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

Adrian R. Pagan
University of Rochester

Hernán Sabau
Operadora de Bolsa

Resumen: Este documento estudia la posibilidad de inconsistencia de los estimadores de máxima verosimilitud para ciertos modelos heteroscedásticos de regresión. Estos incluyen el modelo de regresión de Poisson y los modelos ARCH.

Abstract: This paper studies the possibility of inconsistency of the maximum likelihood estimators for certain heteroskedastic regression models. These include the Poisson regression model and the ARCH models.

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 ordinary least squares (OLS) estimator remains consistent in the presence of heteroskedasticity, while the generalized least squares (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.

* We are grateful to Rob Engle, Ron Gallant, Sastry Pantula and Hashem Pesaran for comments. This paper was written while Sabau was at the Australian National University.

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; this formulation 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. Section 5 draws some conclusions about the advisability of MLE estimation of heteroskedastic models.

2. Consistency of the MLE and Specification Error

The model to be analyzed is the linear model

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

[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.

where x_t is a $(1 \times k)$ vector of weakly exogenous variables and e_t, conditional upon 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 $p \times 1$ vector, denoting the residual $(p - K)$ parameters as α, i.e. $\theta' = (\beta'\alpha')$. At a minimum p would be $(K + 1)$, occurring when h_t was constant.

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

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

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

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 $d_\theta(\hat{\theta}) = \theta$, where $d_\theta = \partial L / \partial\theta$. If the model is correctly specified it is generally the case that $\hat{\theta} \xrightarrow{p} \theta_0$, the true value of θ_0, 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 and will be characterized by Lemma 1.

LEMMA 1. *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*

$$E(d_\theta(\theta^*)) = 0\,, \quad (4)$$

where the expectation is taken with respect to the true probability measure. If $\theta^* = \theta_0$, $\hat{\theta}$ *is a consistent estimator under mis-specification.*

Exactly what conditions upon 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, among others, Domowitz and White (1982) and Gourieroux et al. (1984). 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 expected, in theory 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. In practice one finds that many models are specified such that regressors have been either directly transformed to stationarity by the use of ratios or differences, or have been effectively rendered stationary in estimation by the method of accounting for extensive serial correlation in the error terms; e.g. in Mishkin's (1982) and Barro's and Rush's (1980) work there is close to unit roots in the autoregressive error term.

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_0 . For this purpose it is Lemma 2 that is of greatest import.

LEMMA 2. *If* $d_\beta(\beta_0, \alpha^*) - E(d_\beta(\beta_0, \alpha^*)) \xrightarrow{p} 0$ *and* $H_{\theta\theta}(\bar{\theta}) + I_{\theta\theta}(\bar{\theta}) \xrightarrow{p} 0$ *as* $T \to \infty$, *where* $H_{\theta\theta} = -\partial^2 L / \partial\theta\partial\theta'$, $I_{\theta\theta} = -\lim_{T\to\infty} E(H_{\theta\theta}) > 0$, *and* $\bar{\theta} \xrightarrow{a.s} \theta^*$, *a necessary and sufficient condition for* $\hat{\beta}$ *to consistently estimate* β_0 *is that* $E(d_\beta(\beta_0, \alpha^*)) = 0$.

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

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

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

$$0 = d_\beta(\beta^*, \alpha^*) - I_{\beta\beta}(\beta^*, \alpha^*)(\hat{\beta} - \beta^*) - I_{\beta\alpha}(\beta^*, \alpha^*)(\hat{\alpha} - \alpha^*) + o_p(1) . \tag{6}$$

Since $\hat{\alpha} \xrightarrow{a.s} \alpha^*$, $\hat{\beta} - \beta_0 \xrightarrow{p} 0$ provided $I_{\beta\beta}(\theta^*) > 0$ and

$$d_\beta(\beta_0, \alpha^*) - E(d_\beta(\beta_0, \alpha^*)) \xrightarrow{p} 0, \; E(d_\beta(\beta_0, \alpha^*)) = 0$$

