



Working Papers in Econometrics and Applied Statistics

HETEROSKEDASTICITY

William E. Griffiths

No. 104 - June 1999

**Working Papers in Econometrics and Applied Statistics
Department of Econometrics
University of New England
Armidale, NSW 2351, Australia.**

<http://www.une.edu.au/econometrics/emetwp.htm>

ISSN 0157-0188
ISBN 1 86389 585 X

HETEROSKEDASTICITY

William E. Griffiths

University of New England

Contact information:

Professor W E Griffiths
Econometrics
University of New England
Armidale NSW 2351
Australia

Phone: 02 6773 2319
Fax: 02 6773 3607
Email: wgriffit@metz.une.edu.au

1. Introduction

A random variable y is said to be heteroskedastic if its variance can be different for different observations. Conversely, it is said to be homoskedastic if its variance is constant for all observations. The most common framework in which heteroskedasticity is studied in econometrics is in the context of the general linear model

$$y_i = x_i'\beta + e_i \quad (1.1)$$

where x_i is a K -dimensional vector of observations on a set of explanatory variables, β is a K -dimensional vector of coefficients which we wish to estimate and y_i denotes the i -th observation ($i = 1, 2, \dots, N$) on a dependent variable. The error term e_i is assumed to have zero mean and variance σ_i^2 , the i subscript on σ_i^2 reflecting the heteroskedasticity assumption that the variance of e_i can be different for different observations. Conditional on $x_i'\beta$, the dependent variable y_i will have mean $x_i'\beta$ and variance σ_i^2 . Thus, introducing heteroskedasticity into the general linear model means that we have a random variable y whose mean and variance can both change over observations.

The existence of heteroskedasticity is often encountered when using cross-sectional data on a number of microeconomic units such as firms or households. A common example is the estimation of household expenditure functions. In this case y_i represents expenditure by the i -th household on some commodity group such as food, and the explanatory variables include total expenditure (or income) and household size. It is usually postulated that expenditure on some commodity is more easily explained by conventional variables for households with low incomes than it is for households with high incomes. Low-income households do not have the option of extravagant food tastes; comparatively, they have few choices, and are almost forced to spend a particular portion of their income on food. High-income households, on the other hand, could have simple food tastes or extravagant food tastes. They might dine on caviar or spaghetti, while their low-income counterparts have to take the spaghetti. The lower predictive ability of the model for high incomes can be captured by specifying a variance σ_i^2 which is larger when income is larger. Other possible

examples include data on costs, outputs, and inputs for a number of firms, and data on quantities purchased and prices for some commodity, or commodities, in a number of retail establishments. These kind of data invariably involve observations on economic units of varying sizes. For example, with data on a number of firms, we might measure the size of the firm by the quantity of output it produces. The larger the firm, the more difficult it is likely to be to explain the variation in some outcome variable y_i by the variation in a set of explanatory variables. Larger firms and households are likely to be more diverse and flexible with respect to the way in which values for y_i are determined. This additional diversity is captured through an error term with a larger variance.

In this Chapter we give the fundamentals of sampling-theory and Bayesian estimation, and sampling theory hypothesis testing, for a linear model with heteroskedasticity. For sampling-theory estimation it is convenient to first describe estimation for a known error covariance matrix and to then extend it for an unknown error covariance matrix. No attempt is made to give specific details of developments beyond what we consider to be the fundamentals. However, references to such developments and how they build on the fundamentals are provided. Autoregressive conditional heteroskedasticity (ARCH) which is popular for modeling volatility in time-series is considered elsewhere in this volume and not discussed in this chapter.

2. Sampling Theory Inference with Known Covariance Matrix

Writing $y_i = x_i'\beta + e_i$ so that all N observations are included yields the familiar matrix expression

$$y = X\beta + e \quad (2.1)$$

where y and e are of dimension $(N \times 1)$ and X is of dimension $(N \times K)$. The assumption of heteroskedastic y can be written as

$$E[(y - X\beta)(y - X\beta)'] = E[ee'] = V = \sigma^2 \Lambda \quad (2.2)$$

where

$$\begin{aligned}
V &= \text{diagonal} (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2) \\
&= \sigma^2 \text{diagonal} (\lambda_1, \lambda_2, \dots, \lambda_N) \\
&= \sigma^2 \Lambda
\end{aligned} \tag{2.3}$$

In equation (2.3) a constant σ^2 has been factored out of V yielding a matrix Λ of ratios $\lambda_i = \sigma_i^2 / \sigma^2$. This factoring device is useful (i) when Λ is known, but σ^2 is not, and (ii) if a heteroskedastic specification has a constant component (σ^2 in this case) and a component that varies over observations. The constant that is factored out is arbitrary, and, in practice, is chosen for convenience.

The generalized least squares estimator for β which, from the Gauss-Markov Theorem is known to be the best linear unbiased estimator, is given by

$$\begin{aligned}
\hat{\beta} &= (X'V^{-1}X)^{-1} X'V^{-1}y = \left(\sum_{i=1}^N \frac{x_i x_i'}{\sigma_i^2} \right)^{-1} \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} \\
&= (X'\Lambda^{-1}X)^{-1} X'\Lambda^{-1}y = \left(\sum_{i=1}^N \frac{x_i x_i'}{\lambda_i} \right)^{-1} \sum_{i=1}^N \frac{x_i y_i}{\lambda_i}
\end{aligned} \tag{2.4}$$

The right-hand expressions in equations (2.4) emphasize the *weighted* nature of the generalized least squares estimator. Each observation (x_i and y_i) is weighted by the inverse standard deviation σ_i^{-1} , or a quantity proportional to it, λ_i^{-1} . Observations that are less reliable because they come from a distribution with a large variance are weighted less than more reliable observations where σ_i^2 is smaller. The mean and covariance matrix of the generalized least squares estimator are given by $E[\hat{\beta}] = \beta$ and $V_{\hat{\beta}}$, respectively, where

$$\begin{aligned}
V_{\hat{\beta}} &= (X'V^{-1}X)^{-1} = \left(\sum_{i=1}^N \frac{x_i x_i'}{\sigma_i^2} \right)^{-1} \\
&= \sigma^2 (X'\Lambda^{-1}X)^{-1} = \sigma^2 \left(\sum_{i=1}^N \frac{x_i x_i'}{\lambda_i} \right)^{-1}
\end{aligned} \tag{2.5}$$

Practical application of (2.4) and (2.5) requires knowledge of at least Λ . For inference purposes, an unbiased estimator for σ^2 can be found from

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})' \Lambda^{-1} (y - X\hat{\beta})}{N - K} \quad (2.6)$$

Although most applications proceed by refining the specification of Λ into one that contains a reduced number of parameters that is constant for changing sample size, there are some scenarios where knowledge of Λ is a reasonable assumption. To illustrate one such example, suppose that we are interested in an industry cost function that can be written as

$$y_{ij} = x'_{ij} \beta + e_{ij} \quad (2.7)$$

where the double subscript (i, j) refers to the j -th firm in the i -th industry. Suppose also that the e_{ij} are independent with $\text{var}(e_{ij}) = \sigma^2$ (a constant) and that there are n_i firms in the i -th industry. A model for data obtained by averaging over all firms in each industry is given by

$$\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n_i} \sum_{j=1}^{n_i} x'_{ij} \beta + \frac{1}{n_i} \sum_{j=1}^{n_i} e_{ij}$$

or

$$\bar{y}_i = \bar{x}'_i \beta + \bar{e}_i \quad (2.8)$$

The variance of the error term is

$$\begin{aligned} \text{var}(\bar{e}_i) &= \frac{1}{n_i^2} \sum_{j=1}^{n_i} \text{var}(e_{ij}) \\ &= \frac{1}{n_i^2} n_i \sigma^2 = \frac{\sigma^2}{n_i} \end{aligned}$$

That is, \bar{e}_i is heteroskedastic with its variance depending on the number of firms used to compute the average industry data. Providing this number is available, the matrix Λ is known with inverse given by

$$\Lambda^{-1} = \text{diagonal}(n_1, n_2, \dots, n_N)$$

The generalized least squares procedure can be applied. It recognizes that industry observations obtained by averaging a large number of firms are more reliable than those obtained by averaging a small number of firms.

