"Testing for Serial Correlation in the Presence of ARCH," *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 1986*, 323-328. Washington, DC: American Statistical Association, 1987.

59

TESTING FOR SERIAL CORRELATION IN THE PRESENCE OF ARCH

Francis X. Diebold, Board of Governors of the Federal Reserve System

1) Introduction

The problem of testing for serial correlation arises constantly in time-series econometrics. Sometimes, as with forward premia in efficient markets studies, the time series to be tested for serial correlation is directly observed. Sometimes, as with residuals from an estimated model, the observed series is only an estimate of the true, but unknown, series to be tested for serial correlation. Either way, the presence of heteroskedasticity violates the assumptions upon which tests for serial correlation are built.

This observation is particularly crucial in light of the recent realization that conditional heteroskedasticity may be commonly present in the time-series context. (See, for example, Engle (1982b) and Weiss (1984). There are two approaches to the dilemma. First, one may attempt to develop tests for serial correlation that are robust to heteroskedasticty of unknown form. This is the approach taken by Domowitz and Hakkio (1985) who combine Godfrey's (1978) Lagrange multiplier test for serial correlation with White's (1980) heteroskedasticity-consistent covariance matrix estimator. The advantage of such an approach is its generality; the cost is reduced power in situations when the form of the heteroskedasticity is known or can be well approximated.

The second approach is to parameterize, or approximate, the form of the heteroskedasticity, and develop serial correlation tests specifically taking it into account. This of course has costs and benefits opposite those of the Domowitz-Hakkio approach. To the extent that the heteroskedasticity approximation is accurate, the test will perform well, and vice versa.

The model of autoregressive conditional heteroskedasticty (ARCH) due to Engle (1982b) has been found to provide a parsimonious and descriptively accurate approximation in many contexts (inflation: Engle (1982c); foreign exchange markets: Domowitz and Hakkio (1985), Diebold and Pauly (1986), Diebold and Nerlove (1985, 1986); stock market: diebold, lee and Im (1986), Diebold (1986); term structure of interest rates: Engle, Lillien and Robbins (1985)). In this paper we consider the properties of two important and heavily used time-series model specification tools, the sample autocorrelation function and the Box-Pierce (1970) and Ljung-Box (1978) "portmanteu" statistics, in the presence of ARCH. The theory of the Bartlett standard errors is first developed, and then the portmanteu tests are treated. We build upon the results of Milhoj (1985) to show why the presence of ARCH renders the usual Bartlett standard error bands overly conservative, relative to the nominal 5% test size, and we develop an ARCH-corrected standard error estimate. This leads directly to ARCH-corrected confidence intervals under the null of uncorrelated white noise. We then treat the Box-Pierce and Box-Ljung serial correlation test statistics and show that they do not have the usual χ^2 limiting null distribution. An appropriate normalization is found which does have a limiting χ^2 distribution, however. The results are illustrated with a numerical example.

2) Correcting the Bartlett Standard Errors

Consider a zero-mean time series $\{x_t\}_{t=1}^{T}$. It can be shown (Anderson (1942), Bartlett (1946)) that, under the null of Gaussian white noise, the sample autocorrelation at lag τ :

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$$

where $\hat{\gamma}(\tau) = 1/T \; \Sigma \; x_t \; x_{t-\tau}$ is asymptotically normally distributed with mean 0 and variance:

$$\text{var}\;(\hat{\rho}(\tau)) = \frac{T - \tau}{T\;(T + 2)}$$

or, as a further approximation, 1/T. This result leads to the so-called Bartlett 95% confidence interval under the null:

$$\rho(\tau) = 0.0 \pm \frac{1.96}{\sqrt{T}} .$$

Under ARCH, however, the sample autocorrelations are normal with mean 0 and variance:

$$(1/T)\;(1 + \frac{\gamma_{x^2}(\tau)}{\sigma^4})$$

where $\gamma_{x^2}(\tau)$ is the autocovariance at lag τ for the squared process $\{x_t^2\}_{t=1}^{T}$ and σ^4 is the squared unconditional variance of the x process. (See Milhoj (1985).) Because:

$$\frac{\gamma_{x^2}(\tau)}{\sigma^4} > 0 \quad \text{for all } \tau$$

it is clear that Bartlett's standard error is "too small" in the presence of ARCH. Note, however, that:

