On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models

Pagan, Adrian and H. Sabau

Working Paper No. 104
October 1987

ON THE INCONSISTENCY OF THE MLE IN CERTAIN HETEROSKEDASTIC REGRESSION MODELS*

A. R. Pagan
University of Rochester

and

H. Sabau
Australian National University
and C.I.D.E., Mexico City

Working Paper No. 104

September, 1987

*We are grateful for comments on an earlier version of this paper by Rob Engle, Ron Gallant, Sastry Pantula and Hashem Pesaran.

1. Introduction

One of the conventions that underlies the general linear model is that the error variance is a constant. Acceptance of this convention in applied work is widespread, possibly because it is difficult to specify any alternative deemed plausible by all. Moreover, it is well known that the OLS estimator remains consistent in the presence of heteroskedasticity, while the GLS estimator also shares this property even if the assumed form of heteroskedasticity is incorrect. Where the effects of unknown heterogeneity in the errors is felt is in the second moment but, as a consequence of work by White (1980) and others, inferences from the OLS and GLS estimators may be made robust to this imperfect knowledge.

These properties make OLS and GLS attractive estimators. But there are a number of cases where OLS and GLS have been by-passed in favor of the maximum likelihood estimator (MLE), because the heteroskedasticity is argued to depend upon the parameters entering the conditional mean of the regression function. Amemiya (1973) studied a model in which the error variance changed as the square of the mean part of the regression function, and his MLE has been made an option in the RATS program. A related approach is the Poisson regression model that has the variance as a linear function of the conditional mean, and this forumlation arises naturally in the analysis of count data models of the type studied in Griliches et al (1984). A final example is the development and use of the ARCH class of models in which the variance is made a function of the square of past errors - Engle (1982).

All of the above have two features in common. First, the heteroskedasticity in the linear model is assumed to be dependent, *inter alia*, upon parameters entering into the conditional mean part of the regression function. Second, estimation is generally performed by maximum likelihood, presumably to gain efficiency by exploiting the connection between the conditional mean and variance parameters. However, as observed by Carroll and Ruppert (1982), this link creates the possibility that the MLE of the conditional mean parameters will be inconsistent if the assumed nature of the heteroskedasticity is invalid. Thus, in a bid to improve efficiency, it is possible that the end result is inconsistency.[1]

Section 2 of this paper examines the factors that would lead to such an inconsistency. For the Amemiya and Poisson regression specifications, Section 3 shows that inconsistency is almost always a consequence of mis-specification. For pure ARCH models, however, the outcome is not as definite, and we eventually find in Section 4 that either the presence of non-normality in the errors or particular types of alternative conditional variances is needed for inconsistency to emerge. As we argue later, however, such alternatives are quite likely in empirical modeling.

[1]There are also models in which the conditional variance is assumed part of the conditional mean e.g. the ARCH-M model of Engle et al (1987). For these, mis-specification of the variance *must* lead to inconsistency in estimators of some of the parameters in the conditional mean.

2. Consistency of the MLE and Specification Error

The model to be analysed is the linear model

$$y_t = x_t\beta+e_t \tag{1}$$

where x_t is a (1xK) vector of weakly exogenous variables and e_t, conditional upon $\mathcal{F}_t$, the sigma field generated by $\{x_{t-j}, z_{t-j}, e_{t-j-1}\}_{j=0}^{\infty}$, is *assumed* normal with zero mean and variance h_t. z_t is a process that would be weakly exogenous to a correctly specified model. Its nature will become clearer later. As the heteroskedasticity represented by h_t may be parameterized in a number of different ways, we simply define the complete vector of parameters to be estimated as θ, a px1 vector, denoting the residual (p-K) parameters as α, i.e. $\theta' = (\beta'\alpha')$.

Under the above assumptions the assumed log likelihood for observed data $\{y_t, x_t\}_{t=0}^{T}$, normalized by the sample size, will be

$$L^{*} = -(1/2T)\log 2\pi-(2T)^{-1}\sum_{t=1}^{T} \log h_t-(2T)^{-1} \sum_{t=1}^{T} h_t^{-1}(y_t-x_t\beta)^2+T^{-1}\log(\mathrm{pdf}(y_0)).$$

(2)

$$= (-1/2T)\log 2\pi + L + T^{-1}\log (\mathrm{pdf}(y_0)) \tag{3}$$

In what follows we ignore the first and last terms in (2), assuming that they are dominated by the middle terms L. The MLE of θ, $\hat{\theta}$, is obtained by solving

$d_\theta(\hat{\theta}) = 0$, where $d_\theta = \partial L/\partial\theta$. If the model is correctly specified it is generally the case that $\hat{\theta} \xrightarrow{p} \theta_o$, the true value of θ_o, and we assume that sufficient regularity attaches to the problem for this to be true. When the model is mis-specified $\hat{\theta}$ is the pseudo-MLE and $\hat{\theta} \xrightarrow{p} \theta^*$, where θ^* is the pseudo-true value of θ, which will be characterized by lemma 1.

<u>Lemma 1</u>: *The pseudo-maximum likelihood estimator* $\hat{\theta}$ *is assumed to converge almost surely to the pseudo-true value of* θ, θ^*, *which is the solution of*

$$E(d_\theta(\theta^*)) = 0 \qquad . \tag{4}$$

where the expectation is taken with respect to the true probability measure. If $\theta^* = \theta_o$, $\hat{\theta}$ *is a consistent estimator under mis-specification.*

□

Exactly what conditions upon $\mathcal{F}_t$ are needed to ensure that Lemma 1 holds will not be detailed here, as it forms the basis of a number of papers by Domowitz and White (1982), Gourieroux et al (1984) and others. It is also clear from the use of the average score that we have ruled out the non-ergodic ARIMA processes as generating mechanisms for x_t. As might be anticipated, following the line of argument in Phillips and Durlauf (1986), coefficients associated with any x_t exhibiting such behaviour can be consistently estimated by MLE under certain types of mis-specification of the heteroskedastic pattern.

Now it is clearly impossible that any model can be mis-specified and yet all parameters be consistently estimated. What is at issue here, however, is the possibility of consistently estimating (by MLE) the sub-vector β_o. For this purpose it is lemma 2 that is of greatest import.