Finite Sample Inference

To construct confidence intervals for the elements in β or to test hypotheses about the elements in β , one can assume the error vector e is normally distributed and proceed with finite sample inference procedures, or one can use large sample approximate inference procedures without the assumption of normally distributed errors. When the errors are normally distributed the following results hold:

$$\frac{(N - K)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(N-K)}^2 \quad (2.9)$$

$$R\hat{\beta} \sim N[R\beta, \sigma^2 R(X'\Lambda^{-1}X)^{-1}R'] \quad (2.10)$$

$$\frac{(R\hat{\beta} - R\beta)'[R(X'\Lambda^{-1}X)^{-1}R']^{-1}(R\hat{\beta} - R\beta)}{\sigma^2} \sim \chi_{(J)}^2 \quad (2.11)$$

$$F = \frac{(R\hat{\beta} - R\beta)'[R(X'\Lambda^{-1}X)^{-1}R']^{-1}(R\hat{\beta} - R\beta)/J}{\hat{\sigma}^2} \sim F_{[J, (N-K)]} \quad (2.12)$$

In the above expressions R is a $(J \times K)$ matrix of rank J whose elements depend on the quantity for which inference is sought. For example, if we are interested in a confidence interval for β_2 then $J = 1$ and $R = (0, 1, 0 \dots 0)$. If we are interested in testing an hypothesis about β_1 and $\beta_2 - 3\beta_3$, then

$$R = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -3 & \dots & 0 \end{bmatrix}$$

Equations (2.9) to (2.11) are intermediate steps that involve the unknown σ^2 ; they lead to the F -statistic in (2.12) upon which inferences are based. If $J = 1$, so that R is a row vector, then using F for inferences is equivalent to using the following t statistic

$$t = \frac{R\hat{\beta} - R\beta}{\hat{\sigma}\sqrt{R(X'\Lambda^{-1}X)^{-1}R'}} \sim t_{(N-K)} \quad (2.13)$$

Confidence intervals ($J = 1$) are obtained using (2.13); confidence regions ($J > 1$) are obtained using (2.14).

Hypothesis testing proceeds by replacing $R\beta$ by its hypothesized value, typically denoted by r in the literature. A common way of computing F under the null hypothesis assumption that $R\beta = r$ is through the equivalent expression

$$F = \frac{(\hat{e}'_R \Lambda^{-1} \hat{e}_R - \hat{e}' \Lambda^{-1} \hat{e})/J}{\hat{e}' \Lambda^{-1} \hat{e}/(N-K)} \sim F_{[J, (N-K)]} \quad (2.14)$$

where $\hat{e} = y - X\hat{\beta}$ and \hat{e}_R is the restricted least squares residual

$$\begin{aligned} \hat{e}_R &= y - X\hat{\beta}_R \\ &= y - X\left(\hat{\beta} + (X'\Lambda^{-1}X)^{-1}R'[R(X'\Lambda^{-1}X)^{-1}R']^{-1}(r - R\hat{\beta})\right) \end{aligned}$$

The residual sums of squares $\hat{e}'_R \Lambda^{-1} \hat{e}_R$ and $\hat{e}' \Lambda^{-1} \hat{e}$ are equal to the residual sums of squares from restricted and unrestricted weighted (transformed) least squares regressions, respectively.

Large Sample Inference

When the errors are not assumed to be normally distributed, large sample inference is based on the approximate result

$$\frac{(R\hat{\beta} - R\beta)'[R(X'\Lambda^{-1}X)R']^{-1}(R\hat{\beta} - R\beta)}{\hat{\sigma}^2} \sim \chi^2_{(J)} \quad (2.15)$$

For $J = 1$, the scalar equivalent to (2.14) is

$$\frac{R\hat{\beta} - R\beta}{\hat{\sigma}\sqrt{R(X'\Lambda^{-1}X)R'}} \sim N(0, 1) \quad (2.16)$$

Inference for Nonlinear Functions

The framework discussed so far is only relevant for making inferences about linear functions of the elements of β that can be written as $R\beta$, or put in the context of a null hypothesis written as $H_0 : R\beta = r$. If interest centers on nonlinear functions of

β , say in the form of a null hypothesis $H_0 : g(\beta) = 0$ where $g(\cdot)$ is a J -dimensional vector function, then inference can be based on the approximate result

$$\frac{g(\hat{\beta})' [G(X'\Lambda^{-1}X)^{-1}G']^{-1} g(\hat{\beta})}{\hat{\sigma}^2} \sim X_{(J)}^2 \quad (2.17)$$

where G is the $(J \times K)$ matrix of partial derivatives

$$G = \left. \frac{\partial g(\beta)}{\partial \beta'} \right|_{\hat{\beta}} \quad (2.18)$$

Three categories of tests frequently used in econometrics are the Wald, Lagrange multiplier and likelihood-ratio tests. In the context of the scenarios we have discussed so far (hypothesis tests about β in a model with covariance matrix $\sigma^2\Lambda$, with Λ known), all three testing principles lead to the results given above. The only difference is that, in a Lagrange multiplier test, the estimate for σ^2 is based on the restricted generalized least squares residuals \hat{e}_R , rather than the unrestricted residuals \hat{e} .

Further details on estimation and hypothesis testing for the case of a known error covariance matrix can be found in standard textbooks such as Judge et al. (1988, Ch. 8, 9), Greene (1997, Ch. 12) and Baltagi (1998, Ch. 5, 9). Of particular interest might be the consequences of using the ordinary least squares (OLS) estimator $b = (X'X)^{-1}X'y$ in the presence of heteroskedastic errors. It is well known that, under these circumstances, the OLS estimator is inefficient and that the estimated covariance matrix $\hat{\sigma}^2(X'X)^{-1}$ is a biased estimate of the true covariance matrix $\sigma^2(X'X)^{-1}X'\Lambda X(X'X)^{-1}$. Examples of inefficiencies and bias are given in most textbook treatments. For a more general discussion of OLS properties in linear models where $E(ee') \neq \sigma^2I$, see Fiebig et al. (1992).