$$\lim_{\tau \to \infty} \;(1/T)\;(\;1 + \frac{\gamma_{x^2}(\tau)}{\sigma^4}\;) = 1/T$$

since $\gamma_{x^2}(\tau) \to 0$ as $\tau \to \infty$, by stationarity and ergodicity of $\{x^2\}$. Because $\gamma_{x^2}(\tau)$ and σ^2 are easily consistently estimated, we can construct a consistent estimate of the variance of the sample autocorrelations as:

$$S(\tau) = (1/T)\ (1 + \frac{\hat{\gamma}_{x^2}(\tau)}{\hat{\sigma}^4})$$

which leads to the corrected confidence interval:

$$\rho_x(\tau) = 0.0 \pm 1.96\ (S(\tau))^{1/2}\ .$$

To implement the results over, say, the first K autocorrelations, we first obtain:

$$\hat{\rho}_x(\tau) = \frac{\Sigma\, x_t x_{t-\tau}}{\Sigma\, x_t^2},\quad \tau = 1 \ldots K$$

$$\hat{\sigma}^4 = (\hat{\sigma}^2)^2 = (1/T\ \Sigma\, x_t^2)^2$$

$$\hat{\gamma}_{x^2}(\tau) = 1/T\ \Sigma\, (x_t^2 - \hat{\sigma}^2)\,(x_{t-\tau}^2 - \hat{\sigma}^2)$$

and then construct the bands via the above formula.

To illustrate, 500 observations were generated on the process:

$$x_t = \varepsilon_t,\quad \varepsilon_t \mid \varepsilon_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\, \varepsilon_{t-1}^2\ .$$

The first 20 autocorrelations of x were calculated, along with the Bartlett 1.96 standard error bands and the ARCH-corrected Bartlett 1.96 standard error bands. One thousand replications were performed for each of ten points in the parameter space: α_1 = 0.0, .1, .2, .3, .4, .5, .6, .7, .8, .9. Without loss of generality, we can set $\alpha_0 = 1 - \alpha_1$ (Pantula (1985)), which maintains the unconditional variance at 1.0. The case of $\alpha_1 = 0.0$ of course corresponds to independent white noise. The realizations were generated via the cannonical form:

$$\varepsilon_t = N_t(0,1)(\alpha_0 + \alpha_1\, \varepsilon_{t-1}^2)^{1/2}$$

where we set $\varepsilon_0 = 0$. The same one-thousand sets of 500 innovations $\{N_t(0,1)\}_{t=1}^{500}$ were used to generate the ARCH realization at each explored point of the sample space; this provides powerful variance reduction. The proportions of rejections (in 1000 repetitions over 20 autocorrelations) relative to the uncorrected Bartlett 95% confidence interval are given in Table 1 as P, while rejection frequencies relative to the corrected intervals appear as P_c.

The results speak for themselves. When $\alpha_1 = 0$, of course, the nominal size (5%) approximately equals the actual size (4.6%). This is also true if the ARCH correction is (needlessly) applied. As α_1 rises, however, so too does the empirical size of the uncorrected confidence interval, so that, for example, when $\alpha_1 = .9$, the probability of a type I error is more than twice the nominal probability of 5%. The ARCH-corrected intervals, on the other hand, do a beautiful job of maintaining nominal size.

The problem of spurious "significance" of sample autocorrelations due to ARCH becomes progressively less serious for progressively higher-ordered autocorrelations, due to the earlier mentioned fact that the "correction factor" tends to unity as $\tau \to \infty$. This is of little value in practice, however, because it is precisely the low-order autocorrelations which are typically calculated. The calculation of twenty sample autocorrelations in the simulations reported above was done with the eventual calculation of Box-Pierce statistics in mind; had fewer sample autocorrelations been calculated, the average deviation from nominal test size would have been substantially larger.