Lemma 2: *If* $d_\beta(\beta_o,\alpha^*) - E(d_\beta(\beta_o,\alpha^*)) \xrightarrow{p} o$ *and* $H_{\theta\theta}(\bar{\theta}) + \mathcal{I}_{\theta\theta}(\bar{\theta}) \xrightarrow{p} o$ *as* $T\to\infty$, *where* $H_{\theta\theta} = -\partial^2 L/\partial\theta\partial\theta'$, $\mathcal{I}_{\theta\theta} = -\lim_{T\to\infty} E(H_{\theta\theta}) > o$, *and* $\bar{\theta} \xrightarrow{a.s} \theta^*$, *a necessary and sufficient condition for* $\hat{\beta}$ *to consistently estimate* β_o *is that* $E(d_\beta(\beta_o,\alpha^*)) = 0$.

Proof: Necessity follows from lemma 1. For sufficiency expand $d_\beta(\hat{\beta},\hat{\alpha}) = 0$ around $d_\beta(\beta^*,\alpha^*)$ to get

$$d_\beta(\hat{\beta},\hat{\alpha}) = 0 = d_\beta(\beta^*,\alpha^*) + H_{\beta\beta}(\bar{\theta})(\hat{\beta}-\beta^*) + H_{\beta\alpha}(\bar{\theta})(\hat{\alpha} - \alpha^*), \qquad (5)$$

where $\bar{\theta}$ lies between θ^* and $\hat{\theta}$. Under the assumptions (5) becomes

$$0 = d_\beta(\beta^*,\alpha^*) - \mathcal{I}_{\beta\beta}(\beta^*,\alpha^*)(\hat{\beta}-\beta^*) - \mathcal{I}_{\beta\alpha}(\beta^*,\alpha^*)(\hat{\alpha}-\alpha^*) + o_p(1). \qquad (6)$$

Since $\hat{\alpha} \xrightarrow{a.s} \alpha^*$, $\hat{\beta} - \beta_o \xrightarrow{p} 0$ provided $\mathcal{I}_{\beta\beta}(\theta^*) > 0$ and $d_\beta(\beta_o,\alpha^*) - E(d_\beta(\beta_o,\alpha^*)) \xrightarrow{p} 0$, $E(d_\beta(\beta_o,\alpha^*)) = 0$ is a sufficient condition as well.

□

We now have to introduce the true form of heteroskedasticity, and this is done by assuming that the density of e_t, conditional upon $\mathcal{F}_t$, is actually $\mathcal{N}(0,\bar{h}_t)$. No precise specification of $\bar{h}_t$ will be provided, but the conditions needed for lemmas 1 and 2 to hold clearly restrict it e.g. it would be necessary that $E(\bar{h}_t) < \infty$, and in certain cases higher order moments of the random variable $\bar{h}_t$ would need to be bounded as well. The assumption of conditional normality means that any inconsistency in the MLE is due to pure mis-specification of the heteroskedasticity i.e. postulating it to be h_t when it is really $\bar{h}_t$, although as noted later density and heteroskedasticity mis-specification interact, and the consequences of one depend critically upon the validity of the other assumption.

For a benchmark, it is useful to begin with the case where h_t has been specified solely as a function of α. Theorem 1 deals with that instance.

Theorem 1: *If* e_t *is conditionally normal with* $E(e_t|\mathcal{F}_t) = 0$, $E(e_t^2|\mathcal{F}_t) = \bar{h}_t$, h_t *is not specified as a function of* β, *and the restrictions on* $\mathcal{F}_t$ *from lemma 2 hold*, $\hat{\beta} \xrightarrow{p} \beta_o$.

Proof: The pseudo-score d_β is

$$d_\beta = T^{-1}\sum_t (y_t - x_t\beta)x_t'h_t^{-1} \tag{6}$$

$$\therefore \quad E(d_\beta(\beta_o,\alpha^*)) = E[E(T^{-1}\sum_t(y_t - x_t\beta_o)x_t'(h_t^*)^{-1}|\mathcal{F}_t)] \tag{7}$$

Since x_t and $h_t^* = h(\mathcal{F}_t,\alpha^*)$ are functions of $\mathcal{F}_t$, and $E(y_t - x_t\beta_o|\mathcal{F}_t) = 0$, (7) is zero and the necessary and sufficient condition of lemma 2 is satisfied.

□

Theorem 1 is the well known result that the GLS estimator (which is identical to the MLE under these circumstances) remains consistent in the presence of mis-specified heteroskedasticity. Its proof makes apparent that such a theorem is unlikely to extend to cases where h_t is made a function of

β. For the wider class of problems Theorem 2 below takes the necessary and sufficient condition of lemma 2 and re-states it in a more useful form for isolating cases where $\hat{\beta}$ will be inconsistent.

<u>Theorem 2</u>: *Under the same conditions on* $\mathcal{F}_t$ *as Theorem 1 and lemma 2, except that* $h_t = h(\mathcal{F}_t, \alpha, \beta)$, $\hat{\beta}$ *is an inconsistent estimator of* β *whenever*

$$\lim_{T\to\infty} E(T^{-1}\sum_t(\bar{h}_t - h_t^*)(\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2})\neq 0. \quad (8)$$

<u>Proof</u>: Differentiating L in (3) with respect to β gives

$$d_\beta = (1/2)T^{-1}\sum_t(h^{-1}_t(y_t - x_t\beta)^2 - 1)(\partial h_t/\partial\beta)h^{-1}_t + T^{-1}\sum_t(y_t - x_t\beta)x_t'h_t^{-1}. \quad (9)$$

Therefore $-E(d_\beta(\beta_o, \alpha^*)) = 0$ iff

$$E\{\lim_{T\to\infty} T^{-1}\sum_t((h_t^*)^{-1}\bar{h}_t - 1)(\partial h_t/\partial\beta(\theta^*))(h_t^*)^{-1}\} = 0,$$

using the properties that $E((y_t - x_t\beta_o)^2|\mathcal{F}_t) = \bar{h}_t$ and $E(y_t - x_t\beta_o|\mathcal{F}_t) = 0$. Then, if $\lim_{T\to\infty} T^{-1}E\{\sum_t(\bar{h}_t - h_t^*)(\partial h_t/\partial\beta(\theta^*))(h_t^*)^{-2}\}\neq 0$, $E(d_\beta(\beta_o, \alpha^*))\neq 0$ and the necessary condition for $\hat{\beta}$ to be consistent is violated. □

The remainder of this paper consists of checking (8) for various specifications of $\bar{h}_t$ and h_t.

