Rochester Center for

Economic Research

Testing for Heteroskedasticity

Pagan, A.R. and Y. Pak

Working Paper No. 315
February 1992

University of

Rochester

Testing for Heteroskedasticity*

A. R. Pagan and Y. Pak

Rochester Center for Economic Research
Working Paper No. 315

Testing for Heteroskedasticity*

A. R. Pagan and Y. Pak

Rochester Center for Economic Research
Working Paper No. 315

September 1991

* We would like to thank Colin Cameron, Bruce Hansen and Pravin Trivedi for their comments on an earlier version of this paper.

1. Introduction

Although mathematical statistics has as its foundation the assumption that random variables are identically and independently distributed, it is generally recognized that the distribution functions may well not be identical. Despite the fact that a failure of this restriction could be due to any moments lacking constancy, it has been the second moment which has attracted the most attention, and which has led to interest in detecting whether an assumption of constancy across units of observation is reasonable. In econometrics a constant variance tends to be referred to as a random variable exhibiting "homoskedasticity" whereas a non–constant one is said to be "heteroskedastic."[1] Sometimes one sees the equivalent term of heterogeneous; certainly the latter description is a more meaningful one for those unfamiliar with econometric rhetoric but the two terms will be used interchangeably in this paper.

Early interest in heteroskedasticity arose from the concerns of users of the linear regression model. It was demonstrated that the ordinary least–squares (OLS) estimator was not efficient if the errors in the regression model were heteroskedastic, and, more seriously, any inferences made with standard errors computed from standard formulae would be incorrect. To obviate the latter problem, methods were derived to make valid inferences in the presence of heteroskedasticity—Eicker (1967), White (1980). To address the former problem it was desirable to perform efficient estimation in the presence of heteroskedasticity. If the heteroskedasticity had a known parametric form the generalized least–squares (GLS) estimator could be invoked; if the form of the heteroskedasticity was unknown, Generalized Least Squares based on a non–parametric estimation of the variance could be implemented along the lines of Carroll (1982) and

---

[1]The argument given by McCulloch (1985) regarding the spelling of heteroskedasticity with a "k" rather than a "c" seems to have been almost universally accepted in econometrics in the last decade.

Robinson (1987). One has the feeling that the development of these procedures has mitigated some of the concern about heteroskedasticity in the basic regression model.

The basic regression model is now only one of the techniques used in econometric work, and the development of powerful packages for PC's such as LIMDEP, SHAZAM, GAUSS, and RATS has resulted in a much wider range of methods of examining data. Analyses of binary data, frequently called "discrete choice models" in econometrics, censored data, "count" data in which the random variable is discrete and takes only a limited number of values, and models in which "volatility" affects the conditional mean of a variable to be explained, as in the ARCH–M model of Engle et al (1987), have the characteristic that heterogeneity is an integral part of the model; or, in terms used later, the heteroskedasticity is **intrinsic** to the model. In these instances, what is of interest is whether the pattern of heteroskedasticity in the data differs from that in the model; if so, it is generally the case that the estimators of the specified model parameters would be inconsistent, and therefore, a re–specified model is called for e.g., see Arabmazar and Schmidt (1981) for an analysis of this for a censored regression model. For such cases, the detection of the correct format for any such heteroskedasticity is of fundamental importance and should be a routine part of any empirical investigation.

This paper aims to provide a review of work done on testing for heteroskedasticity. The basic approach taken is that all existing tests can be regarded as "conditional moment tests" (CM Tests) in the sense of Newey (1985a), Tauchen (1985) and White (1987), with the differences between them revolving around the nature of the moments used, how nuisance parameters are dealt with, and the extent to which a full sample of observations is exploited. Given this view, it is natural to devote Section 2 of the paper to a general discussion of CM tests and complications in working with them. Section 3 proceeds to categorize existing tests for the regression model according to this framework. This section is lengthy, largely because most of the existing literature has concentrated upon the regression model. In fact, there are good reasons for working through the diversity of approaches in this area. Binary, count and censored regression models, dealt with in section 4, can be regarded as specialized non–linear regression models, implying that approaches developed for the linear model will have extensions. What differentiates the models in section 4 is that these exhibit intrinsic heteroskedasticity, and the prime question is whether there is any "extra" heteroskedasticity that is not in the maintained model. Many terms to describe this situation are in use– for example over–dispersion, but it is useful to adopt the descriptor of "extrinsic heteroskedasticity," as this is neutral towards the issue of whether there is "too much" or "too little" heteroskedasticity in the maintained model. Section 4 also contains a discussion of specification testing in volatility models; this material falls into the general framework advanced in this section, because it is likely that a simple model of volatility has already been fitted and the question posed is whether this simple model is an adequate explanation of the data. Section 5 reviews work done on the size and power of test statistics proposed in the literature, while section 6 concludes the paper.

2. Conditional Moment Tests and Their Properties

When models are estimated assumptions are made, either explicitly or implicitly, about the behavior of particular combinations of random variables. Let such a combination be denoted as ϕ_i, $i = 1, \ldots, n$, and assume that the restriction is that $E(\phi_i \mid \mathcal{F}_i) = 0$, where $\mathcal{F}_i$ is some sigma field associated with the random variables.[2] This is a *conditional* moment restriction that is either implied by the model or is used in constructing an estimator to quantify it. It is useful to convert this to an *unconditional* moment restriction by denoting z_i as a $(q \times 1)$ vector of elements

[2]Mostly ϕ_i will be a scalar in what follows, but there is no necessity for that.

constructed from $\mathscr{F}_i$. By the Law of Iterated Expectations it then follows that

$$E(z_i\phi_i) = E(m_i) = 0. \tag{1}$$

Of course it is clear that (1) is not unique, as any non–singular transformation of it satisfies the restriction i.e.,

$$E(Az_i\phi_i) = 0, \tag{2}$$

where A is non–singular. A particularly useful choice of A is $(\Sigma z_i z_i')^{-1}$.

Given that (1) and (2) should hold in the population, it is natural to examine the sample moment(s) $\hat{\tau} = n^{-1}\Sigma z_i\phi_i$ or $\hat{\gamma} = \Sigma Az_i\phi_i$ as a test of this restriction. It is clear from this why choosing $A = (\Sigma z_i z_i')^{-1}$ is helpful, since then $\hat{\gamma}$ will be the regression coefficient of ϕ_i on z_i, whereas $\hat{\tau}$ is the regression coefficient of $z_i\phi_i$ against unity. As emphasized in Cameron and Trivedi (1991), the regression of ϕ_i on z_i allows one to think in traditional terms about "null" and "alternative" hypotheses simply by considering whether γ (the population counterpart to $\hat{\gamma}$) is zero or not. Selecting either $\hat{\tau}$ or $\hat{\gamma}$ it is logical to test if (1) holds by testing if either τ or γ is zero. If m_i does not depend upon any nuisance parameters that need to be estimated, one would expect that $\text{var}(\hat{\tau}) \simeq n^{-2}\,\text{var}(\Sigma m_i) = n^{-2}\,\Sigma\text{var}(m_i) = n^{-2}\,V$, if observations are independently distributed. Using a central limit theorem, $n^{1/2}\,\hat{\tau}$ should be $\mathscr{N}(0, \lim_{n\to\infty} n^{-1}V)$ and $S = n^2\hat{\tau}'V^{-1}\hat{\tau} = (\Sigma m_i)'V^{-1}(\Sigma m_i)$ will be $\chi^2(q)$, where all these distributional statements are meant to hold under (1).[3] Hence a large value of this test statistic, relative to a $\chi^2(q)$ random variable, would be grounds for rejection of (1).

[3] In what follows, m_i will either be $z_i\phi_i$ or $Az_i\phi_i$ as the argument does not depend on the specific format. No attempt is made to spell–out what conditions are needed for central limit theorems etc. to apply, but a recent detailed reference would be Whang and Andrews (1991).

Of course, V is an unknown and the issue of its estimation arises. One possibility is to evaluate $\text{var}(m_i) = E(m_i m_i')$ directly, but that may require some auxiliary assumptions about the density of m_i, or the random variables underlying it, which are not directly concerned with (1). This point has been made very forcibly by Wooldridge (1990) and Dastoor (1990); the latter emphasizes that large values of S might simply reflect a violation of the auxiliary assumptions rather than a failure of (1). Such arguments have led to proposals that the test statistics be made *robust* i.e., dependent on as few auxiliary assumptions as possible, and to this end V is replaced by $\tilde{V} = \Sigma m_i m_i'$.[4] Unfortunately, there appears to be a tradeoff between the desire for robustness and the need to use asymptotic theory, as the test statistic $\tilde{S} = (\Sigma m_i)\tilde{V}^{-1}(\Sigma m_i)$ is likely to converge to a $\chi^2(q)$ slowly, because $\tilde{V}$ is a random variable itself. Much depends on how "random" $\tilde{V}$ is, and that in turn depends on the nature of m_i. If the m_i are highly non–linear functions of the basic random variables, for example being quartic or higher polynomials, $\tilde{V}$ will exhibit a good deal of randomness, and this will result in large departures of $\tilde{S}$ from a $\chi^2(q)$. An example of this problem is in the component of White's (1980) information matrix test focusing on excess kurtosis, which involves the fourth power of a normally–distributed random variable—see Chesher and Spady (1991) and Kennan and Neumann (1988). For tests of heteroskedasticity in the basic regression model the m_i are not highly non–linear functions, and therefore no major difficulties have been reported; however, that situation may be modified when more complex models such as censored regression are thoroughly investigated.

It is worth drawing attention to a potential difference between tests of (1) based on $\hat{\gamma}$ and $\hat{\tau}$.[5] In theory, there should be no difference, but the fact that $\hat{\gamma}$ can be

[4]When m_i is dependent this formula would need to be modified. A range of possibilities is set out in Andrews(1991)

[5]The discussion that follows takes γ and τ as scalars for simplicity.

computed by regressing ϕ_i against z_i leads to the temptation to utilize the t–ratio from such a regression as a test of (1). But the correct t–ratio has to use $\mathrm{var}(\hat{\gamma}) = A^2\mathrm{var}(\Sigma z_i\phi_i)$, whereas the regression program takes it to be $\mathrm{var}(\phi_i)A$. Unless it is known that the $\mathrm{var}(\phi_i)$ is a constant, whereupon $\mathrm{var}(\Sigma z_i\phi_i) = \mathrm{var}(\phi_i)A^{-1}$, to obtain the correct test statistic it is necessary to use the option for finding "heteroscedastic consistent standard errors" built into most econometric packages these days; failure to do so would create the potential for differing conclusions based on $\hat{\tau}$ and $\hat{\gamma}$.

In practice it will rarely be the case that m_i can be regarded as solely a function of random variables. Either ϕ_i or z_i will involve nuisance parameters, θ, that are replaced by estimates $\hat{\theta}$ when computing $\hat{\tau}$ i.e., $\hat{\tau} = n^{-1}\,\Sigma m_i(\hat{\theta})$. This feature may create difficulties in evaluating V. To see that, expand $\Sigma m_i(\hat{\theta})$ around the true value of θ, θ_0, retaining only the first terms

$$\Sigma m_i(\hat{\theta}) \approx \Sigma m_i(\theta_0) + M_\theta(\hat{\theta}-\theta_0), \tag{3}$$

where $M_\theta = E[\Sigma\partial m_i/\partial\theta]'$.[6] Accordingly,

$$\begin{aligned}\mathrm{var}(\Sigma m_i(\hat{\theta})) &= V + \mathrm{cov}\{\Sigma m_i(\theta_0)(\hat{\theta}-\theta_0)'\}M_\theta' + M_\theta\mathrm{cov}\{(\hat{\theta}-\theta_0)\Sigma m_i'(\theta_0)\} \\ &\quad + M_\theta\mathrm{var}(\hat{\theta})M_\theta',\end{aligned} \tag{4}$$

and the appropriate term to substitute for the variance of $\hat{\tau}$ will be $n^{-2}\,\mathrm{var}(\Sigma m_i(\hat{\theta}))$ rather than n^{-2} V. Inspection of (4) shows that the two are equal if $M_\theta = 0$ and, happily, for many tests of heteroskedasticity in the basic regression model, that restriction will hold. It is important to observe that M_θ *is to be evaluated under (1)*,

[6]The symbol $\approx$ is meant to indicate that terms have been neglected that will not affect the asymptotic properties of the left hand side variable after it has been appropriately normalized.

since it is the $\mathrm{var}(\Sigma m_i(\hat{\theta}))$ under that restriction which is desired. Some cases of heteroskedasticity feature $M_\theta \neq 0$, except "under the null hypothesis."