3. Sampling Theory Estimation and Inference With Unknown Covariance Matrix

Consider again the linear model $y = X\beta + e$ with error-covariance matrix $V = \sigma^2\Lambda$. In this section we relax the assumption that Λ is known. As we saw in the

previous section, there are some circumstances where such an assumption is reasonable. However, there are also many when it is not. For example, in a household expenditure function, we may be willing to assume the variance of expenditure depends on total expenditure and the demographic composition of the household, but not willing to specify the values of parameters that describe the dependence. Thus, we could write, for example,

$$\sigma_i^2 = \theta_0 + \theta_1 z_{1i} + \theta_2 z_{2i} \quad (3.1)$$

where z_{1i} and z_{2i} are total expenditure and demographic composition, respectively, and $(\theta_0, \theta_1, \theta_2)$ are unknown parameters. If $(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$ are not known, then some kind of reparameterization such as that in (3.1) is necessary to reduce the number of parameters to a manageable number that does not increase with sample size. We will work in terms of the general notation

$$\sigma_i^2 = \sigma^2 h_i(\alpha) = \sigma^2 h(z_i' \alpha) \quad (3.2)$$

where α is an $(S \times 1)$ vector of unknown parameters and $h_i(\cdot)$ is a function of those parameters and an $(S \times 1)$ vector z_i which could be identical to or different from x_i . To write (3.1) in terms of the general notation in (3.2), we re-express it as

$$\begin{aligned} \sigma_i^2 &= \theta_0 \left(1 + \frac{\theta_1}{\theta_0} z_{1i} + \frac{\theta_2}{\theta_0} z_{2i} \right) \\ &= \sigma^2 (1 + \alpha_1 z_{1i} + \alpha_2 z_{2i}) \\ &= \sigma^2 h_i(\alpha) \end{aligned} \quad (3.3)$$

In this example, and others which we consider, $h_i(0) = 1$, implying that $\alpha = 0$ describes a model with homoskedastic errors.

Several alternative specifications of $h_i(\alpha)$ have been suggested in the literature. See Judge et al. (1985, p. 422) for a review. One of these is that given in (3.3), namely

$$h_i(\alpha) = 1 + \alpha_1 z_{1i} + \dots + \alpha_S z_{Si} \quad (3.4)$$

This model has been considered by Goldfeld and Quandt (1972) and Amemiya (1977), and, in the context of a random coefficient model, by Hildreth and Houck (1968), Griffiths (1972), Froehlich (1973), Raj (1975) and Dent and Hildreth (1977).

A specification in terms of the Hadamard product has been provided by Neudecker et al. (1995). Note that, if (z_{1i}, \dots, z_{Si}) are non-overlapping dummy variables, then the specification in (3.4) describes a partition of the sample into $(S + 1)$ sub-samples, each one with a different error variance. Such a model could be relevant if parts of the sample came from different geographical regions or there exists some other way of naturally creating sample separations. Estimation within this framework has been considered by Oberhofer and Kmenta (1974), Taylor (1977, 1978), Magnus (1978), Swamy and Mehta (1979), Griffiths and Judge (1992), and Hooper (1993).

One potential difficulty with the specification in (3.4) is that the requirement $h_i(\alpha) > 0$ can mean that restrictions must be placed on α to ensure that negative variances are not possible. Two possible specifications which avoid this problem are

$$h_i(\alpha) = (1 + \alpha_1 z_{1i} + \dots + \alpha_S z_{Si})^2 \quad (3.5)$$

and

$$h_i(\alpha) = \exp(\alpha_1 z_{1i} + \dots + \alpha_S z_{Si}) \quad (3.6)$$

The specification in (3.5) has received attention from Rutenmiller and Bowers (1968) and Jobson and Fuller (1980). The specification in (3.6) was introduced by Harvey (1976) under the heading “multiplicative heteroskedasticity” and has been applied and extended by Just and Pope (1978), Griffiths and Anderson (1982), and Hill et al. (1997, 1999).

A class of models which has been popular, but which does not fit within the framework of equation (3.2), is that where the location parameter vector β also appears within the variance function. Authors who have considered this class of models under varying degrees of generality include Theil (1971), Amemiya (1973), Jobson and Fuller (1980), Battese and Bonyhady (1981), Carroll and Ruppert (1982), Davidian and Carroll (1987) and Welsh et al. (1994).

Maximum Likelihood Estimation

Two-step estimation of heteroskedastic error models was popular prior to the development of modern software. These techniques use the residuals from least squares estimation to estimate α , and then use the estimate of α in a generalized least

squares estimator. See Judge et al. (1985, p. 431-441) for details. Recently, however, it is more common to assume normally distributed errors and proceed with maximum likelihood estimation. Working in this direction, the log-likelihood function can be written as

$$\begin{aligned} L &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{i=1}^N \ln(h_i(\alpha)) - \frac{1}{2\sigma^2} \sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{h_i(\alpha)} \\ &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \ln|\Lambda| - \frac{1}{2\sigma^2} (y - X\beta)' \Lambda^{-1} (y - X\beta) \end{aligned} \quad (3.7)$$

Differentiating this function with respect to σ^2 and β and setting these derivatives equal to zero gives the results

$$\hat{\beta}(\alpha) = (X' \Lambda^{-1} X)^{-1} X' \Lambda^{-1} y \quad (3.8)$$

$$\hat{\sigma}^2(\alpha) = \frac{1}{N} (y - X\hat{\beta}(\alpha))' \Lambda^{-1} (y - X\hat{\beta}(\alpha)) \quad (3.9)$$

Since these estimators are conditional on knowing α , they resemble those provided in the previous section, the only difference being the divisor N , instead of $(N-K)$, in equation (3.9).

Differentiating L with respect to α yields

$$\frac{\partial L}{\partial \alpha} = -\frac{1}{2} \sum_{i=1}^N \frac{1}{h_i(\alpha)} \frac{\partial h_i}{\partial \alpha} + \frac{1}{2\sigma^2} \sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{[h_i(\alpha)]^2} \frac{\partial h_i}{\partial \alpha} \quad (3.10)$$

Setting this derivative equal to zero does not yield a convenient solution for α , although it does simplify for specific definitions of $h_i(\alpha)$. For example, for the specification in equation (3.6), written in matrix algebra notation as $h_i(\alpha) = \exp\{z_i'\alpha\}$, equating (3.10) to zero yields

$$\sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{\exp(z_i'\alpha)} z_i = \sigma^2 \sum_{i=1}^N z_i \quad (3.11)$$

Substituting (3.8) and (3.9) into (3.7) yields the concentrated log-likelihood function

$$L^*(\alpha) = \text{constant} - \frac{N}{2} \ln[(y - X\hat{\beta}(\alpha))' \Lambda^{-1} (y - X\hat{\beta}(\alpha))] - \frac{1}{2} \ln|\Lambda| \quad (3.12)$$

Thus, maximum likelihood estimation can proceed by numerically finding the value $\hat{\alpha}$ that maximizes L^* , and then substituting that value into equations (3.8) and (3.9).

The information matrix is given by

$$I(\beta, \alpha, \sigma^2) = -E \begin{bmatrix} \frac{\partial^2 L}{\partial \beta \partial \beta'} & \frac{\partial^2 L}{\partial \beta \partial \alpha'} & \frac{\partial^2 L}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \alpha \partial \beta'} & \frac{\partial^2 L}{\partial \alpha \partial \alpha'} & \frac{\partial^2 L}{\partial \alpha \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \beta'} & \frac{\partial^2 L}{\partial \sigma^2 \partial \alpha'} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{X' \Lambda^{-1} X}{\sigma^2} & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^N \frac{1}{[h_i(\alpha)]^2} \frac{\partial h_i}{\partial \alpha} \frac{\partial h_i}{\partial \alpha'} & \frac{1}{2\sigma^2} \sum_{i=1}^N \frac{1}{h_i(\alpha)} \frac{\partial h_i}{\partial \alpha} \\ 0 & \frac{1}{2\sigma^2} \sum_{i=1}^N \frac{1}{h_i(\alpha)} \frac{\partial h_i}{\partial \alpha'} & \frac{N}{2\sigma^4} \end{bmatrix} \quad (3.13)$$

The inverse of this matrix is the asymptotic covariance matrix for the maximum likelihood estimators of β , α and σ^2 . Its block-diagonal nature means the asymptotic covariance matrix for the maximum likelihood estimator $\hat{\beta}$ is given by the familiar expression

$$V_{\hat{\beta}} = \sigma^2 (X' \Lambda^{-1} X)^{-1}$$

Equation (3.13) can be simplified considerably once $h_i(\alpha)$ is specified explicitly. For example, for the case where $h_i(\alpha) = \exp\{z_i' \alpha\}$,

$$I(\beta, \alpha, \sigma^2) = \begin{bmatrix} \frac{X' \Lambda^{-1} X}{\sigma^2} & 0 & 0 \\ 0 & \frac{1}{2} Z' Z & \frac{1}{2\sigma^2} \sum z_i \\ 0 & \frac{1}{2\sigma^2} \sum z_i' & \frac{N}{2\sigma^4} \end{bmatrix} \quad (3.14)$$

Testing Hypotheses about β

To test hypotheses about β , the large sample results given in equations (2.14) to (2.16) can be used. The only differences are that $\hat{\beta}$ and $\hat{\sigma}^2$ become the maximum likelihood estimators and the previously known matrix Λ is replaced by its maximum likelihood estimate.