3. Consistency of the MLE in the Amemiya and Poisson Models

In this section of the paper the assumed heteroskedasticity, h_t, will be either the form adopted by Amemiya (1973) ($h_t = \alpha(x_t\beta)^2$) or the Poisson regression model ($h_t = x_t\beta$). There have been a number of applications of both of these models, and there has also been concern that the form of the heteroskedasticity implied might be too rigid. In particular, in some applications of the Poisson model there appears to be over- or under-dispersion i.e. the exponent of $x_t\beta$ should not be unity (Cox (1984), Cameron and Trivedi (1985)). In the following analysis therefore the true form of heteroskedasticity will be to set $\bar{h}_t = z_t\gamma$, where z_t is a 1xq vector.

Theorem 3: *If $z_t\gamma \neq x_t\beta$, the MLE of β in the Poisson regression model is generally inconsistent.*

Proof: Evaluating (8) with $h_t = x_t\beta$ and $\bar{h}_t = z_t\gamma$ gives

$$\lim_{T\to\infty} E(T^{-1}\sum_t(z_t\gamma - x_t\beta_o)x_t'(x_t\beta_o)^{-2}) \neq 0$$

$$\text{or } \lim_{T\to\infty} E(T^{-1}\sum_t(x_t'z_t\gamma - x_t'x_t\beta_o)(x_t\beta_o)^{-2}) \neq 0. \qquad (10)$$

Let $\bar{x}_t = (x_t\beta_o)^{-1}x_t$, $\bar{z}_t = (x_t\beta_o)^{-1}z_t$. Then (10) is

$$\lim_{T\to\infty} E(T^{-1}\sum_t(\bar{x}_t'\bar{z}_t\gamma - \bar{x}_t'\bar{x}_t\beta_o))\neq 0.$$

$$\text{or } \lim_{T\to\infty} E(T^{-1}\bar{X}'\bar{Z}\gamma - T^{-1}\bar{X}'\bar{X}\beta_o)\neq 0, \tag{11}$$

where $\bar{X}$ and $\bar{Z}$ are TxK and Txq matrices with $\bar{x}_t$ and $\bar{z}_t$ as t'th rows.

Clearly, since $\bar{X}$ and $\bar{Z}$ do not depend on γ, if (11) was zero for some γ, say γ^*, to remain so for arbitrary γ it would be necessary that the derivative of (11) with respect to γ at $\gamma=\gamma^*$ be zero i.e. $\bar{X}'\bar{Z} = 0$, which will generally not be true. □

The analysis for Amemiya's model is more involved, but the conclusion is essentially the same.

<u>Theorem 4</u>: *If $z_t\gamma \neq \alpha(x_t\beta_o)^2$, the MLE of β in Amemiya's model is generally inconsistent.*

<u>Proof</u>: Substituting $\bar{h}_t = z_t\gamma$, $h_t = \alpha(x_t\beta_o)^2$, $\partial h_t/\partial\beta = 2\alpha x_t' x_t\beta$, (8) becomes

$$\lim_{T\to\infty} E(T^{-1}\sum_t 2[x_t' z_t\gamma - x_t' x_t\beta_o\alpha^*(x_t\beta_o)](\alpha^*)^{-1}(x_t\beta_o)^{-3}) \neq 0$$

which could be written as

$$\lim_{T\to\infty} T^{-1}\sum_t 2\,[\bar{x}_t'\bar{z}_t\gamma - \bar{x}_t'\bar{x}_t\beta_o\alpha^*(x_t\beta_o)](\alpha^*)^{-1}(x_t\beta_o)^{-1} \neq 0 \tag{12}$$

where $\bar{x}_t = (x_t\beta_o)^{-1}x_t$ and $\bar{z}_t = (x_t\beta_o)^{-1}z_t$.

$$= \lim_{T\to\infty} (T^{-1}\sum 2[\bar{x}_t'\bar{z}_t\gamma(\alpha^*)^{-1}(x_t\beta_o)^{-1} - \bar{x}_t'\bar{x}_t\beta_o])\neq 0 \tag{13}$$

For (13) to be zero

$$\lim_{T\to\infty} E(T^{-1} \sum_t (\bar{x}_t'\bar{z}_t\gamma)(x_t\beta_o)^{-1} - \alpha^* \sum_t \bar{x}_t'\bar{x}_t\beta_o) = 0, \tag{14}$$

and this is a system of K equations which generally cannot be satisfied by a single value for α^*. In fact, if β^* is to be β_o, $\hat{\alpha} = T^{-1}\Sigma(x_t\hat{\beta})^{-2}(y_t-x_t\hat{\beta})^2$ and $\alpha^* = T^{-1} \Sigma(x_t\beta_o)^{-2}z_t\gamma = T^{-1}\Sigma\bar{z}_t\gamma(x_t\beta_o)^{-1}$

□

There is one situation in which the value of α^* satisfying (14) is equal to $T^{-1}\Sigma\bar{z}_t\gamma(x_t\beta_o)^{-1}$. If K = 1, without loss of generality β_o can be set to unity, and (14) holds for the pseudo true value $\alpha^* = T^{-1}\Sigma\bar{z}_t\gamma x_t^{-1}$, since $\bar{x}_t = 1$. Of course this is not surprising, as the MLE of β is just the weighted least squares estimator with weights x_t^{-1}. For the more realistic multi-dimensional situation, whilst it is not possible to assert that (14) cannot hold it is very unlikely.