When $M_\theta \neq 0$ a separate computation of $\mathrm{var}(\Sigma m_i(\hat{\theta}))$ is needed and, although fairly easy with packages that have matrix manipulation capabilities, it is unlikely that a regression program can be utilized for the computations. Generally, there will be some set of first–order conditions defining $\hat{\theta}$, say $\Sigma h_i(\hat{\theta}) = 0$, and $(\hat{\theta} - \theta_0) \approx -H_\theta^{-1} \Sigma h_i(\theta_0)$, where $H_\theta = E[\Sigma \partial h_i / \partial \theta]$,[7] making the middle terms in (4) depend on the $\mathrm{cov}\{(\Sigma m_i(\theta_0))(\Sigma h_i(\theta_0)\}$. If this turns out to be zero, $\mathrm{var}(\Sigma m_i(\hat{\theta})) \geq \mathrm{var}\ (\Sigma m_i(\theta_0))$, and any tests performed utilizing V would overstate the true value of the test statistic i.e., result in over–rejection of the hypothesis (1). Unless $h(\cdot)$ is specified however, there is no way of knowing if such "directional" statements might be made. By far the simplest procedure to effect an adjustment is to jointly specify the moment conditions to be used for estimation as $E(m_i - \tau)=0$ and $E(h_i(\theta))=0$. By definition the method of moments solutions to this problem will be $\hat{\tau}$ and $\hat{\theta}$, and the $\mathrm{var}(\hat{\tau})$ would be automatically computed by any program getting such estimates. Notice that with $\hat{\theta}$ and $\hat{\tau}$ as starting values iteration will terminate in one step, so there are no convergence problems.

The literature does contain one instance in which any dependence of $\mathrm{var}(\hat{\tau})$ or $\hat{\theta}$ can be accounted for. This is when $\hat{\theta}$ is estimated by maximum likelihood. Then h_i are the scores for θ, and application of the generalized information equality—Tauchen (1985)—yields $E(\partial m_i / \partial \theta) = -E(h_i m_i')$, $E(\partial h_i / \partial \theta) = -E(h_i h_i')$. After substituting these into (4) it is possible to construct the test statistic as the joint test that the intercepts are all zero in the regression of $m_i(\hat{\theta})$ against unity and $h_i(\hat{\theta})$—see Tauchen (1985), Pagan and Vella (1989). A disadvantage of the proposal is that it introduces randomness into the "denominator" of the test statistics owing to the move from M_θ to

[7]If h is not differentiable H_θ will be replaced by $\Sigma \partial E(h_i)/\partial \theta$, see Whang and Andrews (1991).

$\Sigma h_i m_i'$ etc. Some preliminary evidence in Skeels and Vella (1991) using simulated data from a censored regression model is that $\partial m_i/\partial\theta$ is poorly estimated by $n^{-1}\Sigma h_i m_i'$, and that the estimator deteriorates as the degree of censoring increases. However, provided the inaccuracy does not have too great an impact upon the properties of the test it may be a small price to pay for the convenience of the test.

(1) is a very general statement of what might be tested. Another viewpoint is to conceive of an alternative model to the one being investigated which has q extra parameters γ, with the alternative and basic models coinciding when γ takes values γ^*. By specifying a density or a set of moment conditions γ could be estimated and tested to see if it equals γ^*. However, if γ is not of interest per se, the most likely way to perform such a test is with something like the Lagrange Multiplier (LM) Test or Score Test. In this approach the score for γ, d_i, is evaluated at the MLE of θ and γ, given $\gamma = \gamma^*$, and this is tested for whether it is zero. Formally, this is a special case of (1), as the scores should have a zero expectation under the null hypothesis that $\gamma = \gamma^*$. Accordingly, setting $m_i = d_i$ makes the score test a special case of what has already been discussed. The main advantages of the score test are that it yields a very precise moment condition (1) and it also produces a test with optimal properties if the density it is based on is correct; its principal disadvantage is that it introduces an auxiliary assumption pertaining to densities that may be invalid, and such a circumstance would cause it to lose its optimality properties.[8] Nevertheless, the score test is very useful as a benchmark, in that it can suggest suitable moment conditions

[8]The LM test has been shown to possess a number of optimal properties. For local alternatives, it has maximal local power in the class of chi–square criteria, and in some situations, may be the locally best invariant test— see King and Hillier (1985). The situations where this is true center largely on the regression model in which a single parameter entering the covariance matrix of u_i is tested for being zero against the alternative that it is positive. In fact it is not the LM test **per se** which has this property but the one–sided version of it.

which may be modified to allow for unknown densities; a further discussion on this point is given in Section 3.

The presence of nuisance parameters in m_i may actually be converted into an advantage rather than a disadvantage. Suppose that there are two possible estimators of θ, $\hat{\theta}$ and $\tilde{\theta}$, which are the solutions of $\Sigma h_i(\hat{\theta}) = 0$ and $\Sigma g_i(\tilde{\theta}) = 0$ respectively, and that both $\hat{\theta}$, $\tilde{\theta}$ are consistent if what is being tested is true, while they converge to different values for θ if it is false. A comparison of $\hat{\theta}$ with $\tilde{\theta}$ i.e., forming $\psi = \hat{\theta}-\tilde{\theta}$, as recommended in Hausman (1978), enables one to form a test statistic that the hypothesis being tested is true, since ψ will only be close to zero if this is so. Asymptotically, an exactly equivalent procedure is to test if $E(h_i(\theta)) = 0$ using $\Sigma h_i(\tilde{\theta})$ (or $E(g_i(\theta)) = 0$ using $\Sigma g_i(\hat{\theta})$), since $\Sigma h_i(\tilde{\theta}) \approx \Sigma h_i(\hat{\theta}) + H_\theta(\tilde{\theta}-\hat{\theta}) = H_\theta(\tilde{\theta}-\hat{\theta})$, and, if H_θ is non–singular, the tests must give the same outcome. Hence, one can define m_i as g_i and use $\hat{\theta}$ from $\Sigma h_i(\hat{\theta}) = 0$ to produce a moment test. As seen in later sections, such an idea has been a popular way to test for heteroskedasticity.

Ultimately, we are interested in testing for heteroskedasticity in various contexts. If data is ordered either chronologically or by some variable such as (say) firm size, then it is natural to think of heteroskedasticity as involving structural change in whatever constitutes the scale parameter of the model.[9] Defining $E(m_i)=0$ as the moment condition used to estimate this scale parameter, it is therefore reasonable to test for heteroskedasticity by examining the cumulative sums $\sum_{i=1}^{nk} m_i(\hat{\theta})$, where k is a fraction of the sample and θ includes both the scale parameter (θ_1) as well as any others (θ_2) which form part of the model. Since θ is being estimated from the same data as is being used for specification testing, some allowance has to be made for that

[9]To apply the theory that follows the heteroskedasticity cannot arise from the variable by which the data is ordered, since that variable now has a trend associated with it and the asymptotic theory being invoked explicitly rules out the possibility of trending variables.

fact in determining the $\mathrm{var}(\sum_{i=1}^{nk} m_i(\hat{\theta}))$. It will be assumed that $E(\partial m_i/\partial\theta_2)=0$, as this can be shown to be true for estimators of the scale parameter used later, and such a restriction therefore means that $(\hat{\theta}_1-\theta_1)$ asymptotically behaves like $-(\sum_{i=1}^{n} \partial m_i/\partial\theta_1)^{-1}(\sum_{i=1}^{n} m_i)$.

Linearizing $\sum_{i=1}^{nk} m_i(\hat{\theta})$ around θ and applying the assumptions just made gives

$$\sum_{i=1}^{nk} m_i(\hat{\theta}) \approx \sum_{i=1}^{nk} m_i(\theta) + [\sum_{i=1}^{nk} (\partial m_i/\partial\theta_1)](\hat{\theta}_1-\theta_1) \tag{5}$$

$$\approx \sum_{i=1}^{nk} m_i(\theta) + [\sum_{i=1}^{nk} (\partial m_i/\partial\theta_1)]\ (-\sum_{i=1}^{n} (\partial m_i/\partial\theta_1))^{-1}(\sum_{i=1}^{n} m_i(\theta)) \tag{6}$$

$$= \sum_{i=1}^{nk} m_i(\theta) - (nk/n)[(nk)^{-1}\sum_{i=1}^{nk} (\partial m_i/\partial\theta_1)][n^{-1}\sum_{i=1}^{n} (\partial m_i/\partial\theta_1)](\sum_{i=1}^{n} m_i(\theta)) \tag{7}$$

$$\approx \sum_{i=1}^{nk} m_i(\theta) - k(\sum_{i=1}^{n} m_i(\theta)), \tag{8}$$

as $(nk)^{-1}\sum_{i=1}^{nk} (\partial m_i/\partial\theta_1) - n^{-1}\sum_{i=1}^{n} (\partial m_i/\partial\theta_1)$ should be $o_p(1)$ for large enough n and for fixed k as they both estimate $E(\partial m_i/\partial\theta_1)$, which is constant under the null hypothesis. Consequently, the $\mathrm{var}(\sum_{i=1}^{nk} m_i(\hat{\theta}))=\mathrm{var}\{\sum_{i=1}^{nk} m_i(\theta) - k\sum_{i=1}^{n} m_i(\theta)\}$. To evaluate this variance assume that $m_i(\theta)$ is i.i.d. with variance v; the same results hold if it is dependent and Hansen (1990) has a formal proof of that fact. Then

$$\operatorname{var}\{\sum_{i=1}^{nk} m_i(\theta) - k\sum_{i=1}^{n} m_i(\theta)\} = \operatorname{var}(\sum_{i=1}^{nk} m_i(\theta)) - 2k\operatorname{cov}\{[\sum_{i=1}^{nk} m_i(\theta)][\sum_{i=1}^{n} m_i(\theta)\} + k^2\operatorname{var}(\sum_{i=1}^{n} m_i(\theta)) \tag{9}$$

$$= (nkv) - 2k(nkv) + k^2 nv \tag{10}$$

$$= (nkv) - k^2(nv). \tag{11}$$

Defining $V(k) = \operatorname{var}(\sum_{i=1}^{nk} m_i(\theta)) = nkv$, (11) can be expressed as

$$= V(k) - k^2 V(1) \tag{12}$$

$$= V(k) - V(k)V(1)^{-1}V(k) \tag{13}$$

using the fact that $k = V(k)/V(1)$. Consequently, $\operatorname{var}(\Sigma m_i(\hat{\theta})) \approx V(k) - V(k)V(1)^{-1}V(k)$ and $V(k)$ can be estimated by $\hat{V}(k) = \sum_{i=1}^{nk} m_i^2(\hat{\theta})$. Using these pieces of information the following CUSUM test can be constructed:

$$C(k) = (\sum_{i=1}^{nk} m_i(\hat{\theta}))^2 / [\hat{V}(k) - \hat{V}(k)\hat{V}(1)^{-1}\hat{V}(k)] \tag{14}$$

For any k this will be asymptotically a $\chi^2(1)$ if the model is correctly specified.

A number of different tests can be associated with this approach.

(a) Let $k=(1/n)$, $(2/n)$, etc. and define $SC = \sup_k C(k)$ i.e., look at the maximum of $C(k)$ for all values of k from $(1/n)$ to $(n-1)/n$ in increments of $(1/n)$. By definition of

$\hat{\theta}_1$, $\sum_{i=1}^{n} m_i(\hat{\theta}_1)=0$, forcing one to find the sup of C(k) over a restricted range; Andrews (1990) suggests .15 to .85.

(b) $L_W = n^{-1}\sum_{j=1}^{n-1} C(j/n)$. This is an average test. The L comes from the fact that, if the scale parameter is estimated by ML and m_i are therefore the scores with respect to it, the statistic can be thought of as the LM test for the hypothesis that var $(\nu_i)=0$ in the model $\theta_{1i}=\theta_{1i-1}+\nu_i$ (see Hansen (1990)). The distribution is non–standard and is tabulated in Hansen's paper.

(c) $L_C = n^{-1}V(1)^{-1}\sum_{j=1}^{n-1} [\sum_{i=1}^{j} \hat{m}_i]^2$.

This is also an LM test like (b), differing in the covariance matrix assumed for ν_i under the alternative. Hansen also tabulates the distribution of this test. On the basis of simulation studies (for testing constancy of location parameters) he finds it has better correspondence with asymptotic theory than (b) does. Probably this is because one is using less random elements in the denominator of the statistic. The more randomness one induces into denominators of test statistics the slower their convergence to limiting distributions tends to be. In some cases it can be very slow indeed, and leads to substantial over–rejections e.g. Chesher and Spady (1991).