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 F_t , is actually $N(0, \bar{h}_t)$. 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 $E(e_t \mid F_t) = 0$, $E(e_t^2 \mid F_t) = \bar{h}_t$, h_t is not specified as a function of β, and the restrictions on F_t from Lemma 2 hold, $\hat{\beta} \xrightarrow{P} \beta_0$.*

PROOF. The pseudo-score d_β is

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

$$\therefore\ E(d_\beta(\beta_0, \alpha^*)) = E[E(T^{-1}\sum_t (y_t - x_t\beta_0)x_t'(h_t^*)^{-1} \mid F_t)] \ . \tag{7}$$

Since x_t and $h_t^* = h(F_t, \alpha^*)$ are functions of F_t, and $E(y_t - x_t\beta_0 \mid F_t) = 0$, (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.

THEOREM 2. *Under the same conditions on F_t as Theorem 1 and Lemma 2, except that $h_t = h(F_t, \alpha, \beta)$, $\hat{\beta}$ is an inconsistent estimator of β whenever*

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

PROOF. Differentiating L in (3) with respect to β gives

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

Therefore $-E(d_\beta(\beta_0, \alpha^*)) = 0$ iff

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

using the properties that $E((y_t - x_t\beta_0)^2 | F_t) = \bar{h}_t$ and $E(y_t - x_t\beta_0 | F_t) = 0$. Then, if

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

$$E(d_\beta(\beta_0, \alpha^*)) \neq 0$$

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, and 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 $1 \times q$ 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

$$\lim_{T\to\infty} E(T^{-1}\sum_t (z_t\gamma - x_t\beta_0)x_t'(x_t\beta_0)^{-2}) \neq 0$$

$$\text{or } \lim_{T\to\infty} E(T^{-1}\sum_t (x_t'z_t\gamma - x_t'x_t\beta_0)(x_t\beta_0)^{-2}) \neq 0 \; . \quad (10)$$

Let $\bar{x}_t = (x_t\beta_0)^{-1}x_t$, $\bar{z}_t = (x_t\beta_0)^{-1}z_t$. Then (10) is

$$\lim_{T\to\infty} E(T^{-1}\sum_t (\bar{x}_t'\bar{z}_t\gamma - \bar{x}_t'\bar{x}_t\beta_0)) \neq 0$$

$$\text{or } \lim_{T\to\infty} E(T^{-1}\bar{X}'\bar{Z}\gamma - T^{-1}\bar{X}'\bar{X}\beta_0) \neq 0, \quad (11)$$

where $\bar{X}$ and $\bar{Z}$ are $T \times K$ and $T \times q$ 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.

THEOREM 4. *If* $z_t\gamma \neq \alpha(x_t\beta_0)^2$, *the MLE of* β *in Amemiya's model is generally inconsistent.*

PROOF. Substituting $\bar{h}_t = z_t\gamma$, $h_t = \alpha(x_t\beta_0)^2$, $\partial h_t / \partial\beta = 2\alpha\, x_t'x_t\beta$, (8) becomes

$$\lim_{T\to\infty} E(T^{-1}\sum_t 2[x_t'z_t\gamma - x_t'x_t\beta_0\alpha^*(x_t\beta_0)](\alpha^*)^{-1}(x_t\beta_0)^{-3}) \neq 0$$

which could be written as

$$\lim_{T\to\infty} T^{-1}\sum_t 2[\bar{x}_t'\bar{z}_t\gamma - \bar{x}_t'\bar{x}_t\beta_0\alpha^*(x_t\beta_0)](\alpha^*)^{-1}(x_t\beta_0)^{-1} \neq 0 \tag{12}$$

where $\bar{x}_t = (x_t\beta_0)^{-1}x_t$ and $\bar{z}_t = (x_t\beta_0)^{-1}z_t$.

$$= \lim_{T\to\infty} (T^{-1}\sum 2[\bar{x}_t'\bar{z}_t\gamma(\alpha^*)^{-1}(x_t\beta_0)^{-1} - \bar{x}_t'\bar{x}_t\beta_0]) \neq 0 \,. \tag{13}$$

