

# TESTS FOR HETEROSKEDASTICITY<sup>1</sup>

Phoebus J. Dhrymes

Class Notes G6412, February 18, 1992; revised February 28, 1996.

p. 111ff. Replace the entire text of the section by the following text.

## 2.3.1 An *Ad Hoc* Approach

As we observed earlier, **heteroskedasticity** is the case in which the error process of the GLM,  $\{u_t : t = 1, 2, \dots\}$ , is one of independent, nonidentically distributed, zero mean random variables such that

$$E(u_t^2) = \sigma_t^2, \quad t = 1, 2, \dots .$$

Thus, the covariance matrix of the previous section,  $\Sigma$ , is a diagonal matrix. As we also remarked earlier, it is not possible to estimate consistently this matrix, and thus generalized least squares, or feasible Aitken estimators are not available.

This situation arises most commonly in cross-sectional studies. For example, suppose we are studying some aspect of firms' behavior, involving their sales, or profits, or advertising expenditures, utilizing information on individual firms; or, an aspect of household behavior, involving their income and consumption expenditures, utilizing information on individual households. It is clear in both cases that, typically, we have a very large variation in the size of these entities, from firm to firm or household to household.

Regarding the error specification, one may be prepared to accept the proposition that a firm's, or household's, "error term" is independent of that of another's. On the other hand, it would be rather **inappropriate to assert that the error terms in all households or in all firms have a common variance**. Would, for example, the residual variation be of the same order of magnitude for a firm or a household which is 1000

---

<sup>1</sup>©Phoebus J. Dhrymes, 1996.

Preliminary material; not to be reproduced or disseminated without permission of the author.

times the “size” of another? In such cases, empirical practitioners usually resort to normalizing all variables in the model by some “size” variable, and treating the resulting model as if it obeyed all standard assumptions of the GLM. One typically arrives at this procedure by viewing the residuals of the OLS regression in connection with various “size” candidates, formally or informally, and then deciding that a size variable, say  $A_t$ , is appropriate. This is equivalent to the assertion that

$$\Sigma = \sigma^2 A, \quad A = \text{diag}(A_1, A_2, \dots, A_T)$$

where  $A$  a **known matrix**, and then carrying out an Aitken procedure; the latter simply amounts to normalizing corresponding observations by  $A_t^{(1/2)}$ . If this is what one elects to do, the entire discussion of the previous section is fully applicable.

### 2.3.2 A Systematic Approach

It is rather simple to formalize, and generalize, the procedure above in terms of the **variance components** model. In this context, one assumes that

$$\sigma_t^2 = \sigma_0 + \sum_{i=1}^s \sigma_i z_{ti} = \sigma_0 + z_t \sigma, \quad t = 1, 2, \dots$$

The problem is then posed as

$$\min_{\theta} \sum_{t=1}^T \left( \frac{y_t - x_t \beta}{\sigma_t^2} \right)^2, \quad \theta = (\beta', \sigma_0, \sigma')',$$

presumably, subject to the condition  $\hat{\sigma}_0 + z_t \hat{\sigma} > 0$ .

Actually, in order to obviate the necessity of imposing the nonnegativity constraint above, it is just as simple to write the variance component specification as

$$\sigma_t^2 = e^{\sigma_0 + z_t \sigma},$$

so that, e.g., the **homoskedasticity** case corresponds to the statement  $\sigma_1 = \sigma_2 = \dots = \sigma_s = 0$  and, thus,  $\sigma_t^2 = e^{\sigma_0}$ .

Following in this vein, and for the sake of simplicity basing our exposition on the linear specification, one might consider the general model above but impose, through Lagrange multipliers, the conditions that

$\sigma_i = 0$ , for all  $i = 1, 2, \dots, s$ . Tests for the presence of heteroskedasticity, then, are simple implementations of the Lagrange Multiplier test discussed earlier. Unfortunately, in the context of least squares we have no formal derivation for the estimator of  $\sigma_t^2$ , under the null. Thus, the Lagrange multiplier route is not conveniently available in a least squares context. Instead we shall follow here the “diagnostic” approach that we shall also follow in the case of autocorrelation, viz. we first estimate the parameters of the GLM by least squares; then we use the residuals from that regression to estimate the variance component parameters and test the null  $H_0 : \sigma = 0$ .

