: low variance (L), average or medium variance (M), and high variance (H). Lastly a four state Markov chain can be obtained from the three state definition by including the sample mean as another cell boundary. While other state definitions could always be chosen, these are sufficiently broad to illustrate the effects of increasing the number of states. Finally we note that these rules for state definitions are symmetric about the mean. Although this is not required, some limited Monte Carlo experiments indicated that non-symmetric definitions yield poor test results. However, for some probability distribution it is possible that non-symmetric definition would dominate.

4. A SIMPLE MONTE CARLO EXPERIMENT AND AN ILLUSTRSTION

To assess the finite sample performance of the Markov chain test for ARCH errors compared to the LM test, we conduct a simple Monte Carlo experiment. We follow essentially the same experimental design as Engle, Hendry and Trumble (1985). The ARCH model with $p = 1$ used to generate the data and calculate the two tests is:

$$
\begin{aligned}
y_t &= \beta x_t + \epsilon_t \qquad t = 1, \dots, N \\
x_t &= \lambda x_{t-1} + \nu_t \qquad \nu_t \sim IN(0, \sigma_\nu^2) \\
\epsilon_t &= \eta_t(\gamma + \alpha\, \epsilon_{t-1}^2)^{1/2},
\end{aligned}
\tag{15}
$$

where y_t is the dependent variable, x_t is an exogenous forcing process with normal disturbance term ν_t and η_t is the error term drawn from some independent and identical distribution. In this Monte Carlo experiment we maintain the normality assumption and the independence of η_t for the exogenous process. For non-zero values of α, the error term ϵ_t in the regression equation will be ARCH of order 1. To guage the sensitivity of the tests to different distributional assumptions we let η_t follow a (i) standard normal distribution; (ii) t-distribution with 5 degrees of freedom; (iii) lognormal distribution with a mean and variance of 0 and 1 respectively and (iv) an exponential distribution with a mean of 1/2. While other distributions and other parameters of the distribution could have been chosen, we feel that these four accurately characterize the relative performance of the two tests. For all experiments we set $\beta = 1$, $\sigma_\nu^2 = 4$ and the scale parameter $\gamma = 1 - \alpha$ (as in Engle, Hendry and Trumble, 1985).

We also study the effect of varying the autoregressive parameters α and λ and the number of observations N. Following Engle, Hendry and Trumble (1985) we set $\alpha = 0.0$, 0.4, and 0.8, $\lambda = 0.0$, 0.8 and $N = (4 + j)^2$ $j = 1, \dots, 5$. The x_t are held fixed in repeated samples within experiments but are generated separately between experiments. A full factorial design implies 120 experiments.

The data are generated from a pseudo-random generator (G05) in the NAG subroutines. In Table I - Table III we record the number of rejections of the null hypothesis of no ARCH errors in 1000 replications of the two tests at the 10 and 5 percent level of significance. As explained above we choose three different rules for defining states for the Markov chains (labelled 2 state, 3 state, and 4 state in the tables). Both tests use the knowledge that there is no intercept in (15) and that $p = 1$.

First we turn to the results in Table I for $\alpha = 0.0$ (no ARCH errors). The 95 percent confidence intervals for the expected number of rejections in 1000 replications when the null hypothesis is true at the 10 and 5 percent level are [81,119] and [36,64] respectively. We see for all distributions and sample sizes considered, the actual size of the two state Markov chain test is closer to the expected asymptotic value than the corresponding LM test. The LM test appears to be biased towards the null hypothesis of no ARCH effect (this is especially true for the lognormal distribution). While there are occasions for which the two state Markov chain test over-rejects relative to expected (see $N = 36$, $\lambda = 0$ and the exponential distribution), tests are typically close to the correct size by $N = 81$. Reasonable test sizes for the LM test occurs for only the normal distribution. The apparent effect of a finer state definitions for the Markov test is first to cause the test to be biased away from the null hypothesis (3 state chain) and then toward the null (4 state chain). The biases do not appear to diminish that rapidly with sample size. On balance the the four state Markov chain test (with the sample mean as the center boundary condition) appears to have better test sizes than the three state chain; however both are dominated by the two state test. The amount of serial correlation in the forcing process has no appreciable systematic effect on any of the test statistics over the various distributions.