3. Testing Heteroskedasticity in the Regression Model

3.1 Testing Using General Moment Conditions

The basic regression model is

$$y_i = x_i'\beta + u_i, \tag{15}$$

where x_i is a ($p\times1$) vector and u_i will be assumed independently distributed $(0,\sigma_i^2)$, with σ_i^2 being a function of some variable summarized by the field $\mathscr{F}_i$. Initially, x_i will be taken to be weakly exogenous and y_i will be a scalar. If the errors are to be homoskedastic, $E(z_i(\sigma^{-2}u_i^2 - 1)) = E(z_i\phi_i)$ will be zero, where $\sigma_i^2 = \sigma^2$ under the assumption that there is no heteroskedasticity, while z_i is a (qx1) vector drawn from $\mathscr{F}_i$ possessing the property $E(z_i) = 0$. [10] The set of moment conditions $E(z_i\phi_i) = E(m_i) = 0$ will therefore be used to test for heteroskedasticity, and this was the structure set out in (1).

There are obviously many tests for heteroskedasticity that may be generated by selection of z_i, and this was the theme of Pagan and Hall (1983). Examples would be $z_i = (x_i'\beta)^2$ used in MICROFIT (Pesaran and Pesaran (1987)); $z_i = (y_{i-1} - x_{i-1}'\beta)^2$, the test for first-order Autoregressive Conditional Heteroskedasticity (ARCH) introduced in Engle (1982); $z_i = \mathrm{vec}(x_i \otimes x_i')$ (excluding any redundant elements), used in White (1980); and $z_i = i$, which was suggested by Szroeter (1978), and applied by many others after him. All of these tests can be constructed by regressing ϕ_i upon z_i, and all the issues concerning robustness and dependence on nuisance parameters set out in Section 2 apply. Perhaps the issue that occupied most attention in Pagan and Hall was the latter one. As argued in Section 2, a necessary condition enabling one to ignore the fact that $\theta' = (\beta'\ \sigma^2)$ is estimated rather than known, would be $E(\partial m_i/\partial\theta) = 0$.[11] From the definition of m_i, $\partial m_i/\partial\theta = (\partial z_i/\partial\theta)\phi_i + z_i(\partial\phi_i/\partial\theta)$. Both of these two terms will have expectation of zero; the first because $E(\phi_i | \mathscr{F}_i) = 0$ and

[10] Although, z_i will typically be a function of β, this dependence will be suppressed unless its recognition is important. The assumption that $E(z_i) = 0$ means that whatever is selected to represent z_i will need to be mean corrected.

[11] Because z_i has been assumed to have zero mean, there is implicitly another parameter, $E(z_i)$, that needs to be estimated. However, it is easily seen from the same argument as used later for β and σ^2 that it does not affect the distribution of $\hat{\tau} = n^{-1}\Sigma\hat{m}_i$.

$\partial z_i/\partial\theta \in \mathcal{F}_i$; the second because it equals either $E(\sigma^{-2}u_i x_i z_i') = 0$ (for β) from $E(u_i | \mathcal{F}_i) = 0$, or $-E(z_i\sigma^{-4}u_i^2) = -E(z_i)\sigma^{-2} = 0$ (for σ^2), given the assumption $E(z_i) = 0$. Consequently, when testing for heteroskedasticity in the regular regression model, there will be no dependence of $\hat{\tau} = n^{-1}\Sigma\hat{m}_i$ upon $\hat{\beta}$ or $\hat{\sigma}^2$, and the variance is therefore simply found from $\mathrm{var}(m_i)$.

When x_i is not weakly exogenous i.e., some members of x_i are endogenous variables, it will no longer be true that $E(\partial m_i/\partial\theta) = 0$, since $E(\sigma^{-2}u_i x_i z_i') \neq 0$. In these circumstances, an allowance must be made for the estimation of β. The simplest procedure would be to jointly estimate τ and β and to then test if τ takes the value zero. An alternative is to use the variance given in (4), or to explicitly evaluate it in the simultaneous equation context as done in Pagan and Hall (1983, p. 192–194). There are papers in the literature which claim that the distribution does not depend on $\hat{\beta}$—Szroeter (1978) and Tse and Phoon (1985)—but the assumptions made to get this result imply a degeneracy in the reduced form errors which has to be regarded as implausible; a fuller account of this point is contained in Pagan and Hall (1983, p. 195).

Another choice of moment condition that is slightly more complex leads to the Goldfeld–Quandt (1965) test. Essentially they compare the residual variance estimated over a sub–period $i = 1, \ldots, n_1$ with that over $i = n_1 + k, \ldots, n$, with k being the number of observations dropped.[12] One would effect such a comparison by making $m_i = z_i(y_i - x_i'\beta)^2$, with z_i being n_1^{-1} for $i = 1, \ldots, n_1$; $z_i = 0$ for $i = n_1 + 1, \ldots, n_1 + k$; and $z_i = n_2^{-1}$ for $i = n_1 + k+1, \ldots, n$, where $n_2 = n-(n_1+k)$. In fact, because Goldfeld and Quandt estimate β separately for both $i = 1, \ldots, n_1$ and $i = n_1+k+1, \ldots, n$, there are really two sets of parameters β_1 and β_2 in θ, with moment conditions for

[12]In fact, they use the ratio of these variances and order the observations in ascending order; the difference in variances is used here so that $E(m_i) = 0$ when there is no heteroskedasticity.

estimating these being $E[\sum_{i=1}^{n_1} x_i(y_i - x_i'\beta_1)] = 0$, $E[\sum_{i=n_1+k+1}^{n} x_i(y_i - x_i'\beta_2)] = 0$. However, Goldfeld and Quandt did not intend that β_1 and β_2 be different; they were simply trying to make the estimated residual variances independent of one another, and this could not be done if β is estimated utilizing the full sample. But this choice of method to estimate β should not be allowed to disguise the fact that, underlying Goldfeld and Quandt's test, is a very specific moment condition. It would be possible to generalize the Goldfeld–Quandt test to allow for more than one break. Doing so with three contiguous breaks would produce a test statistic emulating Bartlett's (1937) test for homogeneity, popularized by Ramsey (1969) as BAMSET after the OLS residuals are replaced by the BLUS residuals. It is not possible to express Bartlett's test exactly as a conditional moment test, but asymptotically it is equivalent to one in which the z_i are defined as for the Goldfeld Quandt test over the three periods.

Some of the testing literature is concerned with robustness. Finding the variance of m_i could be done by writing $\text{var}(m_i) = \text{var}(z_i)\text{var}(\phi_i) = \text{var}(z_i)\sigma^{-4}\text{var}(u_i^2-\sigma^2)$, after which a distributional assumption concerning u_i would allow $\text{var}(m_i)$ to be quantified exactly. However, a distributional assumption for u_i must be an auxiliary one, and it is not directly connected with a test for heteroskedasticity. Consequently, as discussed in Section 2, it may be desirable to estimate $\text{var}(m_i)$ without making distributional assumptions, in particular, $n^{-1}\Sigma m_i m_i' = n^{-1}\Sigma z_i z_i'\phi_i^2$ could be adopted as the estimate. Alternatively, if $E(\phi_i^2)$ is a constant one might use $(n^{-1}\Sigma\phi_i^2)(n^{-1}\Sigma z_i z_i')$. This was Koenker's (1981) criticism of the test for heteroskedasticity introduced by Breusch and Pagan (1979) and Godfrey (1978), and his version using the second formulation above has become the standard way of implementing a test which stems from the moment condition $E(z_i\phi_i) = 0$.[13] It appears that there can be some major differences in tests

[13]There can be some dangers to this strategy in that one is attempting to estimate $E(u_i^4)$

constructed with different estimates of var(ϕ_i). In Monte Carlo experiments Kamstra (1990) finds that the Koenker variant which adjusts the size of the test using the outer product $n^{-1}\Sigma m_i m_i'$, leads to severe over–rejection.

3.2 Testing Using the Optimal Score

Although the treatment of tests according to the selection of z_i and the extent to which robustness is addressed yields a satisfactory taxonomy, it does not address either the question of the optimal choice of z_i, or the possibility that the distribution of u_i might affect the nature of the moment condition itself. One resolution of this lacuna is to derive the Lagrange Multiplier or score test. To this end, let the density of $\epsilon_i = \sigma_i^{-1}u_i$ be $f(\epsilon_i)$, where the ϵ_i are identically and independently distributed random variables. The log–likelihood of (15) will therefore be

$$L = -1/2\ \Sigma\log\sigma_i^2 + \Sigma\log f(\sigma_i^{-1}(y_i - x_i'\beta)). \tag{16}$$

If γ are parameters such that $\gamma = \gamma^*$ makes $\sigma_i^2 = \sigma^2$, the scores for γ are the basis of the LM test, and these will be

---

robustly. If $E(u_i^4)$ did not exist, Koenker's test might not even be consistent whereas the Breusch/Pagan/Godfrey test would be as it uses $2\sigma^4$ as the variance, i.e only a second moment is used as a divisor. Phillips and Loretan (1990) make this observation in connection with recursive tests of the sort to be discussed in section 3.4. The problem is likely to be particularly acute when one is trying to make tests robust to ARCH errors as the conditions for the existence of a fourth moment in u_i are much more stringent than for a second moment. The problem will only affect the power of such robust tests as it is under the alternative that the moments may not exist. One of Kamstra's (1990) experiments had an ARCH(1) model for which the robust test had very poor power properties. When the density of $\sigma_i^{-1}u_i$ is $\mathcal{N}(0,1)$ the fourth moment fails to exist if the ARCH(1) parameter exceeds .57 and, although Kamstra sets the true parameter to .4, there may be some small sample impact of being near the boundary.

$$\partial L/\partial\gamma = -1/2\ \Sigma\sigma_i^{-2}\partial\sigma_i^2/\partial\gamma - (1/2)\ \Sigma f_i^{-1}(\partial f_i/\partial\epsilon_i)\sigma_i^{-2}(\partial\sigma_i^2/\partial\gamma)\epsilon_i. \tag{17}$$

After re–arrangement and simplification,

$$\partial L/\partial\gamma = 1/2\ \Sigma(\partial\sigma_i^2/\partial\gamma)\sigma_i^{-2}[-\psi_i\epsilon_i - 1], \tag{18}$$

where $\psi_i = f_i^{-1}(\partial f_i/\partial\epsilon_i)$. Under the null hypothesis H_0: $\gamma = \gamma^*$,

$$\partial L/\partial\gamma|_{\gamma=\gamma^*} = 1/2\ \sigma^{-2}\Sigma(\partial\sigma_i^2/\partial\gamma)_{\gamma=\gamma^*}[-f_i^{-1}(\partial f_i/\partial\epsilon_i)\epsilon_i - 1]. \tag{19}$$

Studying (19) it is apparent that the "optimal" choice of z_i should be $(\partial\sigma_i^2/\partial\gamma)_{\gamma=\gamma^*}$, and that the nature of the distribution of ϵ_i impinges directly upon the "optimal" test. For moment conditions having the structure $E(z_i\phi_i)=0$, the best choice of ϕ_i is $-f_i^{-1}(\partial f_i/\partial\epsilon_i)\epsilon_i - 1$, constituting a non–linear function of $\epsilon_i=\sigma^{-1}u_i$ that depends directly upon the density of ϵ_i. Interestingly enough, ϕ_i will only be $(\sigma^{-2}u_i^2 - 1)$ if $f(\bullet)$ is the standard normal density; in that instance $f_i^{-1}(\partial f_i/\partial\epsilon_i) = -\epsilon_i$, revealing that the moment conditions $E[z_i(\sigma^{-2}u_i^2 - 1)] = 0$ implicitly have the assumption of a Gaussian density for ϵ_i underlying their construction. Notice that $E[z_i(\sigma^{-2}u_i^2 - 1)] = 0$ regardless of the nature of the density; it will only be the **power** of the test statistic that is affected by not allowing ϕ_i to vary according to $f(\cdot)$.