Using the maximum likelihood estimator for β and its corresponding covariance matrix to test hypotheses about β requires knowledge of the function $h_i(\alpha)$. That is, the form of the heteroskedasticity is required. As an alternative, tests can be based on the ordinary least squares estimator for β and an estimate of its covariance matrix. Specifically, the least squares estimator $b = (X'X)^{-1} X'y$ has covariance matrix

$$V_b = (X'X)^{-1} X' \hat{V} X (X'X)^{-1}$$

which White (1980) has shown can be consistently estimated by

$$\hat{V}_b = (X'X)^{-1} X' \hat{V} X (X'X)^{-1}$$

where \hat{V} is a diagonal matrix containing the squares of the ordinary least squares residuals. Thus, the result

$$(Rb - R\beta)' (R\hat{V}_b R')^{-1} (Rb - R\beta) \stackrel{a}{\sim} \chi_{(j)}^2$$

can be used to make approximate inferences about β . The finite sample properties of such inferences have been questioned however and ways for improving the test have been investigated. See, for example, MacKinnon and White (1985), Hsieh (1983), Chesher and Jewitt (1987), Chesher (1989), Andrews (1991b), Chesher and Austin (1991), Keener et al. (1991), and Andrew and Monahan (1992).

Testing for Heteroskedasticity

Assuming that $h_i(\alpha)$ is such that $h_i(0) = 1$, tests for heteroskedasticity can be formulated in terms of the hypotheses

$$H_0 : \alpha = 0$$

$$H_1 : \alpha \neq 0$$

We will describe the likelihood ratio, Wald and Lagrange multiplier test statistics for these hypotheses, and then refer to other tests and evaluations that have appeared in the literature.

Using equation (3.12), the *likelihood ratio test* statistic is given by

$$\begin{aligned}\gamma_{LR} &= 2[L(\hat{\alpha}) - L(0)] \\ &= N \ln \left(\frac{\hat{e}'_0 \hat{e}_0}{\hat{e}' \hat{\Lambda}^{-1} \hat{e}} \right) - \sum_{i=1}^N \ln[h_i(\hat{\alpha})]\end{aligned}\quad (3.15)$$

where $\hat{e}_0 = y - Xb$ are the OLS residuals and $\hat{e} = y - X\hat{\beta}(\hat{\alpha})$ are the maximum likelihood residuals. When the null hypothesis of homoskedasticity holds, γ_{LR} has an approximate $\chi^2_{(s)}$ distribution.

The *Wald test* statistic is given by

$$\gamma_W = \hat{\alpha}' \hat{V}_{\hat{\alpha}}^{-1} \hat{\alpha} \quad (3.16)$$

where, applying partitioned-inverse results to equation (3.13), it can be shown that

$$V_{\hat{\alpha}}^{-1} = \frac{1}{2} \sum_{i=1}^N \frac{1}{[h_i(\alpha)]^2} \frac{\partial h_i}{\partial \alpha} \frac{\partial h_i}{\partial \alpha'} - \frac{1}{2N} \left(\sum_{i=1}^N \frac{1}{h_i(\alpha)} \frac{\partial h_i}{\partial \alpha} \right) \left(\sum_{i=1}^N \frac{1}{h_i(\alpha)} \frac{\partial h_i}{\partial \alpha'} \right) \quad (3.17)$$

The statistic γ_W has an approximate $\chi^2_{(s)}$ distribution when $\alpha = 0$.

The *Lagrange multiplier test* statistic is given by

$$\gamma_{LM} = s'_0 I_0^{-1}(\alpha, \sigma^2) s_0 \quad (3.18)$$

where

$$s_0 = \begin{pmatrix} \partial L / \partial \alpha \\ \partial L / \partial \sigma^2 \end{pmatrix} \text{ evaluated at } \alpha = 0, \sigma^2 = \hat{\sigma}^2(0) \text{ and } \beta = b.$$

$I_0^{-1}(\alpha, \sigma^2)$ is the inverse of the bottom-right block in equation (3.13), evaluated at $\alpha = 0$ and $\sigma^2 = \hat{\sigma}^2(0)$.

Recognizing that $\partial h_i / \partial \alpha$ evaluated at $\alpha = 0$ is equal to z_i , equation (3.10) can be used to yield

$$s_0 = \begin{pmatrix} \frac{1}{2} \sum_{i=1}^N z_i \left(\frac{\hat{e}_{0i}^2}{\hat{\sigma}_0^2} - 1 \right) \\ 0 \end{pmatrix} \quad (3.19)$$

where $\hat{\sigma}_0^2 = \sum_{i=1}^N \hat{e}_{0i}^2 / N = \hat{\sigma}^2(0)$.

Then, utilizing (3.17) evaluated at $\alpha = 0$, the Lagrange multiplier statistic becomes

$$\gamma_{LM} = \frac{\sum_{i=1}^N z_i' (\hat{e}_{0i}^2 - \hat{\sigma}_0^2) \left(\sum_{i=1}^N (z_i - \bar{z})(z_i - \bar{z})' \right)^{-1} \sum_{i=1}^N z_i' (\hat{e}_{0i}^2 - \hat{\sigma}_0^2)}{2\hat{\sigma}_0^4}$$

This statistic is conveniently calculated as one-half of the regression sum-of-squares of $\hat{e}_{0i}^2 / \hat{\sigma}_0^2$ on z_i and a constant term. It has an approximate $\chi_{(s)}^2$ distribution under $H_0 : \alpha = 0$. The Lagrange multiplier test statistic was derived by Breusch and Pagan (1978) and Godfrey (1978). To make it more robust to departures from normality, replacement of the denominator $2\hat{\sigma}_0^4$ by $N^{-1} \sum_{i=1}^N (\hat{e}_{0i}^2 - \hat{\sigma}_0^2)^2$ has been suggested (Koenker 1981; Koenker and Bassett 1982).

Many more tests for heteroskedasticity have been suggested in the literature. See Pagan and Pak (1993) for a review and for details on how the various tests can be classified as conditional moment tests. One popular test that we have not yet mentioned is the Goldfeld-Quandt (1965) test which uses the error variances from two separate least squares regressions to construct a finite sample F -statistic. Other classes of tests have been described by Szroeter (1978) and Farebrother (1987). Lee (1992) suggests a test where the mean function is estimated nonparametrically and hence does not have to be precisely specified. Orme (1992) describes tests in the context of censored and truncated regression models. Also, tests for heteroskedasticity in these and other nonlinear models, such as discrete choice models and count data models, are reviewed by Pagan and Pak (1993). Numerous Monte Carlo studies have compared the finite sample size and power of existing and new test statistics. Typically, authors uncover problems with existing test statistics such as poor finite sample size or power, or lack of robustness to misspecification and nonnormality, and suggest alternatives to correct for such problems. A study by Godfrey and Orme (1999) suggests that bootstrapping leads to favorable outcomes. Other examples of

Monte Carlo studies that have appeared are Harvey and Phillips (1974), Ali and Giaccotto (1984), Evans and King (1985, 1988), Griffiths and Surekha (1986), Sharma and Giaccotto (1991), Evans (1992), Griffiths and Judge (1992) and Godfrey (1996). See Farebrother (1987) for some insightful comments on the results of Griffiths and Surekha (1986).