With $\alpha = 0.4$ and 0.8 (Table II and III), the number of rejections for all tests increases in α . However, the two state Markov chain test is much better able to detect a false null hypothesis than the LM test for the lognormal and exponential distributions. This suggests that the LM test is especially sensitive to departures from symmetry. The tables show that the LM test is superior to the Markov chain tests under conditional normality and about the same as the two state under the conditional t-distribution. We conducted a limited set of experiments using a t- distribution with 2 degrees of freedom and found that the number of rejections for the LM test with α not equal to zero fell dramatically. On the other hand, for a t- distribution with 2 degrees of freedom, the two state Markov chain test results were quite similar to those in Table II and III (results are available upon request). As might be expected given their test sizes, the three state Markov chain test has more rejections than the four state test . However, the number of rejections for three states is generally quite similar to the two state and LM tests. Also with $\alpha = 0.4$ and 0.8 for the lognormal and exponential distributions, we encountered singularity problems in running the artificial regression for the LM test and some runs had to be repeated.

In sum the LM test works best with ARCH errors under conditional normality. For other distributions, our limited set of Monte Carlo experiments suggest that the simple two state Markov chain test is preferred.

Finally we wish to illustrate the two tests using an empirical example from Gregory and McCurdy (1984) and investigate whether the forward foreign forecast errors are independent of a subset of current information. Without going into a great deal of detail (see Gregory and McCurdy, 1984), the test relation regression is:

$$\frac{s_{t+1} - f_t}{s_t^T} = \beta_0 + \beta_1 \left[\frac{s_t^T - f_{t-1}}{s_t^T} \right] + \beta_2 \left[\frac{f_t - s_t^T}{s_t^T} \right] + \epsilon_{t+1}, \tag{16}$$

where f_t and s_t^T are the Tuesday closing rates of the thirty-day forward and spot rate respectively and s_{t+1} is the Thursday closing spot rate four weeks and two business days into the future and ϵ_{t+1} is an error term. Thus there are thirty days between f_t and s_{t+1}.

To motivate this regression consider the following orthogonal decomposition of the future spot rate:

$$s_{t+1} = E_t\, s_{t+1} + (s_{t+1} - E_t\, s_{t+1})$$

where E_t is the mathematical expectation operator conditional upon information available at time t and $\epsilon_{t+1} = s_{t+1} - E_t\, s_{t+1}$. Under the rational expectations hypothesis (REH) $E_t\, s_{t+1} = f_t$ so that $s_{t+1} - f_t$ is the forecast error. Normalizing the forecast error by s_t^T yields the dependent variable in equation (16) which under the market efficiency hypothesis (MEH) should be orthogonal to information available at time t. In (16) we choose a subset of the information set namely the normalized lagged forecast errors and forward premium.

The question of whether the (normalized) forecast error is independent of current information may be addressed by testing the hypotheses that the estimated column vector β is not significantly different from zero. However, for our purposes, the most relevant issue is the properties of the error term ϵ_{t+1}. Under the rational expectations hypothesis (REH), ϵ_{t+1} should be serially uncorrelated but need not be homoskedastic (see, for example, Cumby and Obstfeld, 1983, Domowitz and Hakkio, 1985; Hodrick and Srivastava, 1984; and Hsieh, 1984). In fact, Baillie and Bollerslev (1987), Domowitz and Hakkio (1985) and McCurdy and Morgan (1987 and 1985) have tested and estimated a similar equation to (16) in which the disturbance term follows an ARCH process.