Consideration of (19) points to the fact that there are two issues in devising an appropriate test for heteroskedasticity in the regression model. The first of these is how to approximate $(\partial\sigma_i^2/\partial\gamma)_{\gamma=\gamma^*}$; the need to form a derivative of σ_i^2 emphasizes that the alternative has a major role to play in determining what z_i will be. If σ_i^2 has the "single index" form $\sigma_i^2 = g(z_i'\gamma)$, where g is some function, then $(\partial\sigma_i^2/\partial\gamma)|_{\gamma=\gamma^*}$ is g_1z_i, with g_1 being the derivative of g. Setting $\gamma = \gamma^*$, $g(z_i'\gamma^*)$ must be a constant, σ^2, if there is to be no heteroskedasticity, and g_1 will therefore be constant under the null hypothesis. In these circumstances, $(\partial\sigma_i^2/\partial\gamma)|_{\gamma=\gamma^*} = g_1^* z_i$, and the constancy of g_1^* enables it to be eliminated, leaving the appropriate moment condition as $E(z_i\phi_i) = 0$ i.e., the test statistic is invariant to g_1^* and is the same as if σ_i^2 was the linear function $z_i'\gamma$. This was the observation in Breusch and Pagan (1979). It is important to emphasize however, that the result depends critically on the single index format, and it is not true that the test statistic is invariant to general types of heteroskedasticity, an interpretation sometimes given to the result.

Another approach to estimating $\partial\sigma_i^2/\partial\gamma$ is to use non–parametric ideas i.e., since σ_i^2 is an unknown function of elements in $\mathscr{F}_i$, one might take z_i as known functions of these elements and then approximate σ_i^2 by the series expansion $z_i'\gamma$. Examples of z_i would be orthogonal polynomials or Fourier terms such as sines and cosines. Kamstra (1990) explores this idea through the theory of neural networks, which is a procedure for doing non–parametric regression by series methods. As z_i he selects a set of q principal components of $\zeta_{ij} = (1+\exp(-x_i'\delta_j))$, where values of δ_j (j=1,...,r) are found by randomly drawing from [–R,R] and then used to construct ζ_{ij}. The parameters r and q are chosen as 4p and p if $n \leq 50$; for $n > 50$ r is increased according to $2\log(n)n^{-1/6}$ and this rule is also applied to q after $n > 100$. R was always equal to unity in Monte Carlo experiments performed with the test. In his Monte Carlo work he finds that this "nets" test works well in a wide variety of circumstances.

Although the score test is a useful benchmark for suggesting suitable choices for z_i, it is known that it is only "locally" optimal in large samples, and if departures from the null are strong, one might do better using alternative information more directly. Considering the score in (18), $\sigma_i^{-2}\ \partial\sigma_i^2/\partial\gamma = \partial\log\sigma_i^2/\partial\gamma = -2\partial\log\sigma_i^{-1}/\partial\gamma$. For small departures from the null i.e. $\gamma=\bar{\gamma}$, a linear approximation to this quantity is likely to suffice and that can be regarded as being proportional to $\sigma_i^{-1}(\bar{\gamma})-\sigma_i^{-1}(\gamma^*) = \sigma_i^{-1}(\bar{\gamma})-\sigma^{-1}$. Hence, one interpretation of the LM test is that it takes as z_i, $\sigma_i^{-1}(\bar{\gamma})$ for a value of γ, $\bar{\gamma}$ close to γ^*. For larger departures this argument indicates that it would make sense to use $\sigma_i^{-1}(\bar{\gamma})$ as z_i, where $\bar{\gamma}$ is now some specified value of γ thought to reflect the alternative hypothesis. Thus a family of tests, $\tilde{\tau}(\bar{\gamma})$, indexed upon $\bar{\gamma}$, could be formed. Evans and King (1985) make this proposal.[14] They find that choices of $\bar{\gamma}$ can be made that are superior to the LM test in small samples, even if $\bar{\gamma}$ lies away from the true value γ. In the event that $\gamma = \bar{\gamma}$ their test is a point optimal test, in the sense of the Neyman–Pearson lemma applied to testing the simple hypotheses of $\gamma = \gamma^*$ versus $\gamma = \bar{\gamma}$.

Even after a candidate for z_i has been determined, the optimal score in (19) depends upon the function ψ_i, and therefore requires some knowledge of or approximation to the density f(.). Because f(.) will be rarely known exactly, it is of interest to explore the possibility of allowing for general forms of ψ_i. Within the class of generalized exponential densities giving rise to generalized linear models (GLM), $\psi_i\epsilon_i$ is known as the "deviance" function, see McCullagh and Nelder (1983), and one could work within that framework in devising tests for heteroskedasticity, see Gurmu and Trivedi (1990). Alternatively, there is a large literature on estimating β efficiently in the face of unknown density for u_i, and it can be applied to the current situation, see Bickel (1978). Two interesting ways of approximating ψ_i are proposals by Potscher and Prucha (1986) and McDonald and Newey (1988) that the Student's t and the generalized t density be used for f(.), as that allows a diversity of shapes in ψ_i. For the generalized t density $\psi_i=(rs+1)\mathrm{sgn}(u)|u|^{r-1}/(q\sigma^r+|u|^r)$, with r,s being distributional parameters. McDonald and Newey propose that either r,s be estimated by maximizing $\Sigma \log f(r,s)$, where β is replaced by the OLS estimate , or by minimizing

[14] z_i has to be normalized such that γ could be interpreted as a coefficient of variation. In their test σ^2 is estimated as the OLS residual variance while the β appearing in $\phi_i = [\sigma^{-2}(y_i - x_i'\beta)^2 - 1]$ is estimated from the GLS moment condition $E[(1+z_i'\bar{\gamma})^{-1/2}(y_i - x_i'\beta)] = 0$. Because the distribution of $\hat{\tau}$ does not depend on $\hat{\beta}$ or $\hat{\sigma}^2$, this switch is asymptotically of no consequence.

the var($\tilde{\beta}$), where $\tilde{\beta}$ is the MLE, because the latter depends upon (r,s) solely through a scalar.

Within the literature there are examples of ϕ_i functions that can be regarded as performing the same task as ψ_i does, namely adapting the functional form of ϕ_i to the nature of the density. One that appears in many econometric packages is Glejser's (1969) test that sets $\phi_i=|u_i|-E(|u_i|)$. Glejser proposed regressing $|u_i|$ against a constant and z_i, and the intercept in such a regression essentially estimates $E|u_i|$ under the null hypothesis.[15] Since the optimal ϕ_i is $-\psi_i\epsilon_i-1$, Glejser's test will be optimal if $\psi_i=|u_i|/u_i= |\epsilon_i|/\epsilon_i=\mathrm{sgn}(\epsilon_i)$. The density with such an ψ_i is the double exponential $f(\epsilon)=Ce^{-|\epsilon|}$, showing that Glesjer's test is likely to be successful in the situation of fat tailed densities. This would constitute an argument for its use when ARCH is being tested for, as it has been observed that f(.) has fat tails even after an ARCH process has been allowed for, Engle and Bollerslev (1986) and Nelson (1991).

Rather than approximating ψ_i it is tempting to estimate it non–parametrically. For example, one could estimate $f(\cdot)$ and its derivative by a kernel estimator at the points $\hat{\epsilon}_i$, where $\hat{\epsilon}_i$ are the standardized OLS residuals, and then proceed to form the test using this estimated quantity, $\hat{\psi}_i$, in place of ψ_i. Whang and Andrews (1990) provide theorems regarding the distribution of conditional moment tests when a component of the test is estimated non–parametrically. A critical condition in their theorems needs to be verified in order to ensure that the distribution of $\hat{\tau}$ does not

---

[15]In order that the distribution of Glejser's test be independent of $\hat{\beta}$ it will be necessary that $E(z_i x_i' \mathrm{sgn}(u_i))=0$, and this requires conditional symmetry for the distribution of u_i. Thus in Kamstra's (1990) simulations one would expect that referring Glejser's test to a chi square distribution would be in error if the underlying density was an exponential or a gamma, and this is apparent in his results. Conditional symmetry is also required for the optimal score test to be asymptotically independent of $\hat{\beta}$ as the derivative of (19) with respect to β will only be zero if $E(x_i\psi_i)=0$. For independence from $\hat{\sigma}^2$, $E[(\partial\sigma_i^2/\partial\gamma)|_{\gamma=\gamma^*}]=0$ is needed, the analogue of $E(z_i)=0$ used earlier.

depend upon the non–parametric estimator of ψ_i asymptotically. Here that condition requires $n^{-1/2}\{E[z_i(-\zeta_i\epsilon_i-1)]\}|_{\zeta_i=\hat{\psi}_i}$ to be $o_p(1)$, where ζ_i is a function of data preserving whatever features $f(\cdot)$ is known to possess, such as symmetry, while the expectation is taken before the substitution of $\hat{\psi}_i$ for ζ_i. Writing the expectation in the condition to be tested as $n^{-1/2}E\{z_i[-(\zeta_i-\psi_i)\epsilon_i-\psi_i\epsilon_i-1]\}$, independence of z_i and ϵ_i along with $E(z_i)=0$ would ensure that $E(z_i\psi_i\epsilon_i)=0$, reducing the requirement to $n^{-1/2}z_i(\hat{\psi}_i-\psi_i)\epsilon_i$ being $o_p(1)$, which necessitates $n^{1/4}$ consistency for whatever non–parametric estimator of ψ_i is used for $\hat{\psi}_i$.

3.3 Test Statistics Based on Estimator Comparison

As mentioned in section 2, one possible test for a specification error is to compare estimators whose probability limit differs only if there is a mis–specification. One way to effect such a comparison is to substitute the parameter estimates from a set of first–order conditions defining them into those for another estimator. When testing for heteroskedasticity there have been two proposals based on this line of thought.

Koenker and Bassett (1982) estimate β in (15) by a quantile estimator i.e. $\hat{\beta}(\eta)$ was chosen to minimize $\Sigma\rho_\eta(y_i-x_i'\beta)$, where $\rho_\eta(\lambda)=|\eta-1(\lambda<0)||\lambda|$, $1(\cdot)$ is the indicator function, and $0<\eta<1$ defines the η'th quantile. They show that the quantile estimators of the **slope** coefficients, $\hat{\beta}_1$, are consistently estimated when there is no heteroskedasticity, but that plim $\hat{\beta}_1(\eta)$ differs according to η if there is heteroskedasticity. This feature leads to a test for heteroskedasticity based on a comparison of estimators of β_1 at two different quantiles η_1 and η_2, i.e. the test is based on $\hat{\beta}_1(\eta_1)-\hat{\beta}_1(\eta_2)$. Interpreted as a conditional moment test of the form $E(z_i\phi_i)=0$, this would be $E[x_{1i}(\eta_1-1(y_i<x_i'\beta))u_i]=0$, where x_{1i} are the mean corrected regressors corresponding to the slope coefficients β_1. Their computation of the asymptotic local power function of this test when $\sigma_i^2=1+x_{1i}\delta n^{-1/2}$ revealed that power

was larger for this "comparison" test than for the LM test appropriate when $f(\cdot)$ is normal i.e one based on $\phi_i=\sigma^{-2}u_i^2-1$, whenever the true density $f(\cdot)$ was a contaminated normal. As Newey and Powell (1987) point out, these power computations made by Koenker and Bassett exaggerated the power gain due to an error in deriving the non–centrality parameters of the test statistics, but, even after correction, there was still an improvement.

A major disadvantage of working with quantile estimators is that the function $\rho_\eta(\cdot)$ is not differentiable. This feature led Newey and Powell to replace $\rho_\eta(.)$ of the quantile estimators with $\rho_\nu(\lambda)= |\nu-1(\lambda<0)|\lambda^2$; the estimator of β found by minimizing this function is $\hat{\beta}(\nu)$, and was termed Asymmetric Least Squares (ALS). Their recommended test is then based on $\hat{\beta}_1(\nu_1)-\hat{\beta}_1(\nu_2)$. After examining a numerical experiment they find that $\nu_1=.46$, $\nu_2=.54$ seems to give best power. With these values of ν their test performs in a very similar fashion to the Koenker–Bassett test when the density $f(\cdot)$ is contaminated normal, but has much better power if $f(\cdot)$ is normal, leading to their conclusion that the comparison be based on the ALS estimator. Now, the implicit moment condition used in the comparison is $E[x_{1i}(\nu_1-1(y_i<x_i'\beta))u_i]=0$, and, when $\nu_1=.5$, this becomes $E[x_{1i}(.5-1(u_i<0))u_i]= E[.5x_{1i}sgn(u_i)u_i]=E(.5x_{1i}|u_i|)=0$, which is just the moment condition used in constructing Glejser's test, provided z_i is set to x_{1i}. Moreover, when the error density is symmetric, $\hat{\beta}_1(.5)$ will be OLS. This argument points to the fact that the performance of the ALS comparison test $\hat{\beta}_1(.46)-\hat{\beta}_2(.54)$ should be very close to Glejser's test based on $\Sigma x_{1i}|\hat{u}_i|$. Indeed this is what Newey and Powell find, culminating in their conclusion that using Glesjer's test would be a simple way of attaining the benefits of doing the ALS test.