Estimation with Unknown Form of Heteroskedasticity

The work of White (1980) on testing for heteroskedasticity and testing hypotheses about β without specifying the precise form of the heteroskedasticity motivated others to seek *estimators* for β that did not require specification of the form of heteroskedasticity. Attempts have been made to specify estimators which are more efficient than OLS, while at the same time recognizing that the efficiency of GLS may not be achievable (Cragg 1983, 1992; Amemiya 1983; Balestra 1983; Koenker et al. 1993). Carroll (1982) and Robinson (1987) develop adaptive estimators that assume no particular form of heteroskedasticity but nevertheless have the same asymptotic distribution as the generalized least squares estimator that uses a correct parametric specification. These adaptive estimators have been evaluated in terms of a second-order approximation by Linton (1996) and extended to time-series models by Hidalgo (1992) and to panel data by Li and Stengos (1994). Simultaneous confidence bands in the presence of heteroskedasticity of unknown forms have been investigated by Faraway and Sun (1995). Szroeter (1994) suggests weighted least squares estimators that have better finite sample efficiency than OLS when the observations can be ordered according to increasing variances but no other information is available.

Other Extensions

Rilestone (1991) has compared the relative efficiency of semiparametric and parametric estimators of β under different types of heteroskedasticity, whereas Surekha and Griffiths (1984) compare the relative efficiency of some Bayesian and sampling theory estimators using a similar Monte Carlo setup. Donald (1995) examines heteroskedasticity in sample selection models, and provides access to that literature. Heteroskedasticity in the context of seemingly unrelated regressions has been studied by Bewley and Theil (1987) and Mandy and Martins-Filho (1993). More

general error structures for heteroskedasticity are considered by Spanos (1994). Andrews (1991a) examines a number of model-selection procedures in the presence of heteroskedasticity.

4. Bayesian Inference

With Bayesian inference post-sample information about unknown parameters is summarized via posterior probability density functions (pdfs) on the parameters of interest. In the heteroskedastic model that we have been discussing, namely, $y_i = x_i'\beta + e_i$, with $\text{var}(e_i) = \sigma^2 h_i(\alpha)$, the parameters of interest are β , α and σ , with particular interest usually centering on β , although there are applications where σ and α also have a bearing on decision making (e.g., Griffiths 1986). The starting point for Bayesian inference is the specification of prior pdfs for β , σ and α . Since noninformative prior pdfs carry with them the advantage of objective reporting of results, we adopt the conventional ones for β and σ (see Zellner, 1971)

$$f(\beta, \sigma) = f(\beta)f(\sigma) \propto \text{constant} \cdot \frac{1}{\sigma} \quad (4.1)$$

The choice of prior for α is likely to depend on the function $h(\cdot)$. Possible choices are a uniform prior or a prior based on the information matrix. See Zellner (1971, p. 47) for details on the latter. Leaving the precise nature of the prior for α unspecified, the joint prior pdf for all unknown parameters can be written as

$$f(\beta, \sigma, \alpha) = f(\beta, \sigma)f(\alpha) \propto \frac{f(\alpha)}{\sigma} \quad (4.2)$$

Assuming normally distributed observations, the likelihood function can be written as

$$f(y|\beta, \sigma, \alpha) \propto \frac{1}{\sigma^N} |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y - X\beta)' \Lambda^{-1} (y - X\beta)\right\} \quad (4.3)$$

The joint posterior pdf for (β, σ, α) is

$$\begin{aligned} f(\beta, \sigma, \alpha|y) &\propto f(y|\beta, \sigma, \alpha)f(\beta, \sigma, \alpha) \\ &\propto \frac{f(\alpha)}{\sigma^{N+1}} |\Lambda|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y - X\beta)' \Lambda^{-1} (y - X\beta)\right\} \end{aligned} \quad (4.4)$$

Once this joint posterior pdf for all parameters has been obtained, the major task is to derive marginal posterior pdfs for each single parameter. The information in a marginal posterior pdf can then be represented in a diagram or summarized via the moments of the pdf. Where possible, marginal posterior pdfs are obtained by integrating out the remaining parameters. Where analytical integration is not possible, numerical methods are used to estimate the marginal posterior pdfs. There are a variety of ways that one proceed with respect to equation (4.4). The steps for one that is likely to work well are:

1. Integrate σ out to obtain the joint posterior pdf $f(\beta, \alpha|y)$.
2. Integrate β out of the result in step 1 to obtain the posterior pdf $f(\alpha|y)$.
3. Use a Metropolis algorithm to draw observations from the density $f(\alpha|y)$.
4. Construct the conditional posterior pdf $f(\beta|\alpha, y)$ from the joint posterior pdf that was obtained in step 1; note the conditional mean $E[\beta|\alpha, y]$ and conditional variance $\text{var}[\beta|\alpha, y]$.
5. From step 4, note the conditional posterior pdf and corresponding moments for each element, say β_k , in the vector β .
6. Find estimates of the marginal posterior pdf for β_k , ($k = 1, 2, \dots, K$), and its moments, by averaging the conditional quantities given in step 5, over the conditioning values of α drawn in step 3.

We will consider each of these steps in turn.

Step 1

The joint posterior pdf for (β, α) is given by

$$\begin{aligned}
 f(\beta, \alpha|y) &= \int f(\beta, \alpha, \sigma|y) d\sigma \\
 &\propto f(\alpha) |\Lambda|^{-1/2} \left[N\hat{\sigma}^2(\alpha) + (\beta - \hat{\beta}(\alpha))' X' \Lambda^{-1} X (\beta - \hat{\beta}(\alpha)) \right]^{-N/2} \quad (4.5)
 \end{aligned}$$

where $\hat{\sigma}^2(\alpha)$ and (\cdot) are defined in equations (3.8) and (3.9). The pdf in (4.5) is utilized directly; it provides an intermediate step for obtaining $(\cdot | \cdot)$ $f(\beta|\alpha, y)$.

Step 2

The marginal posterior pdf for the parameters in the variance function is given by

$$\begin{aligned} f(\alpha|y) &= \int f(\beta, \alpha|y) d\beta \\ &\propto f(\alpha) |\Lambda|^{-1/2} [\hat{\sigma}(\alpha)]^{-(N-K)} |X'\Lambda^{-1}X|^{-1/2} \end{aligned} \quad (4.6)$$

Step 3

The pdf in equation (4.6) is not of a recognizable form, even when imaginative choices for the prior $f(\alpha)$ are made. Thus, it is not possible to perform further analytical integration to isolate marginal posterior pdfs for single elements such as α_s . Instead, a numerical procedure, the Metropolis algorithm, can be used to indirectly draw observations from the pdf $f(\alpha|y)$. Once such draws are obtained, they can be used to form histograms as estimates of the posterior pdfs for single elements in α . As we shall see, the draws are also useful for obtaining the posterior pdfs for the β_k .