To that end, note that the residuals of the regression obey

$$\hat{u}_t = u_t - x_t'(X'X)^{-1}X'u, \quad \text{so that} \quad (2.1)$$

$$\hat{u}_t^2 = u_t^2 + u'X(X'X)^{-1}x_t'x_t(X'X)^{-1}X'u - 2u'X(X'X)^{-1}x_t'u_t,$$

and it is easily seen that if the sample is large enough  $\hat{u}_t^2 \sim u_t^2$ , in the sense that the left member behaves essentially as the right member. To demonstrate this note that

$$\hat{u}_t^2 - u_t^2 = \frac{u'X}{\sqrt{T}} \left( \frac{X'X}{T} \right)^{-1} \frac{x_t'x_t}{\sqrt{T}} \left( \frac{X'X}{T} \right)^{-1} \frac{X'u}{T} - 2 \frac{u'X}{\sqrt{T}} \left( \frac{X'X}{T} \right)^{-1} \frac{x_t'u_t}{\sqrt{T}},$$

and the right member converges to zero, at least in probability, for every  $t$ , provided  $[(\max_{t \leq T} |x_t|)/\sqrt{T}] \rightarrow 0$ ; this, however is implied by the condition that  $(X'X/T) \rightarrow M_{xx} > 0$ .

The expectation of  $u_t^2$  is given by

$$E(u_t^2) = \sigma_0 + z_t\sigma, \quad \text{which under the null becomes} \quad E(u_t^2) = \sigma_0. \quad (2.2)$$

This provides the motivation for the following diagnostic procedure:<sup>2</sup> Regress the square of the residuals on the vector  $z_t$ , together with a constant term and test the hypothesis that  $\sigma = 0$ . From previous discussion we easily determine that the  $R^2$  of this regression may be represented as

$$R^2 = \frac{\hat{\sigma}'Z^*Z^*\hat{\sigma}}{h'h}, \quad (2.3)$$

---

<sup>2</sup>As we shall see later on, this procedure is essentially a Lagrange Multiplier test in the context of a ML formulation of the problem.

where  $Z^* = (I - ee'/T)Z$ ,  $h = (\hat{u}_1^2 - m, \hat{u}_2^2 - m, \dots, \hat{u}_T^2 - m)'$ , and

$$h'h = \sum_{t=1}^T (\hat{u}_t^2 - m)^2, \quad m = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2.$$

Unfortunately, the distribution of  $R^2$ , or some other suitable function of it, cannot be explicitly established for small samples. We may, however, find its limiting distribution i.e. its distribution as the sample size  $T \rightarrow \infty$ . We shall postpone doing this, i.e. giving a sketch of its derivation, until we have considered the ML formulation of this problem. To this end, we write the loglikelihood function (LF) of the problem as

$$L(\beta, \sigma_0, \sigma; y, X) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(\sigma_0 + z_t \sigma) - \frac{1}{2} (y - X\beta)' \Sigma^{-1} (y - X\beta),$$

$$\Sigma = \text{diag}(\sigma_0 + z_1 \sigma, \sigma_0 + z_2 \sigma, \dots, \sigma_0 + z_T \sigma), \quad (2.4)$$

and maximize it under the constraint  $\sigma = 0$ , i.e. we operate with  $L^* = L + T\lambda'\sigma$ , where  $\lambda$  is the ( $s$ -element) vector of Lagrange multipliers. The first order conditions are

$$\frac{\partial L^*}{\partial \beta} = (y - X\beta)' \Sigma^{-1} X = 0,$$

$$\frac{\partial L^*}{\partial \sigma_0} = -\sum_{t=1}^T \frac{1}{2(\sigma_0 + z_t \sigma)} + \frac{1}{\sigma_0^2} (y - X\beta)' (y - X\beta) = 0,$$

$$\frac{\partial L^*}{\partial \sigma} = -\sum_{t=1}^T \frac{z_t}{2(\sigma_0 + z_t \sigma)} + \frac{1}{2} \sum_{t=1}^T \left( \frac{(y_t - x_t \beta)^2 z_t}{(\sigma_0 + z_t \sigma)^2} \right) + T\lambda' = 0,$$

$$\frac{\partial L^*}{\partial \lambda} = T\sigma' = 0. \quad (2.5)$$

Substituting the last condition in the first two equations above we find

$$\hat{\beta} = (X'X)^{-1} X'y, \quad \hat{\sigma} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2; \quad (2.6)$$