Using the same data as Gregory and McCurdy (1984), we estimate equation (16) by ordinary least squares using four-weekly data for Canada/United States over the period 1973-1981 which gives a sample size of one hundred and seventeen four-weekly observations. The results are identical to those in Table 2 of Gregory and McCurdy (1984). The OLS estimates of equation (16) are:

$$\frac{s_{t+1} - f_t}{s_t} = \underset{(0.0012)}{0.00212} - \underset{(0.091)}{0.210}\left[\frac{s_t^T - f_{t-1}}{s_t}\right] - \underset{(0.64)}{2.143}\left[\frac{f_t - s_t^T}{s_t^T}\right] \qquad R^2 = 0.1$$

where standard errors are given in parentheses.

We test for a first and second order ARCH effect using the LM test. The right side (one tail) prob-values for first and second order are 0.832 and 0.531 respectively, both indicating an absence of an ARCH effect. In view of the Monte Carlo results, we consider the two state Markov chain (above and below sample mean) in detail. Using this state definition we estimate a second order Markov chain for the squared residuals. The estimated transition probabilities, equilibrium distribution and various hypothesis tests are recorded in Table 4. The horizontal column is the conditioning composite state, the vertical is the outcome composite state. Thus, the probability of going from (L,L) - low variances for last period and the period before, to a high variance next period (L,H) is 0.235. Notice that the first letter for the outcome composite state must match the second letter in the conditioning composite state, otherwise the event is not possible and is labelled by an '-' (e.g., (L,H) to (L,H)).

From the transition probabilities in Table 4 we see (casual) evidence that there is a 'gain' in conditioning the forecast of the squared residuals on the current state. For example the unconditional probability of composite state (H,L) occuring is 0.165 (as given by the equilibrium distribution) but conditional on (H,H) occuring it is 0.900. Thus there is some informal support for an autoregressive structure to the squared residuals.

The first hypothesis test in Table 4, tests the independence hypothesis against a second order process ($H_0 : p_k = p_{ijk}$). At the 10 percent level of significance this hypothesis is rejected and suggests that there may be ARCH errors present (of course, without knowing the actual data genereting process we cannot be certain). Testing independence against a first order yields a fairly high prob-value, suggesting no ARCH effect. Such differences highlight the need to consider more than just a first order processes. This point is reinforced in the test of a first order Markov chain against a second with a prob-value of 0.083. Lastly, we arbitrarily split the sample in two and find that the first and second order Markov processes appear to be stable (prob-value of 0.194 and 0.124 respectively). We also estimated the three state and four state Markov chains using the rules given above; the prob-value for the test of a second order chain against independence are 0.086 and 0.401 respectively. Again these two results are consistent with the Monte Carlo evidence; that is there is a greater tendency to reject the null hypothesis using the three state definition than the four.

5. CONCLUSION

In this paper we have developed a test for autoregressive conditional heteroskedasticity (ARCH) in regression errors based upon finite state Markov chains. The principal advantage of the test over the Lagrange multiplier (LM) test is that no moment conditions are assumed and using a simple two state definition, Monte Carlo evidence from a limited set of experiments indicates good finite sample properties over a wide class of probability distributions. Finally in a practical example we showed that the Markov chain test provides some useful information in describing the movements of the variances over time.

References

Anderson, T.W. and Goodman, L.A. (1957), "Statistical Inference About Markov Chains," *The Annals of Mathematical Statistics*, 28, 89-110.

Baillie, R.T. and Bollerslev, T. (1987), "Multivariate GARCH Processes and Models of Time Varying Risk Premia in Foreign Exchange Markets", unpublished paper, Michigan State University, 20 pp.

Billingsley, P. (1961) *Statistical Inference for Markov Processes*, University of Chicago Press.

Bollerslev, T (1985), "A Note on the Correlation Structure for the Generalized Autoregressive Conditional Heteroskedastic Process," Memo 1985-11, Institute of Economics, University of Aarhus, 16 pp.

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

Bollerslev, T.P., Engle, R.F. and Wooldridge, J.M. (1985), "A Capital Asset Pricing Model with Time Varying Covariances", unpublished manuscript, University of California, San Diego, 22 pp.

Chatfield, C. (1973), "Statistical Inference Regarding Markov Chain Models," *Applied Statistics*, 22, 7-20.