The random-walk Metropolis algorithm which we describe below in the context of the heteroskedastic model is one of many algorithms which come under the general heading of Markov Chain Monte Carlo (MCMC). A recent explosion of research in MCMC has made Bayesian inference more practical for models that were previously plagued by intractable integrals. See, for example, Chib and Greenberg (1995, 1996), Albert and Chib (1996) and Geweke (1999).

The first step towards using a convenient random-walk Metropolis algorithm is to define a suitable ‘‘candidate generating function’’. Assuming that the prior $f(\alpha)$ is relatively noninformative, and not in conflict with the sample information, the

maximum likelihood estimate $\hat{\alpha}$ provides a suitable starting value $\alpha_{(0)}$ for the algorithm; and the maximum likelihood covariance matrix V_{α} provides the basis for a suitable covariance matrix for the random-walk generator function. The steps for drawing the $(m+1)$ th observation $\alpha_{(m+1)}$ are as follows:

1. Draw $\alpha^* = \alpha_{(m)} + \varepsilon$ where $\varepsilon \sim N(0, cV_{\alpha})$ and c is scalar set so that α^* is accepted approximately 50% of the time.
2. Compute

$$r = \frac{f(\alpha^*|y)}{f(\alpha_{(m)}|y)}$$

Note that this ratio can be computed without knowledge of the normalising constant for $f(\alpha|y)$.

3. Draw a value u for a uniform random variable on the interval $(0, 1)$.
4. If $u \leq r$, set $\alpha_{(m+1)} = \alpha^*$.
If $u > r$, set $\alpha_{(m+1)} = \alpha_{(m)}$.
5. Return to step 1, with m set to $m+1$.

By following these steps, one explores the posterior pdf for α , generating larger numbers of observations in regions of high posterior probability and smaller numbers of observations in regions of low posterior probability. Markov Chain Monte Carlo theory suggests that, after sufficient observations have been drawn, the remaining observations are drawn from the pdf $f(\alpha|y)$. Thus, by drawing a large number of values, and discarding early ones, we obtain draws from the required pdf.

Step 4

The conditional posterior pdf $f(\beta|\alpha, y)$ is obtained from the joint pdf $f(\beta, \alpha|y)$ by simply treating α as a constant in equation (4.5). However, for later use we also need to include any part of the normalising constant that depends on α . Recognizing that,

when viewed only as a function of β , equation (4.5) is in the form of a multivariate student- t pdf (Judge et al. 1988, p. 312), we have

$$f(\beta|\alpha, y) \propto |X'\Lambda^{-1}X|^{1/2} [\hat{\sigma}(\alpha)]^{N-K} [N\hat{\sigma}^2(\alpha) + (\beta - \hat{\beta}(\alpha))' X'\Lambda^{-1}X(\beta - \hat{\beta}(\alpha))]^{-N/2} \quad (4.7)$$

This pdf has

$$\text{mean} = E(\beta|\alpha, y) = \hat{\beta}(\alpha) = (X'\Lambda^{-1}X)^{-1} X'\Lambda^{-1}y \quad (4.8)$$

$$\text{covariance matrix} = \left(\frac{N}{N-K-2} \right) \hat{\sigma}^2(\alpha) (X'\Lambda^{-1}X)^{-1} \quad (4.9)$$

$$\text{degrees of freedom} = N - K$$

Step 5

Let $a^{kk}(\alpha)$ be the k -th diagonal element of $(X'\Lambda^{-1}X)^{-1}$, and $\hat{\beta}_k(\alpha)$ be the k -th element of $\hat{\beta}(\alpha)$. The conditional marginal posterior pdf for β_k given α is the univariate- t pdf

$$f(\beta_k|\alpha, y) = k^* [\hat{\sigma}(\alpha)]^{N-K} [a^{kk}(\alpha)]^{(N-K)/2} [N\hat{\sigma}^2(\alpha)a^{kk}(\alpha) + (\beta_k - \hat{\beta}_k(\alpha))^2]^{-(N-K+1)/2} \quad (4.10)$$

where k^* is a normalizing constant independent of α . This pdf has

$$\text{mean} = E(\beta_k|\alpha, y) = \hat{\beta}_k(\alpha) \quad (4.11)$$

$$\text{variance} = \left(\frac{N}{N-K+2} \right) \hat{\sigma}^2(\alpha) a^{kk}(\alpha) \quad (4.12)$$

$$\text{degrees of freedom} = N - K$$

Equations (4.8) and (4.11) provide Bayesian quadratic-loss point estimates for β given α . Note that they are identical to the generalized least squares estimator for known α . It is the unknown α case where sampling theory and Bayesian inference results for point estimation of β diverge. The sampling theory point estimate in this case is $\hat{\beta}(\hat{\alpha})$. The Bayesian point estimate is the mean of the marginal posterior pdf $f(\beta|y)$. It can

be viewed as a weighted average of the $\hat{\beta}(\alpha)$ over all α with $f(\alpha|y)$ used as the weighting pdf. The mechanics of this procedure are described in the next step.

Step 6

An estimate of the marginal posterior pdf $f(\beta_k|y)$ is given by

$$\begin{aligned}\hat{f}(\beta_k|y) &= \frac{1}{M} \sum_{m=1}^M f(\beta_k|\alpha_{(m)}, y) \\ &= \frac{k^*}{M} \sum_{m=1}^M \left([\hat{\sigma}(\alpha_{(m)})]^{N-K} [a^{kk}(\alpha_{(m)})]^{(N-K)/2} \right. \\ &\quad \left. \times [N\hat{\sigma}^2(\alpha_{(m)})a^{kk}(\alpha_{(m)}) + (\beta_k - \beta_k(\alpha_{(m)}))^2]^{-(N-K+1)/2} \right)\end{aligned}\quad (4.13)$$

where $\alpha_{(1)}, \alpha_{(2)}, \dots, \alpha_{(M)}$ are the draws from $f(\alpha|y)$ that were obtained in step 3. To graph $\hat{f}(\beta_k|y)$ a grid of values of β_k is chosen and the average in equation (4.13) is calculated for each value of β_k in the grid. The mean and variance of the marginal posterior pdf $f(\beta_k|y)$ can be estimated in a similar way. The mean is given by the average of the conditional means

$$\bar{\beta} = \hat{E}(\beta|y) = \frac{1}{M} \sum_{m=1}^M \hat{\beta}_k(\alpha_{(m)}) \quad (4.14)$$

The variance is given by the average of the conditional variances plus the variance of the conditional means. That is,

$$\begin{aligned}\hat{\text{var}}(\beta|y) &= \left(\frac{N}{N-K+2} \right) \frac{1}{M} \sum_{m=1}^M \hat{\sigma}^2(\alpha_{(m)}) a^{kk}(\alpha_{(m)}) \\ &\quad + \frac{1}{M-1} \sum_{m=1}^M (\hat{\beta}_k(\alpha_{(m)}) - \bar{\beta})^2\end{aligned}\quad (4.15)$$

Presenting information about parameters in terms of posterior pdfs rather than point estimates provides a natural way of representing uncertainty. In the process just described, the marginal posterior pdfs also provide a proper reflection of finite sample uncertainty. Maximum likelihood estimates (or posterior pdfs conditional on $\hat{\alpha}$) ignore the additional uncertainty created by not knowing α .

There are, of course, other ways of approaching Bayesian inference in heteroskedastic models. The approach will depend on specification of the model and prior pdf, and on the solution to the problem of intractable integrals. Gibbs sampling is another MCMC technique that is often useful; and importance sampling could be used to obtain draws from $f(\alpha|y)$. However, the approach we have described is useful for a wide range of problems, with specific cases defined by specification of $h_i(\alpha)$ and $f(\alpha)$. Other studies which utilize Bayesian inference in heteroskedastic error models include Griffiths, Drynan and Prakash (1979), Surekha and Griffiths (1984, 1985) and Boscardin and Gelman (1996).