Finally, from the next to the last equation in the equation set above, we find

$$\lambda = \frac{1}{T} \sum_{t=1}^T (u_t^2 - \sigma_0) z_t', \quad z_t^{*'} = z_t' - \bar{z}. \quad (2.7)$$

Under the null hypothesis  $H_0 : \sigma = 0$ ,  $\lambda$  is the mean of a sequence of **zero mean** independent non identically distributed random vectors; on the other hand under the alternative  $H_1 : \sigma \neq 0$ , the mean of the vectors is **non zero**. This then motivates a test of the hypothesis of **homoskedasticity** as a test that the mean of  $\lambda$  is **zero**. Now,  $\lambda$  as presented in Eq. (2.7) is not a statistic since it involves the unobservables  $u_t^2$  and  $\sigma_0$ . These, however, may be estimated by  $\hat{u}_t^2 = (y_t - x_t \hat{\beta})^2$  and  $\hat{\sigma}_0$ , respectively, where  $\hat{\beta}$ ,  $\hat{\sigma}_0$  are as determined in Eq. (2.6), and thus we may estimate  $\lambda$ , the Lagrange multiplier (LM), by

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^T (\hat{u}_t^2 - \hat{\sigma}_0) z_t^{*'} \quad (2.8)$$

### 2.3.3 Asymptotic Properties of the LM

In order to use the estimator  $\hat{\lambda}$  for testing the null hypothesis of homoskedasticity it is necessary to establish its distribution. The latter, however, is not easily established due to the complexity of the LM representation; we may, however, establish its limiting distribution, i.e. the distribution of  $\sqrt{T} \hat{\lambda}$  as  $T \rightarrow \infty$ . From probability theory it is known that if two sequences of random variables (or vectors)  $X_n, Y_n$  have the properties that  $X_n$  converges in distribution to  $X$  and  $(X_n - Y_n)$  converges **in probability to zero** then  $Y_n$  converges in distribution to  $X$ . We shall now use this result to simplify our problem by noting that

$$\begin{aligned} \phi_T &= \frac{1}{\sqrt{T}} \sum_{t=1}^T [(\hat{u}_t^2 - \hat{\sigma}_0) - (u_t^2 - \sigma_0)] z_t^{*'} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (\hat{u}_t^2 - u_t^2) z_t^{*'} - \frac{1}{\sqrt{T}} (\hat{\sigma}_0 - \sigma_0) \sum_{t=1}^T z_t^{*'} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T (\hat{u}_t^2 - u_t^2) z_t^{*'}, \end{aligned}$$

the last equality being valid since  $z_t^{*'} = z_t' - \bar{z}$ . The entity in the last equation above may be shown to converge to zero in probability; it therefore follows that

$$\sqrt{T} \hat{\lambda} \sim \frac{1}{\sqrt{T}} \sum_{t=1}^T (u_t^2 - \sigma_0) z_t^{*'}, \quad (2.9)$$

and the right member involves a sequence of independent non identically distributed random vectors with mean zero and covariance matrix

$$\text{Cov} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T (u_t^2 - \sigma_0) z_t^{*'} \right] = 2\sigma_0^2 \frac{Z^{*'} Z^*}{T} \rightarrow 2\sigma_0^2 M_{zz}^* > 0.$$

By the Lindeberg Central Limit Theorem we obtain

$$\sqrt{T} \hat{\lambda} \xrightarrow{d} N(0, 2\sigma_0^2 M_{zz}^*).$$

This means that

$$\psi_T = \frac{T \hat{\lambda}' (Z^{*'} Z^* / T)^{-1} \hat{\lambda}}{2\hat{\sigma}_0^2} = \frac{T^2 \hat{\lambda}' (Z^{*'} Z^*)^{-1} \hat{\lambda}}{2\hat{\sigma}_0^2} \xrightarrow{d} \zeta \sim \chi_s^2$$