Cox, D.R. and Miller, H.D. (1965), *The Theory of Stochastic Processes*, Chapman and Hall.

Cumby, R.E. and Obstfeld, M. (1983), "International Interest-Rate and Price-level Linkages Under Flexible Exchange Rates: A Review of Recent Evidence," in J.F.O. Bilson and R.C. Marston, eds., *Exchange Rates: Theory and Practice*, Chicago: University of Chicago Press.

Domowitz, I. and Hakkio, C.S. (1985), "Conditional Variance and the Risk Premium in the Foreign Exchange Market," *Journal of International Economics*, 19, 47-66.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflations," *Econometrica*, 50, 987-1007.

Engle, R.F. (1983), "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model," *Journal of Money, Credit, and Banking*, 15, 286-301.

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

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

Engle, R.F. and Kraft, D. (1983), "Multiperiod Forecast Error Variances of Inflation Estimated from ARCH Models," *Applied Time Series Analysis of Economic Data* (ed. A. Zellner) Bureau of the Census, 293-302.

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

Feller, W. (1950), *An Introduction to Probability Theory and Its Applications*, Vol. 1, John Wiley and Sons.

Gallant R.A., Nychka, D.W. (1987), "Semi-Nonparametric Maximum Likelihood Estimation, *Econometrica*, forthcoming.

Gallant, R.A. and Tauchen, G. (1987), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," unpublished paper, North Caolina State University, 63 pp.

Godfrey, L.J. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors include Lagged Dependent Variables," *Econometrica* 46: 1293-302.

Gregory, A.W. and McCurdy, T.H. (1984), "Testing the Unbiasedness Hypothesis in the Forward Foreign Exchange Market: A Specification Analysis," *Journal of International Money and Finance*, 3, 357-368.

Gregory, A.W. and Sampson, M.J. (1987), "Testing the Independence of Forecast Errors in the Forward Foreign Exchange Market Using Markov Chains: A Cross-Country Comparison," *Journal of International Forecasting*, 3, 97-113.

Hodrick, R.J. and Srivastava, S. (1984), "An Investigation of Risk and Return in Forward Foreign Exchange," *Journal of International Money and Finance*, 3, 5-29.

Hsieh, D.A. (1982), "Tests of Rational Expectations and No Risk Premium in Forward Exchange Markets," *Journal of International Economics*, 17, 173-184.

Kemeny, J. and Snell, J. (1960), *Finite Markov Chains*, D. Van Nostrand Company.

Koenker, R. (1981), "A Note on Studentizing a Test for Heteroskedasticity," *Journal of Econometrics*, 17, 107-112.

McCurdy, T.H. and Morgan, I.H. (1985), "Testing the Martingale Hypothesis in Deutschmark/U.S. dollars Futures and Spot Markets," Discussion paper no. 639, Department of Economics, Queen's University.

McCurdy, T.H. and Morgan, I.H. (1987), "Tests of the Martingale Hypothesis for Foreign Currency Futures with Time-Varying Volatility," *International Journal of Forecasting*, 3, 131-148.

Neftci, S.N. (1984), "Are Economic Time Series Asymmetric over the Business Cycle," *Journal of Policical Economy*, 92, 307-28.

Pagan, A. and Hall, A.D. (1983), "Diagnostic Tests as Residual Analysis," *Econometric Review*, 2, 159-218.

Tauchen, G. (1986), "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions,"*Economic Letters*, 2, 177-181.

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

White, H. (1980), "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-38.