For (13) to be zero

$$\lim_{T\to\infty} E(T^{-1}\sum_t (\bar{x}_t'\bar{z}_t\gamma)(x_t\beta_0)^{-1} - \alpha^*\sum_t \bar{x}_t'\bar{x}_t\beta_0) = 0 \,, \tag{14}$$

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_0, $\hat{\alpha} = T^{-1}\sum(x_t\hat{\beta})^{-2}(y_t - x_t\hat{\beta})^2$ and $\alpha^* = T^{-1}\sum(x_t\beta_0)^{-2}z_t\gamma = T^{-1}\sum\bar{z}_t\gamma\,(x_t\beta_0)^{-1}$. ■

There is one situation in which the value of α^* satisfying (14) is equal to $T^{-1}\sum\bar{z}_t\gamma\,(x_t\beta_0)^{-1}$. If $K=1$, without loss of generality β_0 can be set to unity, and (14) holds for the pseudo true value $\alpha^* = T^{-1}\sum\bar{z}_t\gamma\, x_t^{-1}$, 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 it would not seem a very wise strategy to work with either the Poisson or Amemiya-type models of heteroskedasticity, as the risk of inconsistency in the $\hat{\beta}$ seems high. There are alternative 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 (1986), 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).

4. Consistency of the MLE in ARCH Models

Engle (1982) argued that it was more appropiate in time series models to assume that the variance of the error term was a function of elements in 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 $h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2$. Many applications of this model have been made –Engle and Bollerslev (1986)– but concern has also arisen over whether the class is too restrictive, and a number of alternatives have been proposed in the literature. Weiss (1984) for example estimated patterns of the form $h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \delta_0 (E(y_t|F_t))^2 + \sum_{j=0}^{r} \delta_j y_{t-j}^2$, and found that the estimates of $\delta_k (k = 0, r)$ 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 random variable with density, conditional upon some sigma field $F, f(\xi)$. Let ψ be a Borel function measurable with respect to F such that $\psi(\xi) = -\psi(-\xi)$ i.e. is conditionally odd in ξ, and assume that $E(\psi(\xi) \mid F)$ exists. Then $\xi(\psi(\xi) \mid F) = 0$.*

PROOF.

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

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 $F^e_{t-1} = \{e_{t-j}\}^{\infty}_{j=2}$. 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-1} conditional upon $F^e_{t-1} = \{e_{t-j}\}^{\infty}_{j-2}$, and e_t is symetrically distributed around zero, conditional upon F_t, the MLE of β in (1), when h_t is assumed to exhibit Engle's ARCH(q) process ($h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e^2_{t-j}$), is a consistent estimator of β_0.*

PROOF. From Engle (1982) h_t is a conditionally even function of e_{t-1} while $\partial h_t / \partial\beta$ is a conditionally odd function. Using (8), Lemma 3, and the law of iterated expectations, $E(1/2\sum(\bar{h}_t - h^*_t)\partial h_t / \partial\beta(\theta^*)(h^*_t)^{-2}) = 0$ whenever $\bar{h}_t$ is a conditionally even function of e_{t-1}. ∎

Theorem 5 covers some interesting alternatives, most notably if h_t is ARCH of order higher than that assumed, if it follows Bollerslev's (1986) GARCH process, i.e. $\bar{h}_t = \delta\bar{h}_{t-1} + \alpha e^2_{t-1}$, or Geweke's (1986) suggestion that $\log h_t = \alpha_0 + \alpha_1 \sum \log e^2_{t-1}$. 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 $F^x_t = \{\bar{x}_{t-j}\}^{\infty}_{j=0}$ 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 F^e_{t-1} and F^x_t.