Consider, for example, the ARCH(1) case described above. The reader may verify that:

$$\frac{\gamma_{x^2}(\tau)}{\sigma^4} = \frac{2\alpha_1^\tau}{1-3\ \alpha_1^2},$$

so that the standard error is:

$$\frac{1}{\sqrt{T}}\ (1 + \frac{2\ \alpha_1^\tau}{1-3\ \alpha_1^2})^{1/2}.$$

The corrected and uncorrected confidence intervals are shown in Figure 1 for $\alpha_1 = .5$. Clearly, most of the divergence occurs at the low-order autocorrelations. The deviation from nominal test size is different at each autocorrelation lag, becoming progressively smaller as the lag order gets larger. Thus, to repeat for emphasis, the entries in the first row of Table 1 are very conservative, in the sense that it is not uncommon practice to examine only the first 5 or 10 autocorrelations, which would lead to much higher rejection proportions. This is strongly illustrated in the first row of Table 2, which reports rejection proportions based on only the first 5 sample autocorrelations.

It is of interest to note that the probabilities of type I error may be calculated analytically, as follows. In a Bartlett world,

$$\hat{\rho}_x(\tau) \overset{a}{\sim} N\ (0, \tfrac{1}{T}) = N\ ((0, C_1(T)\ )\ .$$

In reality, however,

$$\hat{\rho}_x(\tau) \overset{a}{\sim} N\ (0, \frac{1}{T}\ (\frac{2\ \alpha_1^\tau}{1-3\ \alpha_1^2})) = N\ (0, C_2(T,\tau))\ .$$

Thus, the probability that $\hat{\rho}_x(\tau)$ exceeds 1.96 Bartlett standard errors of zero is:

$$P\ (|\hat{\rho}_x(\tau)| > 1.96\ \sqrt{C_1(T)})$$

$$= P\ \left[(|\hat{\rho}_x(\tau)|/\sqrt{C_2(T,\tau)}) > 1.96\ \frac{\sqrt{C_1(T)}}{\sqrt{C_2(T,\tau)}} \right]$$

$$= P\ (|Z| > 1.96\ \frac{\sqrt{C_1(T)}}{\sqrt{C_2(T,\tau)}})$$

where Z is a N(0,1) random variable. Since:

$[C_1(T)\ /\ C_2(T,\tau)] < 1$, for all T, τ, it follows that $P(\cdot) > .05$. If $\alpha_1 = .5$ and $T = 500$, for example, the probabilities of type I error are as given in Table 3.

3) On the Existence of EX_t^4

Strictly speaking, the above results require existence of the fourth raw moment of x, μ_4. This is because:

$$\gamma_{x^2}(\tau) = \alpha_1 \gamma_{x^2}(\tau-1) + \ldots + \alpha_p \gamma_{x^2}(\tau-p)$$

with

$$\gamma_{x^2}(0) = EX_t^4 - \sigma^4$$

$$= \mu_4 - \sigma^4.$$

Thus, if μ_4 does not exist (i.e., is infinite) then neither does $\gamma_{x^2}(\tau)$. Milhoj (1985) shows that a necessary and sufficient condition for existence of μ_4 for a pth-order ARCH process is given by:

$$3\,\alpha'\,(I-\Psi)^{-1}\alpha < 1$$

where $\alpha' = (\alpha_1, \ldots, \alpha_p)$ and $\underset{(pxp)}{\Psi}$ is defined by

$\Psi_{ij} = \alpha_{i+j} + \alpha_{i-j}$ where we set

$\alpha_k = 0$ for $k \leqq 0$ and $k > p$.

In actual applications, of course, it is not known whether the condition is satisfied, and the analyst should proceed under the assumption that it is. Even if the true moment of interest has infinite value, the best sample approximation for the purposes of correcting the Bartlett standard errors will still be obtained by following the procedure outlined above.

