Econometrica, Vol. 44, No. 3 (May, 1976)

ESTIMATING REGRESSION MODELS WITH MULTIPLICATIVE HETEROSCEDASTICITY

BY A. C. HARVEY[1]

A regression model in which the disturbances exhibit a certain type of heteroscedasticity is considered. Maximum likelihood methods of estimation are developed and compared with the two-step estimation procedure. A likelihood ratio test for heteroscedasticity is suggested.

1. INTRODUCTION

HETEROSCEDASTIC REGRESSION MODELS in which the variance of the disturbance term is assumed to be proportional to one of the regressors raised to a certain power have been considered by a number of writers, including Geary [**2**], Goldfeld and Quandt [**4**], Kmenta [**8**, pp. 256–264], Lancaster [**9**], and Park [**11**]. The variance of the ith disturbance term in such cases may be written

$$(1)\qquad \sigma_i^2 = \sigma^2 X_i^{\lambda}.$$

When λ is unknown Park [**11**] has proposed a two-step estimation procedure. The first step consists of taking the (natural) logarithms of the squares of the residuals resulting from the application of ordinary least squares (OLS) to the original equation. These are then regressed on $\log X_i$, thereby yielding estimates of λ and $\log \sigma^2$ which can be used to construct feasible generalized least squares (GLS) estimates of the coefficients in the original regression equation.

In this note the two-step procedure is examined in detail for a more general model than that implied by (1). Maximum likelihood methods are also considered and the properties of estimators obtained by various procedures are compared. Details of the application of similar procedures to other classes of heteroscedastic regression models may be found in [**3**, **5**, and **12**].

2. THE TWO-STEP PROCEDURE

A general formulation of a regression model with multiplicative heteroscedasticity is:

$$(2)\qquad y_i = x_i'\beta + u_i \qquad (i = 1, \dots, n),$$

$$(3)\qquad \sigma_i^2 = e^{z_i'\alpha} \qquad (i = 1, \dots, n),$$

where x_i is a $k \times 1$ vector of observations on the independent variables, β is a $k \times 1$ vector of parameters, z_i is a $p \times 1$ vector of observations on a set of variables which are usually, though not necessarily, related to the regressors in (2), and α is a $p \times 1$ vector of parameters. The u_i's are disturbance terms which are independently and normally distributed with zero means. The first element in z_i will always be assumed to be a constant term. One of the conditions for estimators to exist

[1] I should like to thank G. D. A. Phillips and J. D. Sargan for helpful comments on an earlier draft. Any errors are mine, however.

is that all the elements in z_i be bounded from below for all i from 1 to n. In model (1), which is a special case of (3) with $\alpha' = [\log \sigma^2 \quad \lambda]$ and $z_i' = [1 \quad \log X_i]$, fulfilment of this condition requires that X_i be positive for all i.

The two-step estimator of α,

$$(4) \qquad \tilde{\alpha} = \left[\sum_{i=1}^{n} z_i z_i'\right]^{-1} \sum_{i=1}^{n} z_i \log \hat{u}_i^2,$$

is based on the equation

$$(5) \qquad \log \hat{u}_i^2 = z_i'\alpha + w_i \qquad (i = 1, \dots, n),$$

where $\hat{u}_i$ is the ith residual resulting from the application of OLS to (2) and $w_i = \log(\hat{u}_i^2/\sigma_i^2)$.

For a regression model of the form (2) in which the disturbances have constant variance and plim $n^{-1} \sum_{i=1}^n x_i x_i'$ is equal to a fixed positive definite matrix, it may be shown [**13**, p. 379] that the OLS residuals converge in distribution to the true disturbances. This result may be extended to the heteroscedastic case by observing that

$$V(\hat{u}_i - u_i) = n^{-1}\sigma^2 x_i'\left[n^{-1} \sum_{i=1}^{n} x_i x_i'\right]^{-1} n^{-1} \sum_{i=1}^{n} \sigma_i^{-2} x_i x_i' \left[n^{-1} \sum_{i=1}^{n} x_i x_i'\right]^{-1} x_i \qquad (i = 1, \dots, n).$$

The additional assumption that plim $n^{-1} \sum_{i=1}^n \sigma_i^{-2} x_i x_i'$ is a fixed positive definite matrix is then sufficient to ensure that the variance of $\hat{u}_i - u_i$ converges to zero as $n \to \infty$. Since $E(\hat{u}_i - u_i) = 0$, it follows from Chebyshev's inequality that $\hat{u}_i - u_i$ converges in probability to zero. Hence, $\hat{u}_i$ converges in distribution to u_i (which is normally distributed), and so w_i converges in distribution to a variable, w_i^*, which is distributed as the logarithm of a χ^2 variate with one degree of freedom (see Mann and Wald [**10**]).