From the results of this section care needs to be exercised when working with either the Poisson or Amemiya-type models of heteroskedasticity, as the risk of inconsistency in the $\hat{\beta}$ seems high. One alternative strategy for dealing with the heteroskedasticity would be to adopt estimators of β, OLS and GLS, which are consistent, and there are semi-parametric GLS estimators of β that are as asymptotically efficient as the MLE yet presume no knowledge of the heteroskedasticity - Robinson (1987), Newey (1986). Hence, these estimators seem very attractive alternatives, although their small sample performance remains to be investigated. At the very least it would seem important for users of these models to compare the MLE of β with a consistent estimator such as OLS. There is a very close connection between this idea and the residual-based tests for over- and under-dispersion considered by Cameron and Trevidi (1985), and we have analysed such "consistency" tests in Pagan and Sabau (1987).

4. Consistency of the MLE in ARCH Models

Engle (1982) argued that it was more appropriate in time series models to assume that the variance of the error term was a function of elements in $\mathcal{F}_t$, than to presume the traditional view that it was constant. Since many economic models come from orthogonality relations that set conditional expectations to zero values, Hansen and Singleton (1982), this is an important observation. Of course, the nature of the conditioning must be made precise for parametric estimation, and Engle suggested that a useful class to consider would be the ARCH(q) process $h_t = \alpha_o + \sum_{j=1}^{q} \alpha_j e_{t-j}^2$. Many applications of this model have been made - Engle and Bolleslev (1986) - but concern has also arisen over whether the class is too restrictive, and a number of alternatives has been proposed in the literature. Weiss (1984) for example estimated patterns of the form $h_t = \alpha_o + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \delta_o(E(y_t|\mathcal{F}_t))^2 + \sum_{j=1}^{r} \delta_j y_{t-j}^2$, and found that the estimates of $\delta_k (k=1,r)$ were frequently non-zero for economic time series.

To fully analyze the consequences of mistakenly taking the heteroskedastic pattern to be ARCH(q) rather than an alternative candidate requires the following lemma.

Lemma 3: *Let ξ be a symmetrically distributed (around zero) absolutely continuous vector of random variables with joint density, conditional upon some sigma field $\mathcal{F}$, $f(\xi)$. Let ψ be a Borel function measurable with respect to $\mathcal{G}$ such that $\mathcal{F}\subset\mathcal{G}$ and $\psi(\xi) = -\psi(-\xi)$ i.e. is conditionally odd in ξ, and assume that $E(\psi(\xi)|\mathcal{F})$ exists. Then $\xi(\psi(\xi)|\mathcal{F}) = 0$.*

Proof: $$\begin{aligned} E(\psi(\xi)|\mathcal{F}) &= \int_{\infty}^{\infty} \psi(\xi)f(\xi)d\xi \\ &= \int_{-\infty}^{0}\psi(\xi)f(\xi)d\xi+\int_{0}^{\infty}\psi(\xi)f(\xi)d\xi \\ &= \int_{-\infty}^{0}\psi(\xi)f(\xi)d\xi+\int_{-\infty}^{0}\psi(-\xi)f(-\xi)d\xi \\ &= \int_{-\infty}^{0}(\psi(\xi)+\psi(-\xi))f(\xi)d\xi \\ &= 0 \end{aligned}$$

from symmetry of the conditional density around zero and $\psi(\xi) = -\psi(-\xi)$ □

As in the preceding section it is necessary to postulate alternative expressions for the true heteroskedasticity, and then to evaluate (8). It is easiest to understand the impact of mis-specification of the variance of e_t upon $\hat{\beta}$ if the nature of $\bar{h}_t$ is allowed to be more general in stages. First, suppose that the true variance $\bar{h}_t$ is an even function of e_{t-1}, conditional upon $\mathcal{F}^{e}_{t-1} =\{e_{t-j}\}^{\infty}_{j=2}$. Theorem 5 proves that the MLE of β is consistent against such an alternative.

Theorem 5: *If $\bar{h}_t$ is an even function of $\{e_{t-j}\}$ $j = 1,\ldots,q$, e_t is symmetrically distributed around zero, conditional upon $\mathcal{F}_t$, the MLE of β in (1), when h_t is assumed to be Engle's ARCH(q) process ($h_t = \alpha_o + \sum_{j=1}^{q} \alpha_j e_{t-j}^2$), is a consistent estimator of β_o.*

Proof: For $\hat{\beta}$ to be consistent it is necessary to show that $E[(\bar{h}_t - h_t^*)$ $\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2}] = 0 \; \forall \; t$. Considering the second term $E(-\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-1}) =$ $\alpha_j^*(\sum_{j=1}^{q} e_{t-j}x_{t-j})(h_t^*)^{-1}$ it will be zero if $E(e_{t-j}x_{t-j}(h_t^*)^{-1}) = 0$ $(j=1,\ldots,q)$. From Engle (1982) h_t^* is an even function of $(e_{t-1},\ldots,e_{t-q})$, making $e_{t-j}x_{t-j}(h_t^*)^{-1}$ an odd function in $(e_{t-1},\ldots,e_{t-j})$. But the joint density for $e_{t-1},\ldots,e_{t-j}$ conditional upon $\mathcal{F}_{t-j}$ can be built up recursively as the product of symmetric densities and so $f(e_{t-1},\ldots e_{t-j}|\mathcal{F}_{t-j})$ is symmetric. From Lemma 1 therefore $E(e_{t-j}x_{t-j}(h_t^*)^{-1}) = 0(j=1,\ldots,q)$. A similar argument holds for $E(\bar{h}_t$ $\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2})$ provided $\bar{h}_t$ is an even function of $e_{t-1},\ldots,e_{t-q}$.

□

Theorem 5 covers some interesting alternatives, most notably if $\bar{h}_t$ is ARCH of order higher than that assumed, if it follows Bollerslev's (1986) GARCH process, i.e. $\bar{h}_t = \delta\bar{h}_{t-1}+\alpha e_{t-1}^2$, or Geweke's (1986) suggestion that log $\bar{h}_t = \alpha_o+\alpha_1\Sigma \log e_{t-1}^2$. Observe that symmetry in the conditional distribution of e_t is quite crucial. Provided the standard ARCH assumption of normality is valid. Theorem 5 provides the MLE of β with a degree of robustness to mis-specification in the variance, which is a comforting result.