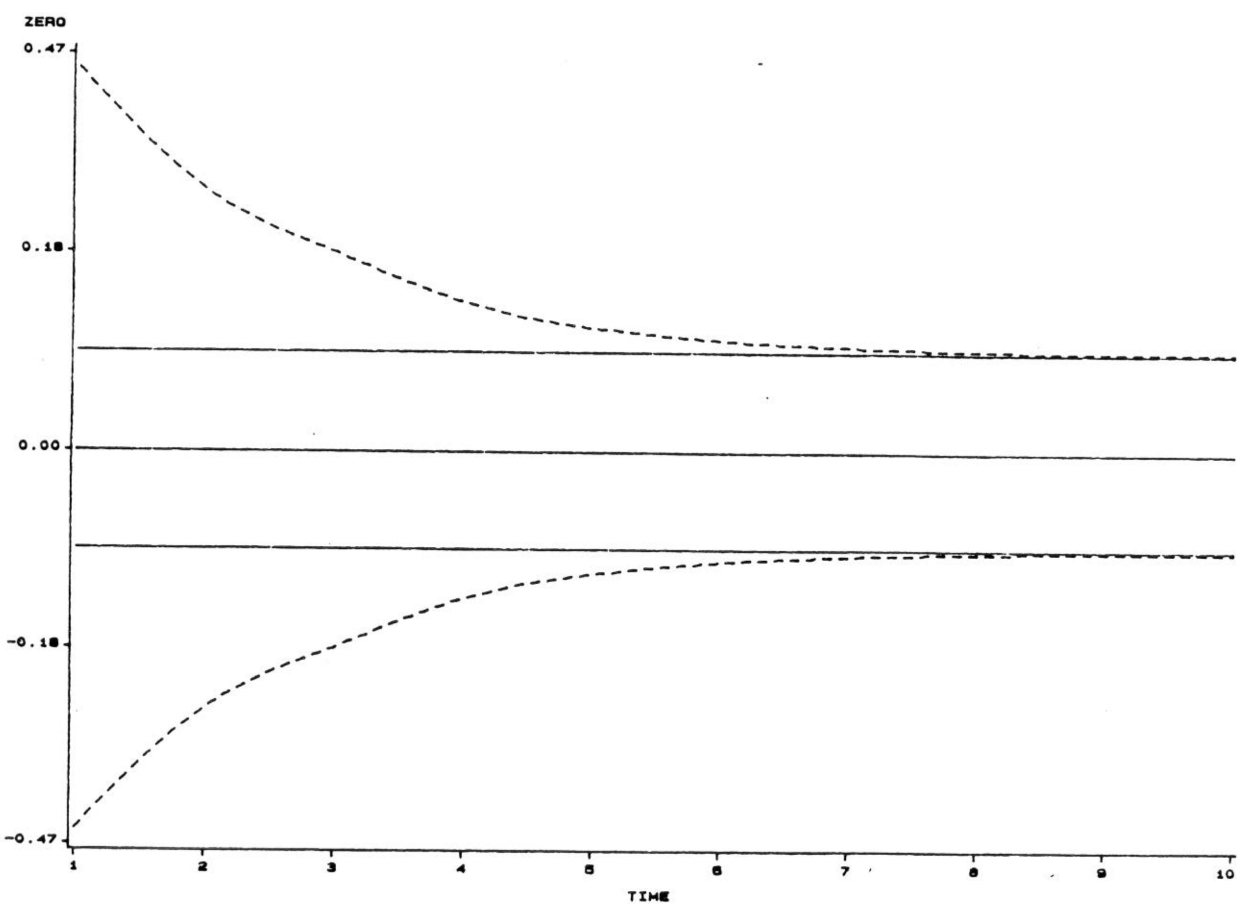
As an example, consider again the ARCH(1) case. Then the existence condition for μ_4 boils down to:

$$\alpha_1 < 1/\sqrt{3} = .577.$$

Thus, in the earlier-tabulated example, the cases of $\alpha_1 = .6$, $.7$, $.8$, and $.9$ all correspond to $\mu_4 = \infty$, yet the ARCH correction continues to work perfectly.

Figure 1

2–SIGMA INTERVALS, CORRECTED AND UNCORRECTED



AUTOCORRELATIONS TO LAG 10

4) The Box-Pierce and Ljung-Box Statistics

The Box-Pierce (1970) serial correlation test statistic (to lag K) is given by:

$$BP(K) = T \sum_{\tau=1}^{K} \hat{\rho}_x^2(\tau).$$

Due to its direct dependence $\hat{\rho}_x^2$, it is also affected by ARCH and must be modified if nominal size is to be maintained. Since under the null of independent white noise we know that:

$$\hat{\rho}(\tau) \xrightarrow{d} N(0,1/T), \quad \tau = 1, 2, 3, \ldots,$$

we have

$$\sqrt{T}\,\hat{\rho}(\tau) \xrightarrow{d} N(0,1).$$

Thus,

$$T\,\hat{\rho}^2(\tau) \xrightarrow{d} \chi_1^2$$

and therefore by asymptotic independence of the sample autocorrelations:

$$T \sum_{\tau=1}^{K} \hat{\rho}^2(\tau) \xrightarrow{d} \chi_K^2,$$ which is the Box-Pierce result.