THEOREM 6. *Let x_t be strongly exogenous variables and the true heteroskedasticity be represented by $\bar{h}_t = \phi(F^x_t, \alpha) + \psi(F^e_{t-1}, F^x_t, \alpha)$, where ψ is an odd function of e_{t-1} conditional upon F^e_{t-1} and F^x_t. If the distribution of e_t, conditional upon 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_0.*

PROOF. 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

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

The term $E(\phi_t \partial h_t / \partial\beta(\theta^*)(h_t^*)^{-2}) = 0$ since $\partial h_t / \partial\beta(0^*)$ is an odd function of e_{t-1} conditonal upon F_{t-1}^e and ϕ_t. (15) therefore holds if

$$E(\psi_t(\partial h_t / \partial\beta(0^*))(h_t^*)^{-2}) = 0 \; .$$

Because ψ_t is a conditionally odd function of e_{t-1}, it is not possible to apply Lemma 3 to the product. However,

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 \, ,$$

so that

$$\partial h_t / \partial\beta = -2\sum_{j=1}^{q} \alpha_j x'_{t-j} \; e_{t-j} \, ,$$

making

$$\partial h_t / \partial\beta(0^*) = -2\sum_{j=1}^{q} \alpha_j^* x'_{t-j} \; (y_{t-j} - x_{t-j}\beta_0) \; .$$

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 $\alpha_j^* = 0 \; (j = 1, \ldots, q)$ necessitates proving that

$$\lim_{T\to\infty} E(d_\alpha(\beta_0 \, , \alpha_0^* \, , \alpha_1^* = 0, \ldots , \alpha_q^* = 0)) = 0 \; ,$$

where

$$\alpha_0^* = \lim_{T\to\infty} (T)^{-1} \sum_{t=1}^{T} E(\phi_t)$$

since the ultimate aim is to show that

$$\lim_{T\to\infty} E(d_\beta(\beta_0 \, , \alpha_0^* \, , \alpha_1^* = 0, \ldots , \alpha_q^* = 0)) = 0 \; ,$$

this means that $\alpha_j^* = 0$ satisfies $d_\theta(0^*) = 0$.

Now

$$\lim_{T\to\infty} T^{-1} E(d_{\alpha_0}(\beta_0 \, , \alpha_0^* \, , 0)) = (2T^{-1}) E(\sum_t ((h_t^*)^{-1} \bar{h}_t - 1)(\partial h_t / \partial\alpha_0)(\theta^*)(h_t^*)^{-1}) \quad (16)$$

where the zero in $d_{\alpha_0}(\bullet)$ represents $\alpha_j = 0 \; (j = 1, \ldots , q)$.

$$= \lim_{T\to\infty} (2T)^{-1} E(\sum_t ((\alpha_0^*)^{-1}((\phi_t + \psi_t) - 1)(\alpha_0^*)^{-1}) \; . \quad (17)$$

(17) is zero if $\alpha_0^* = \lim_{T\to\infty} (T)^{-1}\sum E(\phi_t)$ as $E(\psi_t) = 0$ because it is a conditionally odd function of e_{t-1}. Examining d_{α_j} we get

$$\lim_{T\to\infty} E(d_{\alpha_j}(\beta_0, \alpha_0^*, 0)) = \lim_{T\to\infty} (2T)^{-1}E(\sum_t((h_t^*)^{-1}h_t - 1)e_{t-j}^2(h_t^*)^{-1}) \quad (18)$$

$$= \lim_{T\to\infty} (2T)^{-1}E(\sum_t((\alpha_0^*)^{-1}(\phi_t + \psi_t) - 1)e_{t-j}^2(\alpha_0^*)^{-1}) . \quad (19)$$

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

$$= \lim_{T\to\infty} (2T)^{-1}E(\sum_t((\alpha_0^*)^{-1}(\phi_t - 1)e_{t-j}^2(\alpha_0^*)^{-1}) \quad (20)$$

$$= \lim_{T\to\infty} (2T)^{-1}E(\sum_t((\alpha_0^*)^{-1}E(\phi_t) - 1)\sigma^2(\alpha_0^*)^{-1}) \quad , \quad (21)$$

due to the strong exogeneity of ϕ_t.

$$= 0$$

when $\alpha_0^* = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(\phi_t)$.