Theorem 5 may be extended by regarding $\bar{h}_t$ as composed of two different elements, ϕ_t and ψ_t. ϕ_t will be taken to be a function of $\mathcal{F}_t^x = \{\bar{x}_{t-j}\}_{j=0}^{\infty}$ alone, where $\bar{x}_t$ are those members of x_t *excluding* lagged values of y_t, while ψ_t is an odd function of e_{t-1} conditional upon $\mathcal{F}_{t-1}^e$ and $\mathcal{F}_t^x$.

<u>Theorem 6</u>: *Let $\bar{x}_t$ be strongly exogenous stationary random variables and the true heteroskedasticity be represented by $\bar{h}_t = \phi(\mathcal{F}_t^x,\alpha)+\psi(\mathcal{F}_{t-1}^e,\mathcal{F}_t^x,\alpha)$, where ψ is an odd function of e_{t-1} conditional upon $\mathcal{F}_{t-1}^e$ and $\mathcal{F}_t^x$. If the distribution of e_t, conditional upon $\mathcal{F}_t$, is symmetric around zero, and h_t is invalidly assumed to exhibit Engle's (1982) ARCH(q) process, the MLE of β in (1) is a consistent estimator of β_o.*

<u>Proof</u>: As in Theorem 5 we verify that the necessary and sufficient condition of lemma 2 is satisfied. Substituting for $\bar{h}_t$ in (8) it is necessary that

$$T^{-1}E(\Sigma(\phi_t+\psi_t-h_t^*)\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2}) = 0. \qquad (15)$$

From the proof of Theorem 5 the last of the three terms in (15) is zero. The term $E(\phi_t\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2}) = 0$ for the same reason. (15) therefore holds if $E(\psi_t(\partial h_t/\partial\beta(\theta^*))(h_t^*)^{-2}) = 0$. Because ψ_t is a conditionally odd function of e_{t-1}, it is not possible to apply lemma 3 to the product. However, $h_t = \alpha_o + \sum_{j=1}^{q}\alpha_j e_{t-j}^2$, so that $\partial h_t/\partial\beta = -2\sum_{j=1}^{q}\alpha_j x_{t-j}' e_{t-j}$, making $\partial h_t/\partial\beta\ (\theta^*) = -2\sum_{j=1}^{q}\alpha_j^* x_{t-j}'(y_{t-j}-x_{t-j}\beta_o)$. If it can be shown that $\alpha_j^*=0(j=1,\ldots,q)$, $\psi_t\partial h_t/\partial\beta(\theta^*)(h_t^*)^{-2}$ will be identically zero.

To demonstrate that $\alpha_j^* = 0$ (j=1,...,q) necessitates proving that $\lim_{T\to\infty} E(d_\alpha(\beta_o, \alpha_o^*, \alpha_1^*=0, \ldots, \alpha_q^* = 0)) = 0$, where $\alpha_o^* = \lim_{T\to\infty}(T)^{-1} \sum_{t=1}^{T} E(\phi_t)$ (since the ultimate aim is to show that $\lim_{T\to\infty} E(d_\beta(\beta_o, \alpha_o^*, \alpha_1^*=0, \ldots, \alpha_q^* = 0)) = 0$, this means that $\alpha_j^* = 0$ satisfies $d_\theta(\theta^*) = 0$.

Now $\lim_{T\to\infty} T^{-1}E(d_{\alpha_o}(\beta_o, \alpha_o^*, 0)) = (2T)^{-1}E(\sum_t((h_t^*)^{-1}\bar{h}_t - 1)(\partial h_t/\partial\alpha_o)(\theta^*)(h_t^*)^{-1})$ (16)

where the zero in $d_{\alpha_o}(\cdot)$ represents $\alpha_j=0$(j=1,...,q).

$$= \lim_{T\to\infty}(2T)^{-1} E(\sum_t((\alpha_o^*)^{-1}((\phi_t+\psi_t)-1)(\alpha_o^*)^{-1}). \qquad (17)$$

(17) is zero if $\alpha_o^* = \lim_{T\to\infty}(T)^{-1}\sum E(\phi_t)$ as $E(\psi_t) = 0$ because it is a conditionally odd function of e_{t-1} Examining d_{α_j} we get

$$\lim_{T\to\infty} E(d_{\alpha_j}(\beta_o, \alpha_o^*, 0)) = \lim_{T\to\infty}(2T)^{-1}E(\sum_t((h^*{}_t)^{-1}\bar{h}_t - 1)e_{t-j}^2(h_t^*)^{-1})(j=1,\ldots,q) \qquad (18)$$

$$= \lim_{T\to\infty}(2T)^{-1}E(\sum_t((\alpha_o^*)^{-1}(\phi_t+\psi_t)-1)e_{t-j}^2(\alpha_0^*)^{-1})(j=1,\ldots,q) \qquad (19)$$

But because $\psi_t e_{t-j}^2$ is a conditionally odd function of e_{t-1}

$$= \lim_{T\to\infty}(2T)^{-1}E(\sum_t((\alpha_o^*)^{-1}(\phi_t-1)e_{t-j}^2(\alpha_o^*)^{-1})(j=1,\ldots q) \qquad (20)$$

$$= \lim_{T\to\infty}(2T)^{-1}E(\sum_t((\alpha_o^*)^{-1}E(\phi_t)-1)\sigma^2(\alpha_o^*)^{-1})(\forall j=1,\ldots q) \qquad (21)$$

due to the strong exogeneity of ϕ_t.

$$= 0$$

when $\alpha_o^* = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(\phi_t)$.

Consequently, $\alpha_1^*, \ldots, \alpha_q^*$ are zero and (17) holds, so that the necessary and sufficient condition of Theorem 2 is satisfied making $\hat{\beta}$ consistent.

□

Theorem 6 broadens the range of models that the MLE of β in an ARCH(q) model is robust too, although the heterogeneity described in Theorem 6 may be a rarity. One example however, would be if $\bar{h}_t$ followed the Poisson specification and x_t contained y_{t-1}. Note once again that the assumption of conditional symmetry for the density of the e_t is critical to the outcome, so that it is possible for $\hat{\beta}$ to be inconsistent when $\bar{h}_t$ is conditionally odd in e_{t-1} provided only that the error density is non-symmetric.