The preceding discussion may be summarized in

**Proposition 1.** Consider the standard GLM, with the additional assumption that the errors constitute a sequence of independent, zero mean, normal random variables, and suppose it is desired to test the hypothesis of homoskedasticity as against the alternative of heteroskedasticity of the form,

$$\text{Var}(u_t) = \sigma_t^2 = \sigma_0 + z_t \sigma,$$

where the variables in  $z_t$  are observable, fixed or otherwise stochastically independent of the error terms of the GLM. A test of the hypothesis

$$H_0 : \sigma = 0, \quad \text{as against} \quad H_1 : \sigma \neq 0,$$

may be carried out in terms of the estimator of the LM of Eq. (2.8), through the test statistic

$$\psi_T = \frac{T^2 \hat{\lambda}' (Z^{*'} Z^*)^{-1} \hat{\lambda}}{2\hat{\sigma}_0^2} \xrightarrow{d} \zeta \sim \chi_s^2.$$

**Remark 1.** It is interesting to note that the test statistic above, which has been motivated as a LM test statistic in the context of ML estimation, may be obtained without such motivation and **without** the normality assumption, simply as

$$\psi_T^* = TR^2,$$

where  $R^2$  refers to the regression of  $\hat{u}_t^2$  on a constant term and the vector  $z_t$ . To see this, return to Eq. (2.3) and recall from standard regression theory that

$$\hat{\sigma} = (Z^{*'} Z^*)^{-1} Z^{*'} h, \quad \text{and} \quad TR^2 = \psi_T^* = \frac{h' Z^{*'} (Z^{*'} Z^*)^{-1} Z^{*'} h}{h' h / T}. \quad (2.10)$$

Notice further that  $m = \hat{\sigma}_0$ , and consequently that

$$\hat{\lambda} = \frac{1}{T} Z^{*'} h, \quad \psi_T = \frac{h' Z^{*'} (Z^{*'} Z^*)^{-1} Z^{*'} h}{2\hat{\sigma}_0^2}.$$

A comparison of  $\psi_T$  and  $\psi_T^*$  shows that the two statistics are (numerically) identical in their numerators, but differ in their denominators. By construction, the denominator of  $\psi_T$  converges in probability to  $2\sigma_0^2$ ; to examine the behavior of the denominator of  $\psi_T^*$  we note that, from the argument leading to Eq. (2.9),

$$\frac{h' h}{T} \sim \frac{1}{T} \sum_{t=1}^T (u_t^2 - \sigma_0)^2 \xrightarrow{P} 2\sigma_0^2,$$

if the errors of the GLM are **normally distributed** and otherwise converges to  $E(u_t^4) - \sigma_0^2$ . In the case of normality  $E(u_t^4) = 3\sigma_0^2$ , and we see that in general

$$\frac{h' h}{T} \xrightarrow{P} \mu_4 - \sigma_0^2, \quad \mu_4 = E(u_t^4).$$

Thus, the conformity test based on the test statistic  $\psi_T^*$  and the LM test, based on the test statistic  $\psi_T$  are asymptotically equivalent, **whether or not** the errors of the GLM are normally distributed.

When normality holds, the presumption is that  $\psi_T \geq \psi_T^*$ , in view of the fact that, in all likelihood,  $2\hat{\sigma}_0^2 \leq h' h / T$ .

### 2.3.4 Heteroskedasticity and OLS Based Tests

The problem we examine here is this: suppose we have obtained estimates from a standard GLM, and we wish to carry out tests of significance, but we suspect heteroskedasticity. As we had discussed at an earlier section, the estimated “covariance” matrix obtained through least squares is inappropriate and it does not **systematically either over- or under-estimate** the proper covariance matrix of the OLS estimators.

Wishing very much to avoid the difficulties just discussed, is there any way in which we can use the OLS results and still carry out proper tests? The answer is generally in the affirmative. Although the topic we shall now examine has been in the statistics literature for some time, see, e.g., Eicker (1963), (1967) or Rao (1970), (1973), it has been introduced into the literature of econometrics relatively recently, in the very interesting paper by White (1980). An extension of White's results may be found in Dhrymes (1992).