3.4 Test Statistics Based on CUSUMS of Moments

Under the null hypothesis the scale parameter σ^2 is estimated from some moment condition. If the errors ϵ_i are normally distributed, the moment defining an estimator of σ^2 would be $E(\Sigma\sigma^{-2}u_i^2-1)=0$. For other densities, using the score for σ from (16), with σ_i^2 replaced by σ^2, might produce a more satisfactory estimate. Focusing upon the normal case, the general treatment of testing for structural change in the variance provided in section 2 involved looking at the CUSUMS $\sum_{i=1}^{nk}(\hat{\sigma}^{-2}\hat{u}_i^2-1)$, or just $b(k)=\sum_{i=1}^{nk}\hat{\sigma}^{-2}\hat{u}_i^2$. Harrison and McCabe (1979) proposed b as a test for heteroskedasticity with fixed k, while Breusch and Pagan (1979) adopted the C(k) test in (14). Because the C(k) test is just a transformation of $\Sigma\hat{\sigma}^{-2}\hat{u}_i^2$, there will be no difference in conclusions based upon it or b(k) provided they are referred to their appropriate critical values. An advantage of C(k) is that it is centered and scaled so that asymptotically it is a χ^2 random variable. McCabe (1986) mentions the possibility of using $\max_k b(k)$. However, he did not find the distribution of this test. Instead he ordered $\hat{\sigma}^{-2}\hat{u}_i^2$ and computed a test based on the order statistics for this sequence. As Andrews (1990) has now tabulated the distribution of $\max_k C(k)$ it seems more satisfactory to perform a test in this way. The other two test statistics given earlier—L_W and L_C—do not seem to have been formally used in the literature.

4. Testing Heteroskedasticity in Models Featuring Heteroskedasticity

Section 3 was devoted to procedures for detecting heteroskedasticity when the maintained hypothesis was that there was none. However, the last two decades have seen a proliferation of models incorporating heteroskedasticity as one of their characteristics. Such heteroskedasticity is **intrinsic** to the model and what needs to be tested is not the presence of heteroskedasticity **per se** but whether it departs from that featured in the maintained model, that is it is **extrinsic** heteroskedasticity which is important. Indeed, one might argue that it is rare to have a situation in which there is no intrinsic heteroskedasticity in linear models such as (15). If y_i was a member of the exponential family the density of u_i is rarely homoskedastic, with the normal density being the dominant exception. Moreover, the heteroskedasticity will generally have the characteristic of being a function of the conditional mean $\mu_i=E(y_i|x_i)$. Many examples of models with intrinsic heteroskedasticity might be given, but four representative "types" are set out in this section. Each may be regarded as a regression model with heteroskedasticity, and it is desired to test if the predicted type of heteroskedasticity is sufficient to account for non–constancy in the variance of the errors. These models arise in situations where there is "count," binary or censored data or in which there is interest in explaining volatility in a series. Our enumeration is scarcely exhaustive. Many extensions can be made to the basic models, for example, the type of censoring giving rise to selectivity bias or the possibility of multiple rather than binary responses, but the collection should illustrate the common themes regarding testing that will be found in all such models.

4.1 A General Approach to Testing for Extrinsic Heteroskedasticity

All the models considered in this section can be regarded as being characterized by an error term u_i that has variance $\tilde{\sigma}_i^2$ when the maintained model is correct. The variance $\tilde{\sigma}_i^2$ is a function of some parameters δ. Defining $e_i=\tilde{\sigma}_i^{-1}u_i$, it is e_i which is to be tested for heteroskedasticity i.e., in terms of the analysis of section 3. $\bar{\sigma}_i^2$ will now be the variance of u_i, becoming equal to $\tilde{\sigma}_i^2$ when there is none. Hence, under the null hypothesis, the variance of e_i , σ_i^2, is unity. With this change, the moment conditions used for tests of heteroskedasticity in section 3 will simply be modified by replacing u_i by e_i and by setting σ^2 to unity. Thus the basic moment condition used below (15) becomes $E(z_i(e_i^2-1))=0$.

There are however some complications. A minor one arises because the parameters to be estimated will now include not only those like β in (15) but also δ, the parameters entering into the intrinsic form of heteroskedasticity. Except for a few instances, the distribution of tests for heteroskedasticity in the basic regression model did not depend upon any nuisance parameters such as β and σ^2, but this is unlikely for δ. When $m_i=z_i(e_i^2-1)$, and there is no overlap between δ and β, $\partial m_i/\partial\delta=(\partial z_i/\partial\delta)(e_i^2-1)-z_i u_i^2\sigma_i^{-4}(\partial\sigma_i^2/\partial\delta)$, which has expectation of $-E[z_i\sigma_i^{-2}(\partial\sigma_i^2/\partial\delta)]$, a quantity unlikely to be zero. Accordingly, the var$(\hat{\tau})$ must be computed from (4), or τ and δ must be jointly estimated.

As pointed out in the introduction to this section, there are instances in which the variance of u_i, and hence σ_i^2, will always depend solely upon whatever parameters β enter the conditional mean (along with x_i) i.e. δ coincides with β. In these situations Cameron (1991) points to the possibility of modifying m_i, so as to asymptotically eliminate any distributional dependence. Defining $m_i'=m_i+[E(\partial m_i/\partial\mu_i)\,|\,\mu_i](y_i-\mu_i)$, it is clear that $E(\partial m_i'/\partial\delta)=E(\partial m_i/\partial\delta)+$ $E\{(\partial E[(\partial m_i/\partial\mu_i)\,|\,\mu_i]/\partial\delta))(y_i-\mu_i)\}-E\{E[(\partial m_i/\partial\mu_i)\,|\,\mu_i](\partial\mu_i/\partial\delta)]$.[16] In this expression the middle term is zero and the last is just $-E(\partial m_i/\partial\delta)$, making $E(\partial m_i'/\partial\delta)=0$. Thereupon, adopting $E(m_i')=0$ as the requisite moment condition would allow any distributional dependence upon δ to be eliminated. Cameron puts $m_i=z_i\{(y_i-\mu_i)^2-\tilde{\sigma}_i^2\}$ so that m_i' would be $z_i\{(y_i-\mu_i)^2-\tilde{\sigma}_i^2-(\partial\tilde{\sigma}_i^2/\partial\mu_i)(y_i-\mu_i)\}$. For the moment condition $m_i=z_i(e_i^2-1)$, and $e_i=\tilde{\sigma}_i^{-1}(y_i-x_i'\beta)$, $\mu_i=x_i'\beta$, $E[(\partial m_i/\partial\mu_i)\,|\,\mu_i]=-z_iE[(\partial\tilde{\sigma}_i^2/\partial\mu_i)\tilde{\sigma}_i^{-2}\,|\,\mu_i]$ making $m_i'=z_i(e_i^2-E[(\partial\tilde{\sigma}_i^2/\partial\mu_i)\tilde{\sigma}_i^{-1}\,|\,\mu_i]e_i-1)$.

Inspection of the modified moment m_i' in Cameron's case highlights the fact that the new test involves a component $\zeta_i e_i$, where $\zeta_i=-z_iE[(\partial\tilde{\sigma}_i^2/\partial\mu_i)\tilde{\sigma}_i^{-1}\,|\,\mu_i]$, that is testing for specification error in the conditional mean, which makes sense given that the variance $\tilde{\sigma}_i^2$ is a function solely of the conditional mean (due to the fact that there are

---

[16]The term $E[(\partial m_i/\partial\mu_i)\,|\,\mu_i]$ would actually be evaluated under the **null** hypothesis. It is also obvious that it could be replaced with the unconditional expectation $E(\partial m_i/\partial\mu_i)$ without changing the argument. In most instances it is probably easier to find the conditional moment, but not always.

no parameters in δ not appearing in μ_i). Although in the regression model with normally distributed errors it is possible to make a distinction between whether it is the conditional mean or the conditional variance which is mis–specified, outside of that context it is frequently very difficult to conceive of an alternative model in which changes in μ_i do not impinge upon σ_i^2. Therefore, tests of extrinsic heteroskedasticity inevitably involve a test for the correct specification of the conditional mean. Many of the models analyzed in this section have such a property. Moreover, because tests for correct specification of the conditional mean generally involve lower order moments i.e involve a test of $E(\zeta_i e_i)=0$, it is unclear that tests involving the square of e_i^2 would ever be preferred. Indeed, as will become apparent, score tests for some of the models of this section do in fact involve examining a moment condition like $E(\zeta_i e_i)=0$ and do not involve the squares of e_i at all.

4.2 Testing for Heteroskedasticity in Discrete Choice Models

Discrete choice models are associated with observations on a binary random variable y_i taking values zero or one, and some causal variables x_i, with the two sets related in such a manner that $\Pr(y_i=0|x_i)=F(x_i,\theta)= F_i$, where $F(t)$ is a distribution function. A standard way to motivate this probability is to interpret it as arising from a latent variable model

$$y_i^*=x_i'\beta+u_i^* \tag{20}$$

where $\mathrm{var}(u_i^*)=1$ and $y_i=1(y_i^*\geq 0)$. Then $F(\cdot)$ will be the cumulative distribution function of the errors u_i^* and, using the binary structure of y_i,

$$y_i = (1 - F_i) + u_i, \tag{21}$$

where $F_i = \mathrm{prob}(u_i^* \leq -x_i'\beta) = F(-x_i'\beta)$ and u_i is heteroskedastic with conditional variance $F_i(1-F_i)$. Hence, the conditional variance is always related to the conditional mean.

One might introduce extrinsic heteroskedasticity into this model by setting $\mathrm{var}(u_i^*) = (1+w_i'\gamma)$, thereby modifying F_i to $F(-x_i'\beta/(1+w_i'\gamma)^{1/2})$, where $F(\cdot)$ is now the distribution function of the standardized errors, $(1+w_i'\gamma)^{-1/2}u_i^*$. Intrinsic heteroskedasticity will be $\tilde{\sigma}_i^2 = F_{oi}(1-F_{oi})$, where $F_{oi} = F(-x_i'\beta)$, and the task is to see if there is any extra heteroskedasticity i.e. to test if $\gamma=0$. Following the discussion in section 4.1 one could form a test statistic based on $\Sigma z_i(e_i^2-1)$, where $e_i = \tilde{\sigma}_i^{-1}(y_i - F_{oi})$. Alternatively, following Davidson and Mackinnon (1984), the score test for γ being zero could be used. As the log likelihood of $(y_1,\ldots,y_n)$, conditional upon $\{x_i, z_i\}_{i=1}^{n}$, will be

$$L = \sum_{i=1}^{n} \{(1-y_i)\log(F_i) + y_i \log(1-F_i), \tag{22}$$

under $H_o: \gamma=0$, the scores become

$$d_\gamma = \sum_{i=1}^{n} [(\partial F_i/\partial\gamma)_{\gamma=0}](1-F_{oi})^{-1}F_{oi}^{-1}\,(y_i - F_{oi}), \tag{23}$$

$$= \sum_{i=1}^{n} z_i e_i, \tag{24}$$

where $z_i = [(\partial F_i/\partial\gamma)_{\gamma=0}](1-F_{oi})^{-1/2}F_{oi}^{-1/2} = .5w_i(x_i'\beta)f(x_i'\beta)(1-F_{oi})^{-1/2}F_{oi}^{-1/2}$. As foreshadowed earlier the optimal test therefore does not involve the squares of e_i, and is effectively testing for specification errors in the conditional mean function.

4.3 Testing for Heteroskedasticity in Censored Data Models

There are many types of censored data but the simplest would be the case of left censoring at zero of the latent variable y_i^* in (20) introduced by Tobin (1958).

Observed data is then $y_i=1(y_i^*>0)y_i^*$. For a non-negative random variable it is known that $E(y_i|x_i)=\int_{-x_i'\beta}^{\infty}(1-F_0(\lambda))d\lambda$, where $F_0(\cdot)$ is the distribution function of u_i^*, while the conditional variance of y_i would be $\bar{\sigma}_i^2 = 2\int_{-x_i'\beta}^{\infty}(1-F_0(\lambda))\lambda d\lambda +2x_i'\beta \int_{-x_i'\beta}^{\infty}(1-F_0(\lambda))d\lambda -\{\int_{-x_i'\beta}^{\infty}[1-F_0(\lambda)]d\lambda\}^2$, and these could be used to define e_i for the purpose of constructing the moment $n^{-1}\Sigma z_i(e_i^2-1)$. A disadvantage of the approach is that $F_0(\lambda)$ must either be estimated or specified. One possibility is to estimate it by non-parametric methods, see Whang and Andrews (1991), but no applications in that vein are reported in the literature, and there would seem little benefit in so doing as a mis-specification of either heteroskedasticity or the density for u_i^* essentially has the same impact, and it is going to be very difficult to distinguish between the two types of specification error. For that reason $F_0(\cdot)$ is likely to be specified, and therefore one might as well construct a score test for heteroskedasticity.