Theorem 6 also seems to be of some independent interest since it shows that there exist types of heteroskedasticity that would give zero values for the $\alpha_j^*(j=1,\ldots,q)$, i.e. the ARCH parameter estimates would not reflect this mis-specification at all. In these instances, any ARCH test performed to determine if conditional heteroskedasticity had been accounted for, an option in Hendry's GIVE and Pesarans' DFIT micro-computer packages, would not be powerful, as the deficiency would not be revealed by the estimated values of the ARCH parameters. For robustness of the MLE of β though, this outcome is a good one, as the mis-specification does not contaminate that estimator, provided distributional symmetry for e_t is appropriate.

Finally, the most general type of decomposition of $\bar{h}_t$ would be to add on to ϕ_t and ψ_t above a term η_t that was a conditionally even function of $e_{t-1},\ldots,e_{t-q}$; in many instances it should prove possible to decompose any alternative specification for $\bar{h}_t$ into three such components. For example $\bar{h}_t = \alpha_o+\alpha_1 y_{t-1}^2 = \alpha_o+\alpha_1(x_{t-1}\beta+e_{t-1})^2 = \alpha_o+\alpha_1\beta' x_{t-1}' x_{t-1}\beta+2\alpha_1\beta' x_{t-1}' e_{t-1}+\alpha_1 e_{t-1}^2$ and, if x_t is strongly exogenous, setting $\phi_t = \alpha_o+\alpha_1\beta' x_{t-1}' x_{t-1}\beta, \psi_t = 2\alpha_1\beta' x_{t-1}' e_{t-1}$ and $\eta_t = \alpha_1 e_{t-1}^2$ would define $\bar{h}_t$.

<u>Theorem 7</u>: *If* $\bar{x}_t$ *is strongly exogenous,* $\bar{h}_t = \phi_t+\psi_t+\eta_t$, *where* ϕ_t *and* ψ_t *are as in Theorem 6 while* η_t *is an even function of* $e_{t-1},\ldots,e_{t-q}$ *conditional upon* $\mathcal{F}_{t-q}^e$, *and the other conditions of Theorem 6 are satisfied, the MLE of* β *in (1) is generally an inconsistent estimator of* β_o.

<u>Proof:</u> The proof proceeds by observing that the presence of ψ_t in $\bar{h}_t$ means that $\alpha_1^*\ldots\alpha_q^*$ have to be zero if $\hat{\beta}$ is to be consistent (see the proof of Theorem 6). (17) will then be zero only if α_o^* is $\lim_{T\to\infty} T^{-1}E(\sum_{t=1}^{T}(\phi_t+\eta_t))$ as η_t is a conditionally even function of e_{t-1}, while (20) and (21) will be zero only if $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(((\alpha_o^*)^{-1}(\phi_t+\eta_t)-1)e_{t-j}^2) = 0$ $(j=1,\ldots,q)$. But the value of $\alpha_o^* = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(\phi_t + \eta_t)$ will almost never satisfy this latter requirement as η_t and e_{t-j}^2 $(j=1,\ldots,q)$ are correlated. For example, with $\eta_t=\alpha_1 e_{t-1}^2$, α_o^* from (20) would now involve $E(e_{t-j}^4)$. □

Theorem 7 is a blow against the robustness of the MLE of ARCH models, provided alternatives such as $\bar{h}_t=\alpha_o+\alpha_1 y^2_{t-1}$ are regarded as being plausible alternatives or if it is felt that the presence of conditionally odd and even terms in $\bar{h}_t$ are necessary. In fact there seems to be emerging evidence that this is so for some time series. Nelson (1986) cites Black (1976) and Christie (1982) as showing that positive values of y_{t-1} are associated with a smaller value of $\bar{h}_t$ than negative values are, and he develops a specification for $\bar{h}_t$ that is neither purely conditionally odd nor even to account for financial asset price movements. Weiss (1984) finds that the terms $[E(y_t|\mathcal{F}_t)]^2$ or y^2_{t-1} appear along with an ARCH(q) effect in many of his estimated variances. Since y_t is ARMA in his case, this induces terms such as y^2_{t-1} into the variance specification, whose presence would cause the MLE of β to be inconsistent. Finally, a competing specification to ARCH processes would be random coefficient autoregressions, studied extensively by Nicholls and Quinn (1982), which have terms such as y^2_{t-j} in the variance.

5. Conclusion

In this paper we have demonstrated that the MLE of parameters entering the conditional mean of a regression function can be estimated inconsistently if the presumed heteroskedastic pattern is incorrect. This is in contrast to the OLS and GLS estimators of such parameters, and such an outcome suggests that caution needs to be exercised in the use of the MLE if knowledge of the heteroskedastic pattern is poor. It also stresses the need for good diagnostic tests for specification error in the variance, a point which we have addressed elsewhere (Pagan and Sabau (1987).

Perhaps the most interesting result of the paper was the delineation of circumstances in which Engle's ARCH model would yield a consistent estimator of the conditional mean parameters even if the ARCH specification was invalid. Here it was found that there are types of mis-specification which the MLE would be robust too, although it was argued that there are good reasons for thinking that the most plausible alternatives to an ARCH pattern do not fall into this class. Given the burgeoning applied literature featuring ARCH models, it is imperative that researchers realize that their models could be poorly calibrated if they choose to estimate by MLE under an ARCH specification when in fact the variance exhibits some other form of time dependence. The mis-specification analysis of this paper at least demonstrates the type of alternatives that would create difficulties for the MLE, and therefore checks in that direction would seem sensible.

References

Amemiya, T. (1973), "Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation", *Journal of the American Statistical Association*, 68, 928-934.

Black, F. (1976), "Studies of Stock Market Volatility Changes", *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

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

Carroll, R.J. and D. Ruppert (1982), "A Comparison between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model", *Journal of the American Statistical Association*, 77, 878-882.