Table 1. Testing for First Order ARCH Errors: Number of Rejections at the 10 and 5 percent levels in 1000 replications ($\alpha=0.0$)

Normal

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	119	72	146	61	68	42	54	18
	36	108	57	164	71	73	31	61	28
0.0	49	120	69	132	66	85	32	66	28
	64	109	55	142	78	116	57	61	26
	81	99	55	153	76	120	59	67	35
	25	114	77	148	73	71	43	53	17
	36	114	70	134	72	81	39	62	26
0.8	49	123	74	157	80	96	46	67	33
	64	97	41	121	72	103	50	76	34
	81	112	46	122	63	117	50	81	39

t_5

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	101	63	137	61	73	43	48	20
	36	123	77	152	68	83	32	41	17
0.0	49	111	57	153	78	102	44	40	16
	64	112	59	136	69	114	56	49	30
	81	128	77	133	71	124	59	61	35
	25	99	66	133	59	50	22	46	18
	36	123	54	129	57	61	36	57	24
0.8	49	119	70	152	71	100	46	52	22
	64	102	42	149	92	125	59	57	29
	81	107	44	122	56	123	52	41	26

Lognormal

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	88	44	120	60	79	48	34	13
	36	69	36	103	47	64	33	27	16
0.0	49	94	44	112	50	81	44	31	23
	64	87	37	83	44	73	39	34	28
	81	85	44	77	37	73	32	42	32
	25	66	30	120	66	72	43	33	13
	36	74	32	108	53	78	42	29	15
0.8	49	84	29	116	59	92	46	30	25
	64	92	31	114	53	81	44	32	26
	81	102	52	93	51	84	44	24	18

Exponential

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	108	67	115	54	71	51	47	24
	36	150	71	137	66	54	20	46	29
0.0	49	131	79	152	72	88	42	52	33
	64	128	61	147	70	88	42	62	38
	81	126	66	138	65	125	55	45	31
	25	98	60	106	61	58	35	39	20
	36	116	60	125	64	89	37	40	25
0.8	49	110	49	148	76	90	45	46	33
	64	137	67	142	83	93	49	45	34
	81	103	54	153	79	115	52	53	26

Note: The Markov chain tests for ARCH errors are defined using boundary conditions of the squared residuals obtained from: (i) sample mean (2 state); (ii) plus/minus one quarter standard deviations from mean (3 state) and (iii) mean, plus/minus one quarter standard deviation from mean (4 state). 10% and 5% refer to the number of rejections at the 10 percent and 5 percent level of significance respectively.

Table 2. Testing for First Order ARCH Errors: Number of Rejections at the 10 and 5 percent levels in 1000 replications (α=0.4)

Normal

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	171	131	224	118	95	46	214	151
	36	257	188	270	159	157	84	347	252
0.0	49	283	205	299	190	213	125	452	358
	64	331	219	354	244	283	190	554	466
	81	418	323	404	298	346	237	673	585
	25	175	117	217	121	106	46	216	136
	36	241	173	260	164	159	98	349	265
0.8	49	272	190	279	181	208	111	439	358
	64	374	267	364	273	311	207	555	484
	81	422	314	392	294	359	245	617	531

t_5

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	253	182	277	176	143	83	236	166
	36	362	290	348	252	241	150	375	295
0.0	49	431	345	434	315	336	242	485	420
	64	526	415	485	370	440	320	575	343
	81	628	537	600	488	541	441	660	603
	25	252	178	277	163	145	77	253	182
	36	507	405	419	317	328	218	335	258
0.8	49	424	344	445	338	379	263	502	420
	64	535	439	496	380	441	334	602	517
	81	637	535	576	457	537	410	670	599

Lognormal

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	446	339	438	312	239	162	359	281
	36	593	527	581	465	451	329	500	424
0.0	49	762	688	684	593	635	540	617	552
	64	858	791	729	656	711	609	666	611
	81	919	874	763	700	780	713	723	678
	25	440	321	437	327	234	154	326	229
	36	614	527	558	477	453	344	480	401
0.8	49	747	693	686	589	614	526	578	523
	64	875	812	741	678	713	622	688	629
	81	933	897	800	761	828	760	733	676