From [**1**, p. 943] we find that the expected value of a variable distributed as the logarithm of a χ^2 divided by its degrees of freedom,[2] v, is $\psi(v/2) - \log(v/2)$, where $\psi(s)$ is the psi (digamma) function defined as $d \log \Gamma(s)/d(s) = \Gamma'(s)/\Gamma(s)$, $\Gamma(s)$ being the Gamma function. The second, third, and fourth moments about the mean are $\psi^{(1)}(v/2)$, $\psi^{(2)}(v/2)$, and $\psi^{(3)}(v/2) + 3\psi^{(1)}(v/2)^2$, respectively; $\psi^{(m)}(s)$ denotes the mth derivative of $\psi(s)$. Evaluating these expressions [**1**, p. 260] for $v = 1$ gives

$$(6) \qquad Ew_i^* = \psi(\tfrac{1}{2}) - \log \tfrac{1}{2} = -1.2704,$$

$$(7) \qquad Vw_i^* = \psi^{(1)}(\tfrac{1}{2}) = 4.9348,$$

$$(8) \qquad \frac{E(w_i^* - Ew_i^*)^3}{(Vw_i^*)^{\frac{3}{2}}} = \frac{\psi^{(2)}(\frac{1}{2})}{[\psi^{(1)}(\frac{1}{2})]^{\frac{3}{2}}} = -1.5351, \quad \text{and}$$

$$(9) \qquad \frac{E(w_i^* - Ew_i^*)^4}{(Vw_i^*)^2} = \frac{\psi^{(3)}(\frac{1}{2})}{[\psi^{(1)}(\frac{1}{2})]^2} + 3 = 7.0000 \qquad (i = 1, \dots, n).$$

[2] That is, $\log(\chi^2/v)$.

Let $\tilde{\alpha}_1$ denote the first element in $\tilde{\alpha}$. Expression (6) then implies that this is an inconsistent estimator of α_1 (the first element in α) since plim $\tilde{\alpha}_1 = \alpha_1 - 1.2704$. However, the other $p - 1$ elements in $\tilde{\alpha}$ are, subject to the usual conditions [**13**, Ch. 8], consistent estimators of the corresponding parameters in α. Since α_1 merely introduces a factor of proportionality into (3), the feasible GLS estimator of β constructed from $\tilde{\alpha}$ will, in general, be asymptotically efficient.

It will be seen from (7) that the asymptotic variance-covariance matrix of $\tilde{\alpha}$ is

$$(10) \qquad V\tilde{\alpha} = 4.9348\left[\sum_{i=1}^{n} z_i z_i'\right]^{-1}.$$

Finally, note that expressions (8) and (9) yield interesting information concerning the skewness and kurtosis of the disturbance term in (5).

3. MAXIMUM LIKELIHOOD ESTIMATION

Since the disturbances in the model presented in (2) and (3) are independently and identically distributed, the log-likelihood function is

$$(11) \qquad \log L = \frac{n}{2}\log 2\pi - \tfrac{1}{2}\sum_{i=1}^{n} z_i'\alpha - \tfrac{1}{2}\sum_{i=1}^{n} e^{-z_i'\alpha}(y_i - x_i'\beta)^2.$$

The inverse of the information matrix, which is equal to the asymptotic variance-covariance matrix of the maximum likelihood (ML) estimators, is therefore:

$$(12) \qquad \left\{ E\left[\begin{array}{c:c} \dfrac{\partial^2 \log L}{\partial\beta\,\partial\beta'} & \dfrac{\partial^2 \log L}{\partial\beta\,\partial\alpha'} \\ \hdashline \dfrac{\partial^2 \log L}{\partial\alpha\,\partial\beta'} & \dfrac{\partial^2 \log L}{\partial\alpha\,\partial\alpha'} \end{array}\right]_{\substack{\beta=\hat{\beta}\\ \alpha=\hat{\alpha}}} \right\}^{-1} = \left[\begin{array}{c:c} \left(\sum_{i=1}^{n} \sigma_i^{-2} x_i x_i'\right)^{-1} & 0 \\ \hdashline 0 & 2\left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \end{array}\right].$$

Comparison of (12) with (10) shows immediately that the two-step estimator of α is (asymptotically) inefficient. The standard deviations of the elements in $\tilde{\alpha}$ will exceed the standard deviations of corresponding ML estimators by almost 60 per cent.