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 (or more precisely $E(e_t^3 \mid F_t) = 0$) 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}; in many instances it should prove possible to decompose any alternative specification for $\bar{h}_t$ into three such components. For example

$$\bar{h}_t = \alpha_0 + \alpha_1 y_{t-1}^2 = \alpha_0 + \alpha_1 (x_{t-1}\beta + e_{t-1})^2$$

$$= \alpha_0 + \alpha_1 \beta' x_{t-1}' x_{t-1}\beta + 2\alpha_1 \beta' x_{t-1}' e_{t-1} + \alpha_1 e_{t-1}^2$$

and, if x_t is strongly exogenous, setting $\phi_t = \alpha_0 + \alpha_1 \beta' x_{t-1}' x_{t-1}\beta$,

$$\psi_t = 2\alpha_1 \beta' x_{t-1}' e_{t-1} \text{ and}$$

$\eta_t = \alpha_1 e_{t-1}^2$ would define $\bar{h}_t$.

THEOREM 7. *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} conditional upon F_{t-1}^e, and the other conditions of Theorem 6 are satisfied, the MLE of β in (1) is generally an inconsistent estimator of β_0.*

PROOF. 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_0^* is

$$\lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} (\phi_t + \eta_t)\right)$$

as η_t is a conditionally even function of e_{t-1}, while (20) and (21) will be zero only if

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E(((\alpha_0^*)^{-1}(\phi_t + \eta_t) - 1)e_{t-j}^2) = 0 \; .$$

But the value of

$$\alpha_0^* = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \phi_t + \eta_t)$$

will almost never satisfy this latter requirement as η_t and e_{t-j}^2 are correlated. For example, with $\eta_t = \alpha_1 e_{t-1}^2$, α_0^* from (20) would not involve $E(e_{t-j}^4)$. ■

Theorem 7 is a blow against the robustness of the MLE of ARCH models, provided alternatives such as $\bar{h}_t = \alpha_0 + \alpha_1 y_{t-1}^2$ 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 $[E(y_t \mid F_t)]^2$ or y_{t-1}^2 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_{t-1}^2 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_{t-j}^2 in the variance.

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*, vol. 68, pp. 928-934.

Barro, R. J., and M. Rush (1980). "Unanticipated Money and Economic Activity", in S. Fischer (ed.), *Rational Expectations and Economic Policy*, Chicago, University of Chicago Press.

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

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

Carroll, R. J., and D. Ruppert (1982). "A Comparison between Maximum Likelihood and Generalized Least Squares in a Heteroskedastic Linear Model", *Journal of the American Statistical Association*, vol. 77, pp. 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*, vol. 10, pp. 407-432.

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

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

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

———, and T. Bollerslev (1986). "Modelling the Persistence of Conditional Variances", *Econometric Reviews*, vol. 5, pp. 1-50.

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

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

Goureiroux, C., A. Monfort and A. Trognon (1984). "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, vol. 52, pp. 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*, vol. 52, pp. 909-938.

Hansen, L. P., and K. Singleton (1982). "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, vol. 50, pp. 1269-1286.
Mishkin, F. S. (1982). "Does Anticipated Monetary Policy Matter? An Econometric Investigation", *Journal of Political Economy*, vol. 90, pp. 22-51.
Nelson, D. (1986). "Time Series Behavior of Stock Market Volatility and Returns", manuscript, Massachusetts, MIT.
Newey, W. K. (1986). "Efficient Estimation of Models with Conditional Moment Restrictions", manuscript, New Jersey, Princeton University.
Nicholls, D. F., and B. G. Quinn (1982). *Random Coefficient Autoregressive Models: An Introduction*, New York, Springer-Verlag.
Robinson, P. M. (1986). "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form", manuscript, London, London School of Economics.
Weiss, A. A. (1984). "ARMA Models with ARCH Errors", *Journal of Time Series Analysis*, vol. 5, pp. 129-143.
White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, vol. 48, pp. 817-838.