Under ARCH, however,

$$\rho(\tau) \xrightarrow{d} N\left(0, 1/T\left(1 + \frac{\gamma_{x^2}(\tau)}{\sigma^4}\right)\right).$$

Thus,

$$\{T / (1 + \frac{\gamma_{x^2}(\tau)}{\sigma^4})\}^{1/2}\, \hat{\rho}_x(\tau) \xrightarrow{d} N(0,1),$$

so

$$\{T / (1 + \frac{\gamma_{x^2}(\tau)}{\sigma^4})\}\, \hat{\rho}_x^2(\tau) \xrightarrow{d} \chi_1^2$$

and

$$T \sum_{\tau=1}^{K} \left[\frac{\sigma^4}{\sigma^4 + \gamma_{x^2}(\tau)}\right] \hat{\rho}_\tau^2 \overset{a}{\sim} \chi_K^2.$$

Because the bracketed term is less than or equal to one for all τ, each term in the sum involved in the uncorrected Box-Pierce statistic is "too large," leading to larger than nominal size.

The empirical sizes of the standard and corrected Box-Pierce statistics are shown below in Table 1 (K = 20) and Table 2 (K = 5). Again, the results speak for themselves; the ARCH-corrected statistics perform admirably. It is interesting to note that the very large deviations from nominal size (i.e., much larger than the average deviation of the first 20 sample autocorrelations reported earlier) of the uncorrected Box-Pierce statistics in the presence of ARCH are due to the "cumulation" of errors. This is true regardless of the value of K. Of course, as argued earlier, the problem is made worse as K decreases; this is easily seen by comparing the third rows of Tables 1 and 2.

Similarly, the Ljung-Box (1978) statistic,

$$LB(K) = T(T+2) \sum_{\tau=1}^{K} (T-\tau)^{-1} \rho_x^2(\tau),$$

which is asymptotically equivalent to the Box-Pierce statistic but designed to have better small-sample properties, may be easily corrected for ARCH.

5) Conclusions

In summary, we have shown that the presence of ARCH invalidates the asymptotic distributions of the sample autocorrelations and the Box-Pierce and Box-Ljung test statistics for serial correlation, when computed in the usual fashion. It was shown, both analytically and numerically, that the presence of ARCH renders empirical size much larger than nominal size, leading to spuriously "significant" sample autocorrelations and portmanteu diagnostics. Appropriate correction factors were developed and shown to produce highly satisfactory results, with nominal and empirical sizes being approximately equal.

We have also shown that the error in the Box-Pierce and Box-Ljung statistics, calculated through lag K, is progressively more severe for progressively smaller K. This provides yet another reason, in addition to those given in Box and Pierce (p. 1513) to be wary of test statistics based on small K.

The analysis in the text focused on the case of observed time series. As is well known (Durbin (1970)), the results do not generalize directly to the case of testing for serial correlation in the residuals of estimated models, because the residual autocorrelations are approximately representable as a singular linear transformation of the true disturbance autocorrelations. Box and Pierce (1970) have, however, shown that the dimension of the singularity is equal to d, the degrees of freedom lost in estimating d model parameters. The results remain valid, then, when the statistics are tested against a χ_{K-d}^2 distribution.

Finally, it should be pointed out that the presence of ARCH makes the Bartlett standard errors and the portmanteu tests more conservative; thus, a failure to reject the null of no serial correlation using the uncorrected statistics may be trusted. If the null is rejected, however, and conditional heteroskedasticity is suspected, the corrections should be employed.