Assuming that $\bar{\sigma}_i^2=\text{var}(u_i^*)$ is a function of some parameters γ such that $\gamma=\gamma^*$ produces a constant variance σ^2, $f(\cdot)$ is the density function of $u_i^*/\bar{\sigma}_i$, and $F(\lambda)=\int_{-\infty}^{\lambda}f(u)du$, the log likelihood of the data is

$$L= \sum_{i=1}^{n}(1-y_i)\log F(-x_i'\beta/\bar{\sigma}_i) +\sum_{i=1}^{n}y_i\{-.5\log\bar{\sigma}_i^2 + \log f[(y_i-x_i'\beta)/\bar{\sigma}_i)]\}, \tag{25}$$

with scores evaluated at $\gamma=\gamma^*$

$$1/2\sum_{i=1}^{n}\sigma^{-1}(\partial\bar{\sigma}_i^2/\partial\gamma)|_{\gamma=\gamma^*}\{(1-y_i)f_0(\xi_i)F_0^{-1}(\xi_i)\xi_i-y_i(\psi_i\epsilon_i-1)\}, \tag{26}$$

where $\xi_i=(x_i'\beta)/\sigma$, $e_i=\sigma^{-1}(y_i-x_i'\beta)$, $\epsilon_i=\sigma_i^{-1}u_i^*$, and $\psi_i=f_{0i}^{-1}(\epsilon_i)(\partial f_{0i}/\partial\epsilon_i)$. As expected, if there is no censoring $(y_i=1)$, the test would be identical to that for the uncensored case (see (19)). In general, it does not reduce to a test involving the squares of e_i.

Jarque and Bera (1982) derived this test and Lee and Maddala (1985) interpret it as a conditional moment test. As with the regression case one might base a test on a comparison of the MLE of β and another estimator that is consistent when there is no heteroskedasticity. Powell's (1986) censored quantile estimator minimizing $\sum_{i=1}^{n} \rho[y_i - \max\{0, x_i'\beta\}]$, where $\rho(\lambda) = [\eta - 1(\lambda<0)]\lambda$, could be used as the analogue of the Koenker–Bassett proposal for the linear regression model discussed in section 3.3.

4.4 Testing for Heteroskedasticity in Count Data Models

The modelling of discrete count data, such as the number of patents granted or the number of visits to a medical facility, has become of greater interest to econometricians as large scale data sets containing such information have emerged. Naturally, the "work horse" in the analysis of such data sets has been the Poisson regression model, but, owing to the fact that it forces a restriction, upon the data, that the conditional mean of y_i, μ_i, is equal to the conditional variance, there have been attempts to supplement it with models based on other densities not having such a restriction. An example would be the negative binomial density used by Hausman, Hall and Griliches (1984) and Collings and Margolin (1985). However, sometimes such alternatives are difficult to estimate, and it is not surprising that a literature developed to test for whether the type of heteroskedasticity seen in the data deviated from that intrinsic to the Poisson model i.e., to test whether $\mathrm{var}(y_i)$ equaled the $E(y_i)$.

Since the maintained model is generally the Poisson model the situation is one of a regression such as (15), albeit the conditional mean μ_i may no longer be linear in the x_i but rather may be a non-linear function such as $\exp(x_i'\beta)$. Given some specification for μ_i, the Poisson regression model has $\tilde{\sigma}_i^2 = \mu_i$ and the obvious moment condition to test is $E(z_i(e_i^2 - 1)) = 0$.

An alternative way to write this condition is as $E(z_i(e_i^2-\mu_i^{-1}E(y_i))=0$ which, under the maintained hypothesis of a Poisson model, becomes $E\{z_i\mu_i^{-1}[(y_i-\mu_i)^2-y_i]\}=E\{\bar{z}_i[(y_i-\mu_i)^2-y_i]\}=0$. Defining a class of regression models for count data that are indexed by a parameter γ, and which reduce to the Poisson model when $\gamma=0$, Cameron and Trivedi (1990) point out that the last mentioned condition coincides with the score test based on γ if the wider class derives from the Katz system of densities set out in Lee (1986), or from the "local to Poisson" case of Cox (1983). Differences in score tests therefore reside solely in the nature of $\bar{z}_i$.[17]

Selecting $\bar{z}_i$ requires an alternative model for the variance of u_i, $\bar{\sigma}_i^2 = \tilde{\sigma}_i^2\sigma_i^2 = \mu_i\sigma_i^2$, and may be found as follows. From (18) z_i should be set to $(\partial\sigma_i^2/\partial\gamma)|_{\gamma=0}$ and, if $\sigma_i^2=1+\gamma g_1(\mu_i)$, this would make $z_i=g_1(\mu_i)$ under the Poisson specification. Converting to $\bar{z}_i$ gives $\bar{z}_i=\mu_i^{-1}g_1(\mu_i)=\mu_i^{-2}g(\mu_i)$, adopting Cameron and Trivedi's notation. For their tests they employ $g(\mu_i)$ as either μ_i or μ_i^2 i.e., $g_1(\mu_i)$ is either unity or μ_i, leading to tests based on the moment conditions $E(e_i^2-1)=0$ and $E(\mu_i(e_i^2-1))=0$. In simulation studies reported in their paper the variances of the test statistics are formed in a number of ways, either by explicitly evaluating $E(m_im_i')$ under the maintained hypothesis of a Poisson model or by the adoption of a robust version, and it was the latter which had better performance.

4.5 Specification Tests for Additional ARCH Effects

In recent years the ARCH model and its variants have become a very popular way of modelling heteroskedasticity in econometric models, particularly those concerned with

---

[17]Assuming that μ_i is a function of parameters β, when using the moment $m_i=\bar{z}_i[(y_i-\mu_i)^2-y_i]$ it will be the case that $E(\partial m_i/\partial\beta))=0$ and, consequently, there will be no nuisance parameter dependencies. In contrast, as Cameron and Trivedi observe, if m_i was replaced by $\bar{m}_i=\bar{z}_i[(y_i-\mu_i)^2-\mu_i]$, now there would be nuisance parameter dependencies, even though $E(\bar{m}_i)=0$ under the null hypothesis of a Poisson model.

financial time series—see Bollerslev et al (1990) for a survey. In terms of the structure of section 4.1, u_i is defined by (15) (perhaps with a non–linear rather than linear function of x_i) with variance $\bar{\sigma}_i^2$, becoming $\tilde{\sigma}_i^2$ with a particular maintained type of conditional heteroskedasticity. A number of alternative specifications for $\bar{\sigma}_i^2$ have emerged. A general expression would be $g(\ \alpha_0+\sum_{j=1}^{r}\alpha_j h_{ij}(u_{i-j}))$, which is indexed by three characteristics, $g(\cdot)$, $h_{ij}(\cdot)$ and r. Table 1 provides a list of some of the most popular cases according to the values assigned to these functions and the parameter r. Others would be possible e.g. one could make $\bar{\sigma}_i^2$ functions of a series expansion in u_{t-i} or e_{t-i} e.g. using the Flexible Fourier Form as in Pagan and Schwert (1990) or the neural network approximation in Kamstra (1990).

Table 1

Autoregressive Conditional Heteroskedasticity Formats

Name	r	$g(\psi)$	$h_{ij}(u_{i-j})$	Source
ARCH(p)	finite	identity	u_{i-j}^2	Engle(1982)
GARCH	∞	identity	u_{i-j}^2	Bollerslev(1986)
NARCH	finite	$\psi^{1/\gamma}$	$u_{i-j}^{2\gamma}$	Higgins and Bera(1989)
MARCH	∞	identity	$\sin(u_{i-j}^2)$if $au_{i-j}^2<(\pi/2)$ unity if $au_{i-j}^2\geq(\pi/2)$	Friedman and Laibson(1989)
EGARCH	∞	$\exp(\psi)$	$\delta e_i+\{\lvert e_i\rvert-(2/\pi)^{1/2}\}$	Nelson(1991)

The basic model is $\bar{\sigma}_i^2=g(\psi)$ where $\psi=\alpha_0+\sum_{j=1}^{r}\alpha_j h_{ij}(\cdot)$, where $(\cdot)$ can be either u_{i-j} or e_{i-j}. When $r=\infty$ there are restrictions between the α_j.

One can distinguish two different situations in connection with the above formats. The first is when $\tilde{\sigma}_i^2$ is nested within the alternative, $\bar{\sigma}_i^2$, so that $\bar{\sigma}_i^2=\tilde{\sigma}_i^2$ when a set of parameters γ take the value γ^* e.g. if the maintained model is ARCH and the alternative is GARCH or NARCH. Then, as detailed in section 3.2, the optimal choice of z_i depends upon $(\partial\sigma_i^2/\partial\gamma)_{\gamma=\gamma^*} = \tilde{\sigma}_i^{-2}(\partial\bar{\sigma}_i^2/\partial\gamma)|_{\gamma=\gamma^*}$, because $\sigma_i^2=\bar{\sigma}_i^2/\tilde{\sigma}_i^2$ and γ does not appear in the denominator. Consequently, it is simply a matter of finding the derivatives of $\bar{\sigma}_i^2$ with respect to γ evaluated with $\gamma = \gamma^*$ i.e., $\bar{\sigma}_i^2 = \tilde{\sigma}_i^2$. For the NARCH model as alternative this becomes $z_i= \sum_{j=1}^{r} \hat{\alpha}_j\hat{e}_{t-j}^2\log(\hat{e}_{t-j}^2)-\tilde{\sigma}_i\log(\tilde{\sigma}_i)$ (Higgins and Bera (1989)). If, however, the two conditional variances are non–nested, as occurs with an EGARCH/GARCH comparison, this strategy will not work. Ideally one wants z_i to reflect the "extra" information in $\bar{\sigma}_i^2$ not contained in $\tilde{\sigma}_i^2$. Because the term $\tilde{\sigma}_i^{-2}(\partial\bar{\sigma}_i^2/\partial\gamma)_{\gamma=\gamma^*}$ in the nested case essentially represents the difference between $\tilde{\sigma}_i^2$ and $\bar{\sigma}_i^2$, z_i is ideally very like $\bar{\sigma}_i^2$. However to measure this we would need to estimate under both the null and alternative, which is not in the spirit of a diagnostic test. A simple solution is to think of $\bar{\sigma}_i^2$ as a function of $\tilde{\sigma}_i^2$ and to choose z_i in that way. If one has a precise alternative in mind, one could form $\hat{\bar{\sigma}}_i^2$ from $\hat{\tilde{\sigma}}_i^2$ using whatever transformation defines $\bar{\sigma}_i^2$ e.g. $\exp(\hat{\tilde{\sigma}}_i^2)$ might be used for z_i when EGARCH is thought of as the alternative. Pagan and Sabau (1987a) used $\tilde{\sigma}_i^2$ itself, although, as Sabau (1988) found, the power of such a test is unlikely to be very strong in many instances. The reason is set out in Pagan and Sabau (1987b); the power of their test derives from the extent of the inconsistency in the MLE estimator of β induced by mis–specification of the conditional variance. In the basic regression model there will be no inconsistency, but, because β also enters into the determination of $\tilde{\sigma}_i^2$, the possibility of inconsistency arises in ARCH models. Pagan and Sabau determine that inconsistency only eventuates if the true conditional variance is an asymmetric function of u_i, which it would be if the alternative model to an ARCH was (say) EGARCH, but would not be if the alternative was GARCH. Apart from this research, the question of good choices for z_i does not seem to have been explored in the literature, and it is worthy of some further study.