5. Concluding Remarks

Recent sampling-theory research on heteroskedastic models seems to be concentrated on methods for estimation and hypothesis testing that do not require specification of a particular parametric form of heteroskedasticity. They are motivated by our inability to be certain about the most appropriate variance specification. However, methodology suggested along these lines is generally asymptotic and may not perform well in finite samples. What is likely to be important, and what seems to have been neglected, is whether the types of inferences we make in practice are very sensitive to the assumed form of heteroskedasticity. If they are not, then efforts to develop alternative methods, that do not require an explicit variance function, may be misplaced.

Bayesian estimation has several advantages. Results are presented in terms of intuitively meaningful posterior pdfs. Marginal posterior pdfs reflect all the parameter uncertainty in a model and do not condition on point estimates of nuisance parameters. Predictive pdfs for future values can also be constructed without conditioning on point estimates (Boscardin and Gelman 1996). The advent of MCMC techniques means that many more practical applications of Bayesian inference to heteroskedastic models is now possible.

REFERENCES

- Albert, J. and S. Chib (1996), "Computation in Bayesian Econometrics: An Introduction to Markov Chain Monte Carlo," in R.C. Hill, editor, *Advances in Econometrics, Volume 11A: Bayesian Computational Methods and Applications*, Greenwich: JAI Press, 3-24.
- Ali, M.M. and C. Giacotto (1984), "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, 26, 355-373.
- Amemiya, T. (1973), "Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation," *Journal of the American Statistical Association*, 68, 928-934.
- Amemiya, T. (1977), "A Note on a Heteroscedastic Model," *Journal of Econometrics*, 6, 365-370; and "Corrigenda," *Journal of Econometrics*, 8, 275.
- Amemiya, T. (1983), "Partially Generalized Least Squares and Two-Stage Least Squares Estimators," *Journal of Econometrics*, 23, 275-283.
- Andrews, D.W.K. (1991a), "Asymptotic Optimality of Generalized C_L , Cross-Validation, and Generalized Cross-Validation in Regression with Heteroskedastic Errors," *Journal of Econometrics*, 47, 359-378.
- Andrews, D.W.K. (1991b), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- Andrews, D.W.K. and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica*, 60, 953-966.
- Balestra, P. (1983), "A Note on Amemiya's Partially Generalized Least Squares," *Journal of Econometrics*, 23, 285-290.
- Baltagi, B.H. (1998), *Econometrics*, New York: Springer-Verlag.
- Battese, G.E. and B.P. Bonyhady (1981), "Estimation of Household Expenditure Functions: An Application of a Class of Heteroscedastic Regression Models," *The Economic Record*, 57, 80-85.
- Bewley, R. and H. Theil (1987), "Monte Carlo Testing for Heteroskedasticity in Equation Systems," in T.B. Fomby and G.F. Rhodes, editors, *Advances in Econometrics, Volume 6*, Greenwich: JAI Press, 1-15.
- Boscardin, W.J. and A. Gelman (1996), "Bayesian Computation for Parametric Models of Heteroscedasticity in the Linear Model", in R.C. Hill, editor, *Advances in Econometrics Volume 11A: Bayesian Computational Methods and Applications*, Greenwich: JAI Press.
- Breusch, T.S. and A.R. Pagan (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

- Carroll, R.J. (1982), "Adapting for Heteroscedasticity in Linear Models", *Annals of Statistics*, 10, 1224-1233.
- Carroll, R.J. and D. Ruppert (1982), "A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model," *Journal of the American Statistical Association*, 77, 878-882.
- Chesher, A. (1989), "Hajek Inequalities, Measures of Leverage and the Size of Heteroskedasticity Robust Wald Tests," *Econometrica*, 57, 871-977.
- Chesher, A. and G. Austin (1991), "The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics," *Journal of Econometrics*, 47, 153-174.
- Chesher, A. and I. Jewitt (1987), "The Bias of a Heteroscedasticity Consistent Covariance Matrix Estimator," *Econometrica*, 55, 1217-1222.
- Chib, S. and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327-335.
- Chib, S. and E. Greenberg (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409-431.
- Cragg, J.G. (1983), "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica*, 51, 751-764.
- Cragg, J.G. (1992), "Quasi-Aitken Estimation for Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 54, 179-202.
- Davidian, M. and R.J. Carroll (1987), "Variance Function Estimation", *Journal of the American Statistical Association*, 82, 1079-1091.
- Delgado, M.A. (1992), "Semiparametric Generalized Least Squares in the Multivariate Nonlinear Regression Model," *Econometric Theory*, 8, 203-222.
- Dent, W. and C. Hildreth (1977), "Maximum Likelihood Estimation in Random Coefficient Models," *Journal of the American Statistical Association*, 72, 69-72.
- Donald, S.G. (1995), "Two-Step Estimation of Heteroskedastic Sample Selection Models," *Journal of Econometrics*, 65, 347-380.
- Evans, M. (1992), "Robustness of Size of Tests of Autocorrelation and Heteroskedasticity to Nonnormality," *Journal of Econometrics*, 51, 7-24.
- Evans, M.A. and M.L. King (1985), "A Point Optimal Test for Heteroscedastic Disturbances," *Journal of Econometrics*, 27, 163-178.
- Evans, M.A. and M.L. King (1988), "A Further Class of Tests for Heteroscedasticity," *Journal of Econometrics*, 37, 265-276.
- Faraway, J.J. and J. Sun (1995), "Simultaneous Confidence Bands for Linear Regression with Heteroskedastic Errors", *Journal of the American Statistical Association*, 90, 1094-1098.

- Farebrother, R.W. (1987), "The Statistical Foundations of a Class of Parametric Tests for Heteroskedasticity," *Journal of Econometrics*, 36, 359-368.
- Fiebig, D.G., M. McAleer and R. Bartels (1992), "Properties of Ordinary Least Squares Estimators in Regression Models with Nonspherical Disturbances", *Journal of Econometrics*, 54, 321-334.
- Froehlich, B.R. (1973), "Some Estimators for a Random Coefficient Regression Model," *Journal of the American Statistical Association*, 68, 329-334.
- Geweke, J. (1999), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, 18, 1-74.
- Godfrey, L.G. (1978), "Testing for Multiplicative Heteroskedasticity," *Journal of Econometrics*, 8, 227-236.
- Godfrey, L.G. (1996), "Some Results on the Glejser and Koenker Tests for Heteroskedasticity," *Journal of Econometrics*, 72, 275-299.
- Godfrey, L.G. and C.D. Orme (1999), "The Robustness, Reliability and Power of Heteroskedasticity Tests and the Use of the Bootstrap," *Econometric Reviews*, forthcoming.
- Goldfeld, S.M. and R.E. Quandt (1965), "Some Tests for Homoscedasticity," *Journal of the American Statistical Association*, 60, 539-547.
- Greene, W. (1977), *Econometric Analysis*, 3rd edition, Upper Saddle River: Prentice Hall.
- Griffiths, W.E. (1972), "Estimation of Actual Response Coefficients in the Hildreth-Houck Random Coefficient Model," *Journal of the American Statistical Association*, 67, 633-635.
- Griffiths, W.E. (1986), "A Bayesian Framework for Optimal Input Allocation with an Uncertain Stochastic Production Function," *Australian Journal of Agricultural Economics*, 30, 128-152.
- Griffiths, W.E. and J.R. Anderson (1982), "Using Time-Series and Cross-Section Data to Estimate a Production Function with Positive and Negative Marginal Risks," *Journal of the American Statistical Association*, 77, 529-536.
- Griffiths, W.E., R.G. Drynan and S. Prakash (1979), "Bayesian Estimation of a Random Coefficient Model," *Journal of Econometrics*, 10, 201-220.
- Griffiths, W.E. and G.G. Judge (1992), "Testing and Estimating Location Vectors When the Error Covariance Matrix is Unknown," *Journal of Econometrics*, 54, 121-138.
- Griffiths, W.E. and K. Surekha (1986), "A Monte Carlo Evaluation of the Power of Some Tests for Heteroscedasticity," *Journal of Econometrics*, 31, 219-231.
- Harvey, A.C. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, 44, 461-465.