The "method of scoring" (see, for example, [**12**]), provides a fairly straightforward means of obtaining ML estimators in this case.[3] The iterative equations for estimating β and α may, in view of (12), be written separately as

$$(13) \qquad \begin{aligned} \hat{\beta}^{(t+1)} &= \hat{\beta}^{(t)} + \left\{ E\left[\frac{\partial^2 \log L}{\partial\beta\,\partial\beta'}\right]_{\substack{\beta=\hat{\beta}^{(t)}\\ \alpha=\hat{\alpha}^{(t)}}} \right\}^{-1} \left[\frac{\partial \log L}{\partial\beta}\right]_{\substack{\beta=\hat{\beta}^{(t)}\\ \alpha=\hat{\alpha}^{(t)}}} \\ &= \hat{\beta}^{(t)} + \left(\sum_{i=1}^{n} e^{-z_i'\hat{\alpha}^{(t)}} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i\, e^{-z_i'\hat{\alpha}^{(t)}}(y_i - x_i'\hat{\beta}^{(t)}) \end{aligned}$$

[3] Cf. Kmenta [**8**, p. 264] who suggests a trial and error (search) procedure to estimate λ in equation (1).

and

$$
\begin{aligned}
\hat{\alpha}^{(t+1)} &= \hat{\alpha}^{(t)} + \left\{ E\left[\frac{\partial^2 \log L}{\partial\alpha\,\partial\alpha'}\right]_{\substack{\beta=\hat{\beta}^{(t)}\\ \alpha=\hat{\alpha}^{(t)}}} \right\}^{-1} \left[\frac{\partial \log L}{\partial\alpha}\right]_{\substack{\beta=\hat{\beta}^{(t)}\\ \alpha=\hat{\alpha}^{(t)}}} \\
&= \hat{\alpha}^{(t)} + \left(\sum_{i=1}^{n} z_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i[e^{-z_i'\hat{\alpha}^{(t)}}(y_i - x_i'\hat{\beta}^{(t)})^2 - 1],
\end{aligned} \tag{14}
$$

where $\hat{\alpha}^{(t)}$ and $\hat{\beta}^{(t)}$ are the estimates of α and β, respectively, obtained at the tth iteration. As regards starting values, $\hat{\beta}^{(0)}$ may be set equal to b, the OLS estimate of β in (2). Following Rutemiller and Bowers [**12**, p. 555], $\alpha_1^{(0)}$ may be set equal to $\log s^2$, where $s^2 = \Sigma_{i=1}^{n} (y_i - x_i'b)^2/n$, all the other elements in the vector $\alpha^{(0)}$ being zero. Alternatively we may have $\alpha^{(0)} = \tilde{\alpha}$, the two-step estimator defined in (4).

If a consistent estimator of α is used to give starting values for (14), only one iteration will be needed to produce an estimator with the same asymptotic distribution as the ML estimator [**6**, p. 240]. Provided, therefore, that α_1 is estimated by $\tilde{\alpha}_1 + 1.2704$, the two-step procedure produces a consistent estimator of α and (14) yields the expression

$$
\alpha^* = \tilde{\alpha} + \phi + 0.2807\left[\sum_{i=1}^{n} z_i z_i'\right]^{-1} \sum_{i=1}^{n} z_i\, e^{-z_i'\tilde{\alpha}}\hat{u}_i^2 \tag{15}
$$

where ϕ is a $p \times 1$ vector in which the first element is 0.2704, and the remaining elements are zero. This modified three-step estimator of α (cf. [**5**]) is asymptotically efficient. If it is used to form a feasible GLS estimator it seems reasonable to suppose that such an estimator will have better small sample properties than the corresponding two-step estimator.