Tests for different types of volatility specifications also encounter the same set of difficulties as arose when testing for whether there is any heteroskedasticity in the basic regression model viz. possible dependence upon estimated nuisance parameters and the fact that the optimal form of the test will depend upon the density of e_i. With regard to the first, $\tilde{\sigma}_i^2$ depends upon both β and other parameters δ (as seen in Table 1). Taking $m_i=z_i(e_i^2-1)$ it is necessary that $E(z_i(\partial e_i^2/\partial\theta))=0$ if there is to be no dependence upon estimates of θ. It is easily seen that this is unlikely to be true for $\theta=\delta$, and is even problematic for β. To appreciate the complications with the latter, differentiate $\tilde{\sigma}_i^{-2}(y_i-x_i'\beta)^2$ with respect to β, giving $-\tilde{\sigma}_i^{-2}x_iu_i-\tilde{\sigma}_i^{-4}u_i^2(\partial\tilde{\sigma}_i^2/\partial\beta)$. This will only have zero expectation if $\partial\tilde{\sigma}_i^2/\partial\beta$ is an odd function of u_i and u_i is in turn symmetrically distributed around zero (conditional upon $\mathcal{F}_i$). For the GARCH model, the errors u_i are generally taken to be conditionally normal, while $\partial\tilde{\sigma}_i^2/\partial\beta$ is an odd function, thereby satisfying both conditions, but that would not be true of the EGARCH model. It would also not be true of the GARCH model if the density of u_i was not symmetric. However, since the parameters are most likely to have been estimated by MLE, one could always regress $\hat{m}_i$ upon unity and the estimated scores to allow for the effects of prior estimation. Turning to the second question of the selection of ϕ_i in $m_i=z_i\phi_i$, there has been little research into using the alternative moment condition, mentioned in section 3.2, in which ϕ_i is set to $(-\psi_i e_i - 1)$ rather than $(e_i^2 - 1)$, although Engle and Gonzalez–Rivera (1991) attempted estimation in this way.

Many applications of the ARCH technology are not to pure ARCH models but to ARCH–M (ARCH in mean) contexts, in which a function of $\tilde{\sigma}_i^2$ appears among the regressors in (15). Thus a specification error in the conditional variance now affects the mean and the situation is reminiscent of the models discussed in sections 4.2 and 4.3. Therefore, although a test might be based upon $\Sigma z_i(\hat{e}_i^2 - 1)$, it is likely to be better to directly test the mean using $n^{-1}\Sigma z_i \hat{e}_i$. Oddly enough, in Pagan and Sabau's(1987a) application of these tests to the ARCH–M model estimated in Engle et al(1987), the test based on the squares of $\hat{e}_i^2$ gave much stronger rejection than that based on $\hat{e}_i$.

5. The Size and Power of Heteroskedasticity Tests

5.1 The Size of Tests

The moment tests outlined in section 2 are based on asymptotic theory, raising the possibility that the asymptotic results may fail to be a reliable indicator of test performance in small samples. Many simulation studies have shown that this is true for score tests of heteroskedasticity in the regression model, especially for those based upon normality in u_i, with a frequent finding being that the actual size of the tests is less than the nominal one available from asymptotic theory. When robust tests are utilized it is equally common to find that the nominal is less than the actual size e.g. Skeels and Vella (1991) find that the true size for tests of heteroskedasticity in the censored regression model obtained by robustly estimating $E(\partial m_i/\partial\gamma)$ and $\mathrm{var}(m_i)$ can be twice the nominal size. Only rarely can the exact distribution of any of these tests be analytically determined, an exception being the Goldfeld–Quandt test, creating an interest in approximating the finite sample distributions or modifying the test statistics to more closely approximate the asymptotic distribution. There are four broad approaches in the literature and these are summarized below.

(i) Approximation by Expansion.

When the m_i are the scores the test statistic will be a score test and a general formula for the finite sample distribution of score tests (S) was provided by Harris (1985):

$$P[S\le c]=P_q+(24n)^{-1}\{\alpha_3P_{q+6}+(\alpha_2-3\alpha_3)P_{q+4}+(3\alpha_3-2\alpha_2+\alpha_1)P_{q+2} \\ +(\alpha_2-\alpha_1-\alpha_3)P_q\} + o(n^{-1}) \tag{27}$$

where $P_q=\text{Prob}(\chi^2_q\le c)$, q is the dimension of the extra parameters being tested by the score test, and $\alpha_1,\alpha_2,\alpha_3$ are functions of cumulants of derivatives of the log likelihood. Computing α_1,α_2 and α_3 can be very complex and mostly needs to be done with a symbolic differentiation package. Honda (1988) specialized this general formula to the case where one was testing for heteroskedasticity in the linear regression model. A serious reservation with this approach is that the **exact** score test needs to be used; in particular, the information matrix needs to be employed as the var(m_i), thereby precluding the use of robust estimates. In practice the exact score test is rarely used e.g the modification in Koenker (1981) of the score test for heteroskedasticity in the linear regression model is what appears in most regression packages, and the distribution of that statistic would be different to what is presented in (27). Of course, if the robust statistic has the same asymptotic distribution as the score test we might hope that (27) would be a reliable guide to its distribution in finite samples. Recently, Smith (1990) has shown that the form of (27) remains valid even when Var(Σm_i) is estimated by $\Sigma m_i m_i'$, except that the α's need to be re-defined. To date his results have not been applied to heteroskedasticity tests.

(ii) Approximation by Distribution

An alternative to an asymptotic expansion is to regard the small sample distribution as being well approximated by some density such as the beta, and to estimate the parameters of the latter by matching up the finite sample moments with those of the approximating distribution. The approximate Prob$[\hat{\tau} \leq c]$ may then be found from the beta distribution. Harrison (1980) did this and found it worked quite well for tests of heteroskedasticity in the linear model.

(iii) Numerical Determination of p–Values

Rather than approximate the complete distribution of the statistic $\hat{\tau}$ it is only the p–value for a given estimate (c) which is sought. In the linear regression model with normally distributed errors and a single variable z_i, the statistic $\hat{\tau}=\Sigma z_i(\hat{\sigma}^{-2}\hat{u}_i^2-1)$ can be written as the ratio of quadratic forms in $\sigma^{-1}u_i$, allowing Imhof's (1961) method of computing p–values for such test statistics to be applied. With more than a single z_i this is no longer true, but the fact that the quadratic form is in $\sigma^{-1}u_i$, an $\mathcal{N}(0,1)$ random variable, means that realizations of $\sigma^{-1}u_i$ can be drawn from an $\mathcal{N}(0,1)$ random number generator. Counting the fraction of times for which $\hat{\tau}$, computed with these numbers, is greater than c constitutes an estimator of the p value. Breusch and Pagan(1979) advocated this approach and regarded it as a simple and cheap way of finding p–values. The idea was originally mooted by Barnard (1963). In the context of testing for heteroskedasticity it has been adopted by Bewley and Theil (1987) when looking for this problem in systems of demand equations. Working with PC's the computation of p–values by simulation is very easy and quite cheap, although it does require re–estimation of the basic model, and, in the context of censored regression or discrete choice models, the cost may still be too high (although it should be remembered that one is only dealing with cases where the sample size is small and the iterative procedures to compute $\hat{\theta}$ can always be started with values from the previous replication).

The procedure also applies to tests modified to gain robustness or those designed to emulate the optimal score tests, since the latter are still functions of $\sigma^{-1}u_i$. However, if the density of u_i is unknown, one cannot simulate from it, and bootstrapping is now the obvious alternative numerical procedure, with the u_i(or m_i)

being drawn from the empirical rather than the assumed density function. Technically, the conditions for the success of the bootstrap are not exactly satisfied here as the statistics are not "pivotal," being dependent upon the estimated parameters θ. However, as this dependence disappears asymptotically one would expect that the method would work well.

(iv) Modifying the Test Statistic

Instead of finding an approximation to the small sample distribution of $\hat{\tau}$ it is sometimes more useful to modify the test statistic to make it correspond more closely to the asymptotic distribution. A simple adjustment in this vein is to form $\hat{\tau}' = \hat{\tau} - \bar{\tau}$, where $\bar{\tau}$ is the $E(\hat{\tau})$ in finite samples, and to refer $\hat{\tau}'$ to the asymptotic distribution of $\hat{\tau}$. Essentially this is an attempt to correct for the fact that $E(\hat{\tau})=0$ only in large samples, and therefore $\hat{\tau}$ will not be centered on zero in finite samples. Conniffe (1990) reports some success with this adjustment for score tests generally while Ara and King (1991) find that it works well for tests of heteroskedasticity in the linear regression model based on $\Sigma z_i(\hat{\sigma}^{-2}\hat{u}_i^2-1)$. The major difficulty in applying the idea is to determine $E(\hat{\tau})$, particularly if z_i is stochastic or the data is censored, although one might employ simulation methods to do this.

5.2 The Power of Tests

A number of studies have been developed to investigate the power of tests mentioned in the preceding sections at detecting heteroskedasticity. Most effort has been concentrated upon the linear regression model, although Skeels and Vella (1991), working with n=600, found that score tests were good in the censored regression model but very poor in the Probit model. Ali and Giacotto (1984), Griffiths and Surekha (1986), Kamstra (1990) and Evans and King (1985), (1988) are perhaps the most comprehensive studies available of the linear regression case. Sometimes it is difficult to draw lessons from these reports as only overall results are provided, involving averaging across many experiments, and, unfortunately, some of these experiments fail to satisfy the assumptions needed to apply either asymptotic or finite sample theory when determining the distributions under the null hypothesis. Worst in this respect is Ali and Giacotto who have an extremely large number of different experiments, some of which feature moments of u_i that do not exist, while others are done with non–symmetric densities for u_i which would invalidate the reference of tests such as Glejser's to Chi Square distributions. It is almost impossible therefore to draw any conclusions from their work, as one does not know which experiments are responsible for the poor performance of any test. In other instances, for example Kamstra's demonstration that the power of robust tests using $\Sigma m_i m_i'$ as an estimate of $var(\Sigma m_i)$ is very weak when testing for ARCH, no explanation for the phenomenon has emerged similar to those provided by Chesher and Spady (1991) and Kennan and Neumann (1988) for kurtosis tests. Until one fully understands the causes of these results it is hard to know if one should recommend against the use of the associated tests.

Some experiments come up with clear cut results, but the experimental design seems to be too restricted or the conclusions are not sufficiently qualified. For example, Griffiths and Surekha conclude that the use of $z_i=i$ is to be recommended over the Goldfeld Quandt and Breusch/Pagan/Godfrey (B/P/G) test if it is possible to order the observations by increasing variance, and that one should use BAMSET if it is not possible to do so.[18] Because the ability to order data by increasing variance means that the heteroskedasticity must be a monotonic function of i, it follows that setting $z_i=i$ rather than to dummy variables, as in Goldfeld–Quandt and BAMSET, should be advantageous. What is surprising is their conclusion concerning BAMSET, since the

[18]It may be worth emphasising that there is no such thing as a B/P/G test without specifying what z_i is. What one specifies the heteroskedasticity to be when devising the test, and what it actually is, are two different matters. For example, in Breusch and Pagan (1979) a test was given with z_i as a dummy variable, so one might even refer to it as the Goldfeld–Quandt test, showing how meaningless the appellation is.

B/P/G test is invariant to ordering. In fact, the power of B/P/G can be directly compared to BAMSET in their Table 1, and it is clearly much larger. Hence, their conclusion is contradicted by their own results. It may be that the objective was to conclude that, within the class of tests that used no information about the variables forcing the conditional variance, BAMSET was best, and indeed the body of the text seems to read this way, but that is not the conclusion stated in the paper. Even adopting such an interpretation would be odd, since the ability to order the data by increasing variance means that one knows the variance is driven by a time trend.

A similar set of comments can be made about the studies by Evans and King (1985), (1988), and their conclusion that "... the emphasis on the B and P test in the recent econometric literature is probably mis-specified." Their preference is for the point optimal test described in section 3.2 over either the B/P/G or Goldfeld–Quandt tests. Again the data is generated so that the true heteroskedasticity is always a monotonic transformation of a trend term and the z_i used is also a monotonic transform of that variable. This feature produces a bias against the Goldfeld–Quandt test, but there are two further factors in the experiment that help the performance of the point optimal test. First, the true heteroskedasticity is generated with a single unknown parameter that is positive, and their test statistic takes account of the positivity of that parameter, whereas the B/P/G test does not. Second, as seen in section 3.2, the point optimal test uses information about the alternative, rather than just local information as in the score test, so the fact that it has superior power when there is a high degree of heteroskedasticity is quite consistent with the nature of each of the tests. Even with these advantages the power differential is only moderate, of the order of 10%.