Cameron, A.C. and P.K. Trivedi (1985), "Regression Based Tests for Overdispersion", *Technical Report No.* 9, Econometric Workshop, Stanford University.

Christie, A.A. (1982), "The Stochastic Behaviour of Common Stock Variances", *Journal of Financial Economics*, 10, 407-432.

Cox, D.R. (1984), "Some Remarks on Overdispersion", *Biometrika*, 70, 269-274.

Domowitz, I. and H. White (1982), "Misspecified Models with Dependent Observations", *Journal of Econometrics*, 20, 35-58.

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

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

Engle, R.F., D.M. Lillien and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH Model", *Econometrica*, 55, 391-407.

Geweke, J. (1986), "Comment", *Econometric Reviews*, 5, 57-61.

Goureiroux, C., A.Monfort and A. Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 52, 681-700.

Griliches, Z., J. Hausman, and B.H.Holland (1984), "Econometric Models for Count Data with an Application to the Patents R and D Relationship", *Econometrica*, 52, 909-938.

Hansen, L.P. and K. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286.

Nelson, D. (1986), "Time Series Behavior of Stock Market Volatility and Returns", (mimeo, M.I.T.).

Newey, W.K. (1986), "Efficient Estimation of Models with Conditional Moment Restrictions", (mimeo, Princeton University).

Nicholls, D.F. and B.G. Quinn (1982), *Random Coefficient, Autoregressive Models: An Introduction*, (Lecture Notes in Statistics, 11, Springer-Verlarg).

Pagan, A. R. and H. Sabau, "Consistency Tests for Heteroskedastic and Risk Models", (mimeo)

Phillips, P.C.B. and S. Durlauf, (1986), "Multiple Time Series with Integrated Variables", *Review of Economic Studies*, 53, 473-496.

Robinson, P.M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", *Econometrica*, 55, 875-891.

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.

Rochester Center for Economic Research
University of Rochester
Department of Economics
Rochester, NY 14627

1986-87 DISCUSSION PAPERS

WP#33 OIL PRICE SHOCKS AND THE DISPERSION HYPOTHESIS, 1900 - 1980
by Prakash Loungani, January 1986

WP#34 RISK SHARING, INDIVISIBLE LABOR AND AGGREGATE FLUCTUATIONS
by Richard Rogerson, (Revised) February 1986

WP#35 PRICE CONTRACTS, OUTPUT, AND MONETARY DISTURBANCES
by Alan C. Stockman, October 1985

WP#36 FISCAL POLICIES AND INTERNATIONAL FINANCIAL MARKETS
by Alan C. Stockman, March 1986

WP#37 LARGE-SCALE TAX REFORM: THE EXAMPLE OF EMPLOYER-PAID HEALTH INSURANCE PREMIUMS
by Charles E. Phelps, March 1986

WP#38 INVESTMENT, CAPACITY UTILIZATION AND THE REAL BUSINESS CYCLE
by Jeremy Greenwood and Zvi Hercowitz, April 1986

WP#39 THE ECONOMICS OF SCHOOLING: PRODUCTION AND EFFICIENCY IN PUBLIC SCHOOLS
by Eric A. Hanushek, April 1986

WP#40 EMPLOYMENT RELATIONS IN DUAL LABOR MARKETS (IT'S NICE WORK IF YOU CAN GET IT!)
by Walter Y. Oi, April 1986

WP#41 SECTORAL DISTURBANCES, GOVERNMENT POLICIES, AND INDUSTRIAL OUTPUT IN SEVEN EUROPEAN COUNTRIES
by Alan C. Stockman, April 1986

WP#42 SMOOOTH VALUATIONS FUNCTIONS AND DETERMINANCY WITH INFINITELY LIVED CONSUMERS
by Timothy J. Kehoe, David K. Levine and Paul R. Romer, April 1986

WP#43 AN OPERATIONAL THEORY OF MONOPOLY UNION-COMPETITIVE FIRM INTERACTION
by Glenn M. MacDonald and Chris Robinson, June 1986

WP#44 JOB MOBILITY AND THE INFORMATION CONTENT OF EQUILIBRIUM WAGES: PART 1, by Glenn M. MacDonald, June 1986

WP#45 SKI-LIFT PRICING, WITH APPLICATIONS TO LABOR AND OTHER MARKETS
by Robert J. Barro and Paul M. Romer, May 1986, revised April 1987

WP#64 INCREASING RETURNS, SPECIALIZATION, AND EXTERNAL ECONOMIES: GROWTH AS DESCRIBED BY ALLYN YOUNG, By Paul M. Romer, December 1986

WP#65 THE QUIT-LAYOFF DISTINCTION: EMPIRICAL REGULARITIES by Kenneth J. McLaughlin, December 1986

WP#66 FURTHER EVIDENCE ON THE RELATION BETWEEN FISCAL POLICY AND THE TERM STRUCTURE, by Charles I. Plosser, December 1986

WP#67 INVENTORIES AND THE VOLATILITY OF PRODUCTION by James A. Kahn, December 1986

WP#68 RECURSIVE UTILITY AND OPTIMAL CAPITAL ACCUMULATION, I: EXISTENCE, by Robert A. Becker, John H. Boyd III, and Bom Yong Sung, January 1987

WP#69 MONEY AND MARKET INCOMPLETENESS IN OVERLAPPING-GENERATIONS MODELS, by Marianne Baxter, January 1987

WP#70 GROWTH BASED ON INCREASING RETURNS DUE TO SPECIALIZATION by Paul M. Romer, January 1987

WP#71 WHY A STUBBORN CONSERVATIVE WOULD RUN A DEFICIT: POLICY WITH TIME-INCONSISTENT PREFERENCES by Torsten Persson and Lars E.O. Svensson, January 1987

WP#72 ON THE CONTINUUM APPROACH OF SPATIAL AND SOME LOCAL PUBLIC GOODS OR PRODUCT DIFFERENTIATION MODELS by Marcus Berliant and Thijs ten Raa, January 1987

WP#73 THE QUIT-LAYOFF DISTINCTION: GROWTH EFFECTS by Kenneth J. McLaughlin, February 1987