For all the estimators of α described in this section (large sample) test of significance may be carried out using (12) and these tests will, of course, be more powerful than corresponding tests based on $\tilde{\alpha}$. In fact, if we are considering a (two-sided) test on one of the elements in the vector α, the asymptotic relative efficiency of the test based on the two-step estimator will be equal to the ratio of the appropriate variances in (10) and (12) (see [**7**, Ch. 25]). For any element in α this is simply $2/4.9348 = 0.41$.

A likelihood ratio test may be used to test the hypothesis that the disturbances in the model are homoscedastic; i.e., $\alpha_2 = \alpha_3 = \ldots \alpha_p = 0$. This is based on the statistic

$$
-2\log\{L(b, s^2)/L(\hat{\beta}, \hat{\alpha})\} = n\log s^2 + n - \sum_{i=1}^{n} z_i'\hat{\alpha} - \sum_{i=1}^{n} e^{-z_i'\hat{\alpha}}(y_i - x_i'\hat{\beta})^2. \tag{16}
$$

However, when the first element in the vector $\partial \log L/\partial\alpha$ is evaluated for $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$ and set equal to zero, it will be observed that the second and fourth terms in (16) cancel out. Therefore the test statistic is simply

$$
n\log s^2 - \sum_{i=1}^{n} z_i'\hat{\alpha} \tag{17}
$$

and this is asymptotically distributed as χ^2 with $p - 1$ degrees of freedom under the null hypothesis.

4. CONCLUSION

By making the assumption that the disturbances in the regression model (2) are normally distributed it is possible to obtain estimates of the heteroscedasticity parameters, α, which are considerably more efficient than those obtained by a two-step procedure. This means that more powerful tests can be carried out on the estimates of α and, in addition, the small sample properties of the estimator of β are likely to be improved.

From the point of view of estimation, the multiplicative heteroscedasticity model considered here appears to be rather more attractive than the "additive" model in which either the variance or standard deviation of the ith disturbance term is assumed to be related to a linear combination of known variables [**3**, **5**, and **12**]. There are three reasons for this. Firstly, the likelihood function is bounded and no problems arise due to estimated variances being negative or zero. Secondly, the error terms in the two-step equation (5) are (asymptotically) homoscedastic and so the estimated covariance matrix of the two-step estimator, $\tilde{\alpha}$, is consistent. Finally, the likelihood ratio test has a much simpler form in the multiplicative model.

University of Kent at Canterbury

Manuscript received January, 1975; *revision received April*, 1975.

REFERENCES

[**1**] Abramowitz, M., and I. A. Stegum (eds.): *Handbook of Mathematical Functions*. New York: Dover Publications, Inc., 1965.
[**2**] Geary, R. C.: "A Note on Residual Heterovariance and Estimation Efficiency in Regression," *American Statistician*, 20 (1966), 30–31.
[**3**] Glejser, H.: "A New Test for Heteroscedasticity," *Journal of the American Statistical Association*, 64 (1969), 316–323.
[**4**] Goldfeld, S. M., and R. E. Quandt: *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland, 1972.
[**5**] Harvey, A. C.: "Estimation of Parameters in a Heteroscedastic Regression Model," Paper presented at the European Meeting of the Econometric Society, Grenoble, September, 1974.
[**6**] Jorgenson, D. W.: "Multiple Regression Analysis of a Poisson Process," *Journal of the American Statistical Association*, 56 (1961), 235–245.
[**7**] Kendall, M. G., and A. Stuart: *The Advanced Theory of Statistics*, Vol. 2, 3d ed. London: Charles Griffin and Company, Limited, 1973.
[**8**] Kmenta, J.: *Elements of Econometrics*. New York: Macmillan, 1971.
[**9**] Lancaster, T.: "Grouping Estimators on Heteroscedastic Data," *Journal of the American Statistical Association*, 63 (1968), 182–191.
[**10**] Mann, H. B., and A. Wald: "On Stochastic Limit and Order Relationships," *Annals of Mathematical Statistics*, 14 (1943), 217–226.
[**11**] Park, R. E.: "Estimation with Heteroscedastic Error Terms," *Econometrica*, 34 (1966), 888.
[**12**] Rutemiller, H. C., and D. A. Bowers: "Estimation in a Heteroscedastic Regression Model," *Journal of the American Statistical Association*, 63 (1968), 552–557.
[**13**] Theil, H.: *Principles of Econometrics*. Amsterdam: North-Holland, 1971.