Exponential

		Markov						LM	
		2 State		3 State		4 State			
λ	N	10%	5%	10%	5%	10%	5%	10%	5%
	25	427	317	414	285	205	130	282	205
	36	544	451	477	386	393	311	367	300
0.0	49	669	601	541	438	475	389	471	397
	64	772	718	620	547	591	490	545	481
	81	868	811	633	569	666	550	608	551
	25	365	286	346	248	184	117	237	185
	36	510	425	429	335	344	228	365	296
0.8	49	619	551	508	389	439	337	449	383
	64	783	711	558	486	559	446	516	453
	81	831	783	614	542	642	540	601	539

Table 3. Testing for First Order ARCH Errors: Number of Rejections at the 10 and 5 percent levels in 1000 replications ($\alpha=0.8$)

Normal

λ	N	Markov 2 State 10%	2 State 5%	3 State 10%	3 State 5%	4 State 10%	4 State 5%	LM 10%	LM 5%
	25	327	254	358	261	201	134	368	271
	36	484	406	506	388	384	257	512	423
0.0	49	617	524	625	521	536	427	679	601
	64	784	627	694	601	639	526	778	712
	81	777	709	785	707	747	654	848	792
	25	374	286	376	260	225	130	403	297
	36	485	400	489	385	381	274	532	441
0.8	49	572	468	565	451	491	390	660	579
	64	717	618	715	613	646	537	780	708
	81	777	695	770	687	730	633	864	807

t_5

λ	N	Markov 2 State 10%	2 State 5%	3 State 10%	3 State 5%	4 State 10%	4 State 5%	LM 10%	LM 5%
	25	446	345	443	335	247	157	400	318
	36	560	505	588	494	466	348	520	439
0.0	49	725	646	697	579	615	500	661	574
	64	818	737	776	705	736	651	742	695
	81	915	857	848	793	838	766	815	749
	25	451	349	474	369	276	182	381	303
	36	576	508	581	473	473	356	545	470
0.8	49	717	647	686	582	605	500	638	556
	64	828	763	786	714	735	653	758	692
	81	892	849	845	784	840	757	827	781

Lognormal

λ	N	Markov 2 State 10%	2 State 5%	3 State 10%	3 State 5%	4 State 10%	4 State 5%	LM 10%	LM 5%
	25	630	532	614	504	338	264	528	439
	36	828	772	800	730	711	606	704	652
0.0	49	919	889	858	812	831	770	771	737
	64	958	937	872	841	855	826	820	790
	81	986	974	868	847	901	859	850	811
	25	622	533	626	529	404	286	540	452
	36	809	742	761	685	657	559	658	595
0.8	49	905	846	842	778	808	756	746	694
	64	965	936	855	817	871	831	821	791
	81	986	979	871	850	916	871	856	828

Exponential

λ	N	Markov 2 State 10%	2 State 5%	3 State 10%	3 State 5%	4 State 10%	4 State 5%	LM 10%	LM 5%
	25	501	343	482	385	271	168	297	299
	36	645	548	542	484	472	372	381	323
0.0	49	774	701	568	483	520	451	439	371
	64	877	829	582	523	594	505	522	450
	81	938	903	556	518	645	529	593	526
	25	478	341	427	326	222	132	296	232
	36	615	501	511	426	387	293	355	295
0.8	49	774	711	555	455	484	405	453	395
	64	867	829	569	508	551	452	504	439
	81	928	895	589	542	637	537	604	533

Table 4. Second Order Two State Markov Chain: Transition Matrix, Equilibrium Distribution and Hypothesis Tests

	composite state	(L,L)	(L,H)	(H,L)	(H,H)
Transition Matrix	(L,L)	0.765	0.235	-	-
Q =	(L,H)	-	-	0.526	0.474
	(H,L)	0.833	0.167	-	-
	(H,H)	-	-	0.900	0.100
Equilibrium Distribution					
p		0.584	0.165	0.165	0.086

Hypothesis Tests	Prob Value
Independence against second order	0.083
Independence against first order	0.190
First order against second order	0.083
Stability (split sample): first order	0.194
second order	0.124

Note: State transitions that are not possible (i.e., (L,L) (H,L)) are denoted by '-'. The division point for the stability tests was arbitrarily chosen at one half of the sample.