6. Conclusion

This paper has surveyed methods of testing for heteroskedasticity in a variety of models. Some of these tests are routinely provided as the standard output of computer packages, whilst others still only have spasmodic use. Our strategy was to treat existing tests as focusing on the validity of certain conditional moment restrictions, since the general results from that literature can be brought to bear on this specific problem. Our inquiry also revealed that quite a deal of work remains to be done to understand the performance of tests. Some of the issues raised relate to finite sample performance, and others to the poor performance of tests designed to be robust to departures from the auxiliary assumptions made in constructing them. To date, simulation studies have not addressed these questions very effectively, even neglecting to exploit what existing theory gives as the expected outcomes. As mentioned at various points in section 5, asymptotic theoretical analysis can predict whether certain tests would be expected to work well in a given experiment, yet such results have rarely been incorporated into the analysis. For example, the power of any conditional moment test against a sequence of local alternatives can be computed with formulae in Newey (1985a). Combining together these different sources of information seems mandatory if we are to fully understand the sampling behavior of tests of heteroskedasticity.

References

Ali, M.M. and Giacotto, C. (1984). A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model. J. Econometrics. 26, 355–373.

Arabmazar,A. and Schmidt, P. (1981). Further Evidence on the Robustness of the Tobit Estimator to Heteroscedasticity. J. Econometrics. 17, 253–258.

Ara,I. and King, M.L. (1991). Marginal Likelihood Based Tests of Regression Disturbances. Mimeo, Monash University.

Andrews, D.W.K. (1990). Tests for Parameter Instability and Structural Change with Unknown Change Point. Cowles Foundation Discussion Paper No. 943, Yale University.

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica, 59, 817–858.

Barnard, G.A. (1963). Comment. J. Roy. Statist. Soc. Ser. B. 25, 294.

Bartlett, M.S. (1937). Properties of Sufficiency and Statistical Tests. Proc. Roy. Soc. London Ser. A. 160, 268–282.

Bewley,R. and Theil, H. (1987). Monte Carlo Testing for Heteroscedasticity in Equation Systems. In T. B. Fomby and G.F. Rhodes, Adv. Econometrics, Vol 6 (J.A.I. Press), 1–15.

Bickel, P. (1978). Using Residuals Robustly: Tests for Heteroscedasticity and Nonlinearity. Ann. Statist. 6, 266–291.

Breusch, T.S. and Pagan, A.R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica. 47, 1287–1294.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. J. Econometrics. 31, 307–327.

Bollerslev, T., R.Y. Chou and Kroner, K.F. (1990). ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. Working Paper No. 97, Kellogg Graduate School of Management, Northwestern University.

Cameron, C. (1991). Regression Based Tests of Heteroscedasticity in Models where the Variance Depends on the Mean. Working Paper No. 379, University of California, Davis.

Cameron, A.C. and Trivedi, P. K. (1990). Regression Based Tests for Overdispersion in the Poisson Model." J. Econometrics. 46, 347–364.

Cameron, A.C. and P.K. Trivedi (1991). Conditional Moment Tests with Explicit Alternatives. Mimeo, University of Indiana.

Carroll, R. J. (1982). Adapting for Heteroscedasticity in Linear Models. An. Statist. 10, 1224–1233.

Chesher, A. and Spady, R. (1990) Asymptotic Expansions of The Information Matrix Test Statistic. Econometrica. 59, 787–815.

Collings, B.J. and Margolin, B.H. (1985). Testing Goodness of Fit for the Poisson Assumption When Observations Are Not Identically Distributed. J. Amer. Statist. Assoc. 80, 411–418.

Conniffe, D. (1990). Applying Estimated Score Tests in Econometrics. Paper presented to the Sixth World Congress of the Econometric Society, Barcelona.

Cox, D.R. (1983). Some Remarks on Overdispersion. Biometrika. 70, 269–274.

Dastoor, N.K. (1990). The Asymptotic Behaviour of Some Tests for Heteroskedasticity. Mimeo, University of Alberta.

Davidson,R. and Mackinnon, J.G. (1984). Convenient Specification Tests for Logit and Probit Models. J. Econometrics. 25, 241–262.

Effron, B. (1986). Double Exponential Families and their Use in Generalized Linear Regression. J. Amer. Statistical Assoc. 81, 709–721.

Eicker, F. (1967). Limit Theorems for Regressions with Unequal and Dependent Errors. Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 59–82, Berkeley: University of California

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

Engle, R.F. and Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances. Econometric Rev. 5, 1–50.

Engle, R.F., Lillien, L.M. and Robins, R.P. (1987). Estimating Time Varying Risk Premia in the Time Structure. Econometrica. 55, 391–407.

Engle, R. F. and Gonzalez–Rivera, G. (1991). Semiparametric ARCH Models. Journal of Business and Economic Statistics, 9, 345–359.

Evans,M.A. and King, M.L. (1985). A Point Optimal Test for Heteroscedastic Disturbances. J. Econometrics. 27, 163–178.

Evans, M.A. and King, M.L. (1988). A Further Class of Tests for Heteroscedasticity. J. Econometrics. 37, 265–276.

Friedman, B.M. and Laibson, D. I. (1989). Economic Implications of Extraordinary Movements in Stock Prices. Brookings Papers in Economic Activity. 2/89, 137–189.

Glesjer, H. (1969). A New Test for Heteroscedasticity. J. Amer. Statist. Assoc. 64, 316–323.

Godfrey, L. (1978). Testing for Multiplicative Heteroscedasticity. J. Econometrics, 8, 227–236.

Goldfeld, S. and Quandt, R. (1965). Some Tests for Heteroscedasticity. J. Amer. Statist. Assoc. 60, 539–547.

Gurmu, S. and Trivedi, P. K. (1990). Variable Augmentation Specification Tests in the Exponential Family. Mimeo, University of Indiana.

Griffiths, W.E. and Surekha, K. (1986). A Monte Carlo Evaluation of the Power of Some Tests for Heteroscedasticity. J. Econometrics. 31, 219–231.

Hansen, B.E. (1990). Lagrange Multiplier Tests for Parameter Instability in Non–linear Models. Mimeo, University of Rochester.

Hansen, L.P (1982). Large Sample Properties of Generalized Method of Moments Estimators. Econometrica. 50, 1029–1054.

Harris, P. (1985). An Asymptotic Expansion for the Null Distribution of the Efficient Score Statistic. Biometrika. 72, 653–659.

Harrison, M.J. (1980). The Small Sample Performance of the Szroeter Bounds Test for Heteroskedasticity and A Simple Test for Use When Szroeter's Test is Inconclusive. Bulletin of Economic and Social Research. 42, 235–250.

Harrison, M.J. and McCabe, B.P.M. (1979). A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals. J. Amer. Statist. Assoc. 74, 494–499.

Hausman, J. (1978). Specification Tests in Econometrics. Econometrica. 46, 1251–1271.

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

Higgins, M.L. and Bera, A.K. (1989). A Class of Nonlinear ARCH Models: Properties, Testing and Applications. Working paper 89–1554, Bureau of Economic and Business Research, University of Illinois, Urbana–Champaign.

Honda, Y. (1988). A Size Correction to the Lagrange Multiplier Test for Heteroskedasticity. J. Econometrics. 38, 375–386.

Imhof, J.P. (1961). Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika. 48, 419–426.

Jorgenson, B. (1987). Exponential Dispersion Models. J. Roy. Statist. Soc. Ser. B. 49(2), 127–162.

Jarque, C.M. and Bera, A. K. (1982). Efficient Specification Tests for Limited Dependent Variable Models. Econom. Lett. 9, 153–160.

Kamstra, M. (1990). A Neural Network Test for Heteroskedasticity. Discussion Paper No. 91–06 Simon Fraser University.

Kennan, J. and Neumann, G. R. (1988). Why Does the Information Matrix Test Reject So Often? Working Paper in Economics E–88–10, Hoover Institution, Stanford University.

King, M.H. and Hillier, G. H. (1985). Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model. J. Roy. Statist. Soc. Ser. B. 47, 98–102.

Koenker, R. (1981). A Note on Studentizing a Test for Heteroscedasticity. J. Econometrics. 17, 107–112.

Koenker, R. and Bassett, G. S. (1982). Robust Tests for Heteroscedasticity Based on Regression Quantiles. Econometrica. 50, 43–61.

Lee, L–F, (1986). Specification Tests for Poisson Regression Models. Internat. Econom. Rev. 27, 689–706.

Lee, L.–F and Maddala, G.S. (1985). The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroscedasticity and Nonnormality in the Tobit Model. Internat. Econom. Rev. 26, 1–20.

McCullagh, P. and Nelder, J.A. (1983). Generalized Linear Models. London: Chapman and Hall.

McCulloch, J.H. (1985). On Heteros*edasticity. Econometrica. 53, 483.

McDonald, J.B. and Newey, W. (1988). Partially Adaptive Estimation of Regression Models via the Generalized t Distribution. Econometric Theory. 4, 428–457.

McCabe, B.P.M. (1986). Testing for Heteroscedasticity Occurring at Unknown Points. Comm. Statist. Theory and Methods. 15, 1597–1613.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica. 59, 347–370.

Newey, W.D. (1984). A Method of Moments Interpretation of Sequential Estimators. Econom. Lett. 14, 201–206.

Newey, W.K. (1985a). Maximum Likelihood Specification Testing and Conditional Moment Tests. Econometrica. 53, 1047–1070.

Newey, W.K. (1985b). Generalized Method of Moments Specification Testing. J. Econometrics. 29, 229–256.

Newey, W.K. and Powell, J. L. (1987). Asymmetric Least Squares Estimation and Testing. Econometrica. 55, 819–847.

Pagan, A.R. (1986). Two Stage and Related Estimators and their Applications. Rev. Econom. Stud. 53, 517–538.

Pagan, A.R. and Hall, A.D. (1983). Diagnostic Tests as Residual Analysis. Econometric Rev. 2, 159–218.

Pagan, A.R. and Sabau, H. (1987a). Consistency Tests for Heteroskedasticity and Risk Models. Mimeo, University of Rochester.

Pagan, A.R and Sabau, H. (1987b). On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models. Working Paper No. 104, Center for Economic Research, University of Rochester.

Pagan, A.R. and Schwert, G. W. (1990). Alternative Models for Conditional Stock Volatility. J. Econometrics. 45, 267–290.

Pagan, A.R. and Vella, F.X. (1989). Diagnostic Tests for Models Based On Individual Data. Journal of Applied Econometrics. S4, S29–S52.

Pesaran, M. H. and Pesaran, B. (1987). MICROFIT: An Interactive Econometric Software Package. Oxford University Press, Oxford)

Phillips, P.C.B. and Loretan, M. (1990). Testing Covariance Stationarity Under Moment Condition Failure With an Application to Common Stock Returns. Cowles Foundation Discussion Paper No. 947, Yale University.

Potscher, B.M. and Prucha, I.R. (1986). A Class of Partially Adaptive One–step M Estimators for the Nonlinear Regression Model with Dependent Observations. J. Econometrics. 25, 219–251.

Powell, J.L. (1986). Censored Regression Quantiles. J. Econometrics. 32, 143–155.

Ramsey, J. (1969). Tests for Specification Error in the General Linear Model. J. Roy. Statist. Soc. Ser. B. 31, 250–271.

Robinson, P.M. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. Econometrica. 55, 875–891.

Sabau, H.C.L. (1988). Some Theoretical Aspects of Econometric Inference With Heteroskedastic Models. Unpublished doctoral dissertation, Australian National University.

Skeels, C.L. and Vella, F. (1991). A Monte Carlo Investigation of the Performance of Conditional Moment Tests in Tobit and Probit Models. Mimeo, Australian National University.

Smith, R.J. (1990). An Edgeworth Approximation for the Null Distribution of the Outer Product Form of the Score or Lagrange Multiplier Statistic. Mimeo, University of Cambridge.

Szroeter, J. (1978). A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models. Econometrica. 46, 1311–1328.

Tauchen, G. (1985). Diagnostic Testing and Evaluation of Maximum Likelihood Models. J. Econometrics. 30, 415–443.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. Econometrica. 26,24–36.

Tse, Y.K. and Phoon, C.T. (1985). Testing for Heteroscedasticity in a Dynamic Simultaneous Equation Model. Comm. Statist. Theory and Methods. 14, 1283–1300.

Whang, Y.J. and Andrews, D.W.K. (1991). Tests of Specification for Parametric and Semiparametric Models. Cowles Foundation Discussion Paper No. 968, Yale University.

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

White, H. (1987). Specification Testing in Dynamic Models." in T. Bewley (Ed.) Advan. Econometrics. Fifth World Congress, Vol 1, 1–58, Cambridge University Press, New York.

Wooldridge, J.M. (1990). A Unified Approach to Robust, Regression–Based Specification Tests. Econometric Theory. 6, 17–43.