Table 1
Empirical Size Results, Box-Pierce Tests
And Bartlett Standard Errors, Based on First 20 Autocorrelations*

α_1 =	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
P	.047	.048	.051	.057	.058	.059	.074	.084	.096	.106
Pc	.048	.048	.048	.051	.046	.046	.049	.048	.047	.044
BP	.053	.052	.063	.074	.095	.127	.215	.280	.378	.429
BPc	.052	.052	.054	.054	.044	.042	.051	.060	.063	.055

* Based on 1000 repetitions

Table 2
Empirical Size Results, Box-Pierce Test
And Bartlett Standard Errors, Based on First 5 Autocorrelations*

α_1 =	0	.1	.2	.3	.4	.5	.6	.7	.8	.9
P	.047	.062	.065	.076	.085	.113	.147	.178	.246	.285
Pc	.049	.054	.051	.048	.046	.050	.046	.042	.049	.047
BP	.049	.066	.074	.112	.151	.213	.299	.366	.523	.610
BPc	.048	.047	.048	.040	.041	.048	.047	.040	.052	.047

* Based on 1000 repetitions

Table 3
Analytic Probabilities of Type-I Error, uncorrected Bartlett Standard Errors

	τ = 1	τ = 3	τ = 5	τ = 10
P (Type I error)	.378	.164	.100	.051

References

Anderson, R.L. ,1942, "Distribution of the Serial Correlation Coefficient" Annals of Mathematical Statistics, 13, 1-13.

Bartlett, M.S., 1946, "On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series" Journal of The Royal Statistical Society, Series B, 8, 27-41.

Box, G.E.P., and G.M. Jenkins, 1970, Time Series Analysis, Forecasting And Control, San Francisco: Holden-Day.

Box, G.E.P., and D.A. Pierce, 1970, "Distribution of The Residual Autocorrelations in ARIMA Time Series Models" Journal of The American Statistical Association, 65, 1509-1526.

Diebold, F.X., 1986, "Temporal Aggregation of ARCH Processes and the Distribution of Asset Returns," Econometrics Working Paper Series, University of Pennsylvania.

Diebold, F.X., C.W.J. Lee, and J. Im, 1986, "A New Approach to the Detection and Treatment of Heteroskedasticity in the Market Model," Working Paper, University of Pennsylvania.

Diebold, F.X. and Nerlove, M., 1985, "ARCH Models of Exchange Rate Fluctuations," Paper presented at the 1985 NBER-NSF Conference on Time-Series Analysis.

Diebold, F.X. and Nerlove, M., 1986, "Factor Structure in a Multivariate ARCH Model of Exchange Rate Fluctuations," Paper presented at the 1986 NBER/NSF Conference on Time-Series Analysis.

Diebold, F.X. and Pauly, P., 1986, "Endogenous Risk in a Rational Expectations Portfolio-Balance Model of the DM/Dollar Rate," Econometrics Working Paper Series, University of Pennsylvania.

Domowitz, I. and Hakkio, C.S., 1986, "Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity with Applications to Exchange Rate Models," Working Paper, Northwestern University.

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.

Durbin, J., 1970, "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables" Econometrica, 38, 410-421.

Engle, R.F., 1982a, "A General Approach to Lagrange Multiplier Model Diagnostics" Journal of Econometrics, 21, 83-104.

Engle, R.F., 1982b, "Autoregressive Conditional Heteroskedasticity With Estimates of The Variance of U.K. Inflation" Econometrica, 50, 987-1008.

Engle, R.F., 1982c, "Estimates of the Variance of US Inflation Based on the ARCH Model" Journal of Money, Credit and Banking, 15, 286-301.

Engle, R.F., Lillien, D.M. and Robins, R.P., 1985, "Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model," Working Paper, UCSD, forthcoming Econometrica.

Godfrey, L.G., 1978, "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables" Econometrica, 46, 1303-1310.

Godfrey, L.G., 1979, "Testing the Adequacy of a Time-Series Model" Biometrika, 66, 67-72.

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

Milhoj, A. ,1985, "The Moment Structure of ARCH Processes," Scandinavian Journal of Statistics, 12, 281-292.

Nemec, A.F.L., 1985, "Conditionally Heteroskedastic Autoregressions," Technical Report #43, Department of Statistics, University of Washington.

Nerlove, M., D.M. Grether, and J.L. Carvalho, 1979, Analysis of Economic Time Series: A Synthesis, New York: Academic Press.

Pantula, S.G., 1985, "Estimation of Autoregressive Models with ARCH Errors," Working Paper, Department of Statistics, North Carolina State University.

Weiss, A.A., 1984, "ARMA Models With ARCH Errors" Journal of Time Series Analysis, 5, 129-143.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48, 817-838.