- Harvey, A.C. and G.D.A. Phillips (1974), "A Comparison of the Power of Some Tests for Heteroscedasticity in the General Linear Model," *Journal of Econometrics*, 2, 307-316.
- Hidalgo, J. (1992), "Adaptive Estimation in Time Series Regression Models with Heteroskedasticity of Unknown Form," *Econometric Theory*, 8, 161-187.
- Hildreth, C. and J.P. Houck (1968), "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63, 584-595.
- Hill, R.C., J.R. Knight and C.F. Sirmans (1997), "Estimating Capital Asset Price Indexes," *Review of Economics and Statistics*, 80, 226-233.
- Hill, R.C., C.F. Sirmans and J.R. Knight (1999), "A Random Walk Down Main Street?," *Regional Science and Urban Economics*, 29, 89-103.
- Hooper, P.M. (1993), "Iterative Weighted Least Squares Estimations in Heteroscedastic Linear Models," *Journal of the American Statistical Association*, 88, 179-184.
- Hsieh, D.A. (1983), "A Heteroscedasticity-Consistent Covariance Matrix Estimator for Time Series Regression," *Journal of Econometrics*, 22, 281-290.
- Jobson, J.D. and W.A. Fuller (1980), "Least Squares Estimation when the Covariance Matrix and Parameter Vector are Functionally Related," *Journal of the American Statistical Association*, 75, 176-181.
- Judge, G.G., W. E. Griffiths, R.C. Hill and T.-C. Lee (1985), *The Theory and Practice of Econometrics*, New York: John Wiley and Sons.
- Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl and T.-C. Lee (1988), *An Introduction to the Theory and Practice of Econometrics*, New York: John Wiley and Sons.
- Just, R.E. and R.D. Pope (1978), "Stochastic Specification of Production Functions and Economic Implications," *Journal of Econometrics*, 7, 67-86.
- Keener, R. W., J. Kmenta and N.C. Weber (1991), "Estimation of the Covariance Matrix of the Least-Squares Regression Coefficients When the Disturbance Covariance Matrix is of Unknown Form," *Econometric Theory*, 7, 22-43.
- Koenker, R. (1981), "A Note on the Studentizing a Test for Heteroskedasticity," *Journal of Econometrics*, 17, 107-112.
- Koenker, R. and G. Bassett, Jr. (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43-61.
- Koenker, R., J.A.F. Machado, C.L. Skeels and A.H. Welsh (1993), "Amemiya's Form of the Weighted Least Squares Estimator," *Australian Journal of Statistics*, 35, 155-174.

- Lee, B.-J. (1992), "A Heteroskedasticity Test Robust to Conditional Mean Specification," *Econometrica* 60, 159-172.
- Li, Q. and T. Stengos (1994), "Adaptive Estimation in the Panel Data Error Model with Heteroskedasticity of Unknown Form," *International Economic Review*, 35, 981-1000.
- Linton, O.B. (1996), "Second Order Approximation in a Linear Regression with the Heteroskedasticity of Unknown Form," *Econometric Reviews*, 15, 1-32.
- Mackinnon, J.G. and H. White (1985), "Some Heteroskedasticity-Robust Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305-325.
- Magnus, J.R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," *Journal of Econometrics*, 7, 281-312.
- Mandy, D.M. and C. Martins-Filho (1993), "Seemingly Unrelated Regressions Under Additive Heteroscedasticity: Theory and Share Equation Applications," *Journal of Econometrics*, 58, 315-346.
- Neudecker, H. W. Polasek and S. Liu (1995), "The Heteroskedastic Linear Regression Model and the Hadamard Product: A Note," *Journal of Econometrics*, 68, 361-366.
- Oberhofer, W. and J. Kmenta (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," *Econometrica*, 42, 579-590.
- Orme, C. (1992), "Efficient Score Tests for Heteroskedasticity in Microeconometrics," *Econometric Reviews*, 11, 235-252.
- Pagan, A. and Y. Pak (1993), "Testing for Heteroskedasticity," in G.S. Maddala, C.R. Rao and H.D. Vinod, editors, *Handbook of Statistics II: Econometrics*, Amsterdam: North-Holland, 489-518.
- Raj, B. (1975), "Linear Regression With Random Coefficients: The Finite Sample and Convergence Properties," *Journal of the American Statistical Association*, 70, 127-137.
- Rilestone, P. (1991), "Some Monte Carlo Evidence on the Relative Efficiency of Parametric and Semiparametric EGLS Estimators," *Journal of Business and Economic Statistics*, 9, 179-187.
- Robinson, P.M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica*, 55, 875-891.
- Rutemiller, H.C. and D.A. Bowers (1968), "Estimation in a Heteroscedastic Regression Model," *Journal of the American Statistical Association*, 63, 552-557.

- Sharma, S.C. and C. Giaccotto (1991), "Power and Robustness of Jackknife and Likelihood Ratio Tests for Group Heteroskedasticity," *Journal of Econometrics*, 49, 343-372.
- Spanos, A. (1994), "On Modeling Heteroskedasticity: The Student's t and Elliptical Linear Regression Models," *Econometric Theory*, 10, 286-315.
- Surekha, K. and W.E. Griffiths (1984), "A Monte Carlo Comparison of Some Bayesian and Sampling Theory Estimators in Two Heteroscedastic Error Models," *Communications in Statistics B*, 13, 85-105.
- Surekha, K. and W.E. Griffiths (1985), "A Note on a Posterior Approximation in a Heteroscedastic Model", *Economics Letters*, 19, 249-252.
- Swamy, P.A.V.B. and J.S. Mehta (1979), "Estimation of Common Coefficients in Two Regression Equations," *Journal of Econometrics*, 10, 1-14.
- Szroeter, J. (1978), "A Class of Parametric Tests for Heteroskedasticity in Linear Econometric Models," *Econometrica*, 46, 1311-1328.
- Szroeter, J. (1994), "Exact Finite-Sample Relative Efficiency of Sub-Optimality Weighted Least Squares Estimators in Models with Ordered Heteroscedasticity," *Journal of Econometrics*, 64, 29-44.
- Taylor, W.E. (1977), "Small Sample Properties of a Class of Two-Stage Aitken Estimators," *Econometrica*, 45, 497-508.
- Taylor, W.E. (1978), "The Heteroscedastic Linear Model: Exact Finite Sample Results," *Econometrica*, 46, 663-675.
- Theil, H. (1971), *Principles of Econometrics*, Wiley, New York.
- Welsh, A.H. and R.J. Carroll and D. Ruppert (1994), "Fitting Heteroscedastic Regression Models," *Journal of the American Statistical Association*, 89, 100-116.
- White, H. (1980), "A Heteroscedasticity-Consistent Covariance Matrix Estimators and a Direct Test for Heteroscedasticity," *Econometrica*, 48, 817-838.
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley and Sons.