WP#74 SOCIAL SECURITY, LIQUIDITY, AND EARLY RETIREMENT by James A. Kahn, March 1987

WP#75 THE PRODUCT CYCLE HYPOTHESIS AND THE HECKSCHER-OHLIN-SAMUELSON THEORY OF INTERNATIONAL TRADE by Sugata Marjit, April 1987

WP#76 NOTIONS OF EQUAL OPPORTUNITIES by William Thomson, April 1987

WP#77 BARGAINING PROBLEMS WITH UNCERTAIN DISAGREEMENT POINTS by Youngsub Chun and William Thomson, April 1987

WP#78 THE ECONOMICS OF RISING STARS by Glenn M. MacDonald, April 1987

WP#79 STOCHASTIC TRENDS AND ECONOMIC FLUCTUATIONS by Robert King, Charles Plosser, James Stock, and Mark Watson, April 1987

WP#80 INTEREST RATE SMOOTHING AND PRICE LEVEL TREND-STATIONARITY
by Marvin Goodfriend, April 1987

WP#81 THE EQUILIBRIUM APPROACH TO EXCHANGE RATES
by Alan C. Stockman, revised, April 1987

WP#82 INTEREST-RATE SMOOTHING
by Robert J. Barro, May 1987

WP#83 CYCLICAL PRICING OF DURABLE LUXURIES
by Mark Bils, May 1987

WP#84 EQUILIBRIUM IN COOPERATIVE GAMES OF POLICY FORMULATION
by Thomas F. Cooley and Bruce D. Smith, May 1987

WP#85 RENT SHARING AND TURNOVER IN A MODEL WITH EFFICIENCY UNITS OF HUMAN CAPITAL
by Kenneth J. McLaughlin, revised, May 1987

WP#86 THE CYCLICALITY OF LABOR TURNOVER: A JOINT WEALTH MAXIMIZING HYPOTHESIS
by Kenneth J. McLaughlin, revised, May 1987

WP#87 CAN EVERYONE BENEFIT FROM GROWTH? THREE DIFFICULTIES
by Herve' Moulin and William Thomson, May 1987

WP#88 TRADE IN RISKY ASSETS
by Lars E.O. Svensson, May 1987

WP#89 RATIONAL EXPECTATIONS MODELS WITH CENSORED VARIABLES
by Marianne Baxter, June 1987

WP#90 EMPIRICAL EXAMINATIONS OF THE INFORMATION SETS OF ECONOMIC AGENTS
by Nils Gottfries and Torsten Persson, June 1987

WP#91 DO WAGES VARY IN CITIES? AN EMPIRICAL STUDY OF URBAN LABOR MARKETS
by Eric A. Hanushek, June 1987

WP#92 ASPECTS OF TOURNAMENT MODELS: A SURVEY
by Kenneth J. McLaughlin, July 1987

WP#93 ON MODELLING THE NATURAL RATE OF UNEMPLOYMENT WITH INDIVISIBLE LABOR
by Jeremy Greenwood and Gregory W. Huffman

WP#94 TWENTY YEARS AFTER: ECONOMETRICS, 1966-1986
by Adrian Pagan, August 1987

WP#95 ON WELFARE THEORY AND URBAN ECONOMICS
by Marcus Berliant, Yorgos Y. Papageorgiou and Ping Wang, August 1987

WP#96 ENDOGENOUS FINANCIAL STRUCTURE IN AN ECONOMY WITH PRIVATE INFORMATION
by James Kahn, August 1987

WP#97 THE TRADE-OFF BETWEEN CHILD QUANTITY AND QUALITY: SOME EMPIRICAL EVIDENCE
by Eric Hanushek, September 1987

WP#98 SUPPLY AND EQUILIBRIUM IN AN ECONOMY WITH LAND AND PRODUCTION
by Marcus Berliant and Hou-Wen Jeng, September 1987

WP#99 AXIOMS CONCERNING UNCERTAIN DISAGREEMENT POINTS FOR 2-PERSON BARGAINING PROBLEMS
by Youngsub Chun, September 1987

WP#100 MONEY AND INFLATION IN THE AMERICAN COLONIES: FURTHER EVIDENCE ON THE FAILURE OF THE QUANTITY THEORY
by Bruce Smith, October 1987

WP#101 BANK PANICS, SUSPENSIONS, AND GEOGRAPHY: SOME NOTES ON THE "CONTAGION OF FEAR" IN BANKING
by Bruce Smith, October 1987

WP#102 LEGAL RESTRICTIONS, "SUNSPOTS", AND CYCLES
by Bruce Smith, October 1987

WP#103 THE QUIT-LAYOFF DISTINCTION IN A JOINT WEALTH MAXIMIZING APPROACH TO LABOR TURNOVER
by Kenneth McLaughlin, October 1987

WP#104 ON THE INCONSISTENCY OF THE MLE IN CERTAIN HETEROSKEDASTIC REGRESSION MODELS
by Adrian Pagan and H. Sabau, October 1987

To order copies of the above papers complete the attached invoice and return to Christine Massaro, W. Allen Wallis Institute of Political Economy, RCER, 109B Harkness Hall, University of Rochester, Rochester, NY 14627. Three (3) papers per year will be provided free of charge as requested below. Each additional paper will require a $5.00 service fee which must be enclosed with your order. For your convenience an invoice is provided below in order that you may request payment from your institution as necessary. Please make your check payable to the **Rochester Center for Economic Research**. **Checks must be drawn from a U.S. bank and in U.S. dollars**.

W. Allen Wallis Institute for Political Economy

Rochester Center for Economic Research, Working Paper Series

OFFICIAL INVOICE

Requestor's Name ______________________________

Requestor's Address ______________________________

____

____

____

Please send me the following papers free of charge (**Limit: 3 free per year**).

WP# ______ WP# ______ WP# ______

I understand there is a $5.00 fee for each additional paper. Enclosed is my check or money order in the amount of $__________. Please send me the following papers.

WP# ______	WP# ______	WP# ______
WP# ______	WP# ______	WP# ______
WP# ______	WP# ______	WP# ______
WP# ______	WP# ______	WP# ______