We recall that the covariance matrix of the OLS estimator, in the case of heteroskedasticity, is given by  $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$ , while the typical computer program will print out the (estimate of the) covariance as  $\hat{\sigma}^2(X'X)^{-1}$ , where  $\hat{\sigma}^2$  is the sum of squares of the OLS residuals divided by  $T$ , or more typically, by  $(T - n - 1)$ . As we have shown earlier the difference

$$C = (X'X)^{-1}X'\Sigma X(X'X)^{-1} - E(\hat{\sigma}^2)(X'X)^{-1}$$

is an **indefinite matrix** so that, in general, the “ $t$ -ratios” **can not be asserted to be consistently higher or lower** than they would have been, had we used **the appropriate** covariance matrix. Thus, such “ $t$ -ratios” are **completely useless** from a **formal** point of view, and it is highly desirable to obtain at least a consistent estimator of the proper covariance matrix. But, the reader might interject, why not obtain the feasible Aitken, or the GLS estimator? A little reflection will show that it is impossible to obtain such estimators, if we are only prepared to assert that  $E(u_t^2) = \sigma_t^2$ . This is so since in such a case we have **only one observation with which to estimate** the parameter  $\sigma_t^2$ ! Perhaps, it is possible do the next best thing, viz., to estimate the covariance matrix of the OLS estimator. At first glance this might appear to be impossible as well, since it involves precisely the same parameters we have just proclaimed impossible to estimate (consistently). However, if we write

$$X'\Sigma X = \sum_{t=1}^T \sigma_t^2 x_t' x_t,$$

we realize that to estimate the covariance matrix of the OLS estimator, we do not **actually need (consistent) estimators of the individual**

**variances**,  $\sigma_t^2$ , but only a consistent estimator of the **weighted sum of such variances**, and the weights are known! It may be shown that under the standard assumptions on the explanatory variables, including the case of exogenous and lagged dependent variables, the estimator

$$\hat{\Phi} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2 x_t' x_t.$$

is at least a consistent estimator of

$$\Phi = \text{p} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sigma_t^2 x_t' x_t.$$

Thus, we may use the matrix  $(X'X)^{-1}T\hat{\Phi}(X'X)^{-1}$ , in forming the “ $t$ -ratios” for carrying out tests of significance. The reader should bear in mind, however, that such tests are **only asymptotically valid** and that, in the limit, the “ $t$ -ratio” in question is **normally distributed** with mean zero and variance one, so what we have is a “normal” test, **not** a  $t$  test.

**Remark 2.** If we were to adapt the results in Dhrymes (1992), to the case of nonstochastic explanatory variables,  $x_t$ , in the context of the model  $y_t = x_t\beta + y_{t-1}\lambda + u_t$ , with  $\lambda \in [0, 1)$ , or  $\lambda(-1, 1)$  we have that<sup>3</sup>

$$\hat{\theta} \xrightarrow{P} \theta_0, \quad \theta = (\beta', \lambda)', \quad \hat{\theta} = (Z'Z)^{-1}Z'y, \quad Z = (X, y_{-1}).$$

The limiting distribution of OLS estimator above is given by

$$\sqrt{T}(\hat{\theta} - \theta_0) \sim N(0, M_{zz}^{-1}\Phi M_{zz}^{-1}), \quad M_{zz} = \text{p} \lim_{T \rightarrow \infty} \frac{1}{T} Z'Z,$$

and the estimator of the covariance matrix of the limiting distribution is

$$\left(\frac{Z'Z}{T}\right)^{-1} \hat{\Phi} \left(\frac{Z'Z}{T}\right)^{-1} \xrightarrow{P} M_{zz}^{-1}\Phi M_{zz}^{-1}.$$

Occasionally, the estimator above is referred to in the literature as the “heteroskedasticity corrected” estimator of the covariance matrix.

---

<sup>3</sup>Actually the convergence of the estimators discussed below is convergence a.c. (or with probability one); since we have not introduced this concept, we express the results in terms of convergence in probability, **which is implied by convergence a.c.**