ESTIMATION OF HOUSEHOLD EXPENDITURE FUNCTIONS: AN APPLICATION OF A CLASS OF HETEROSCEDASTIC REGRESSION MODELS

George E. Battese and Bruce P. Bonyhady

No. 7 - September 1979

ISSN 0157-0188
ISBN 0 85834 259 6

Estimation of Household Expenditure Functions: An Application of a Class of Heteroscedastic Regression Models

George E. Battese and Bruce P. Bonyhady*

A class of heteroscedastic regression models is applied to the estimation of household expenditure functions using Australian data. The model permits the testing of the hypothesis that the variances of expenditure observations are proportional to the square of their expectations. Additive expenditure functions with observations having normal, gamma and lognormal distributions are compared with the traditional multiplicative model.

I. Introduction

The estimation of expenditure relationships has been the subject of much econometric research since the path-breaking work of Engel in 1857. Houthakker (1957) comments on some of this work and reports empirical results from a large number of countries. Many empirical investigations consider that expenditure on a consumable item is a function of total household expenditure (a proxy variable for income) and household size. Although the problem of heteroscedasticity of expenditure observations has been recognized, it appears that most econometric applications have considered functional forms for which the dependent variable is assumed to be homoscedastic. The double-logarithmic expenditure function has been widely used because it is claimed to provide a "good fit, ease of computation, and automatic correction for heteroscedasticity" [Houthakker (1957, p.543)].

The claim that the logarithmic transformation of expenditure observations corrects for heteroscedasticity is consistent with the belief that expenditure observations are such that their standard deviations are proportional to their corresponding means [see Bartlett (1947) or Kempthorne (1952, pp. 153-7)]. Theil (1951), Prais and

Houthakker (1955), Jorgenson (1965), and others have suggested that such an hypothesis is plausible for many expenditure items.[1]

More recently, Amemiya (1973) considered the regression model in which the variances of observations on the dependent variable are proportional to the square of their expectations. A small sample of biometric data was used to illustrate the methods of estimation and testing procedures presented in the paper. To our knowledge the statistical model considered by Amemiya has not been applied in household expenditure studies, nor has the hypothesis that variances of expenditure observations are proportional to the square of their expectations been tested.

In this paper we apply the heteroscedastic regression model considered by Amemiya (1973) to the estimation of expenditure functions using Australian data. The model is estimated for each of the assumptions that expenditure observations have normal, lognormal or gamma distributions. We apply the test statistics derived by Amemiya (1973) for testing the hypothesis that the dependent variable has lognormal distribution against the hypothesis that it has gamma distribution, and vice versa. In addition, we consider a more general heteroscedastic model, in which the variances of the observations on the dependent variable are proportional to an unknown power of their expectations. This model permits the testing of two particular hypotheses:

(i) the observations on the dependent variable are homoscedastic; and

(ii) the observations have variances that are proportional to the square of their expectations.

The data used in our empirical study are obtained from the Australian Survey of Consumer Expenditures and Finances that was conducted under the auspices of Macquarie University in the period 1966-68. Podder (1971) has reported results from estimating several different expenditure functions using the methods of ordinary least-squares and instrumental variables. We estimate the parameters for several heteroscedastic expenditure models for five of the expenditure variables considered by Podder (1971).

II. Heteroscedastic Expenditure Models

The variables considered in our empirical study are annual household expenditure on a given category of items, total annual household expenditure, and the number of individuals in the household. These are represented by Y_t, x_{1t} and x_{2t}, respectively, for the t-th household. For convenience, we consider that the Y-observations are independently distributed random variables and the x-observations are fixed constants. We obviously fail to consider possible correlations between the Y-observations resulting from interviewer effects in the survey procedure, and the errors-in-variables and simultaneity relationships in the variables involved. We use sample data for which nonzero observations are available for each category of expenditure.

The expenditure models considered are defined by

Model 1:

$$Y_t = \beta_0 + x_{1t}\beta_1 + x_{2t}\beta_2 + U_t, \qquad t=1,2,\ldots,T, \qquad (2.1)$$

where

$$E(Y_t) = \beta_0 + x_{1t}\beta_1 + x_{2t}\beta_2, \qquad t=1,2,\ldots,T, \qquad (2.2)$$

$$Var(Y_t) = \sigma^2(\beta_0 + x_{1t}\beta_1 + x_{2t}\beta_2)^p \qquad t=1,2,\ldots,T, \qquad (2.3)$$

β_0, β_1, β_2, σ^2 and p are parameters; and T represents the number of households for which survey data are considered for the given category of expenditure.

Five special cases of this heteroscedastic expenditure model are considered. These are defined by different distributional assumptions and values of the variance parameter, p, as follows:

Model 1a: Y_t has normal distribution with $p = 0$;

Model 1b: Y_t has gamma distribution with $p = 2$;

Model 1c: Y_t has lognormal distribution with $p = 2$;

Model 1d: Y_t has normal distribution with $p = 2$;

Model 1e: Y_t has normal distribution with p unknown.

Clearly the classical (homoscedastic) linear model is defined by Model 1a. The heteroscedastic models with the variance parameter, p, equal to 2.0 were considered by Amemiya (1973). Bonyhady (1975) considered the heteroscedastic model in which the variance parameter, p, is unknown and the dependent variable had normal or lognormal distributions. Basic theoretical results for the case of the normal distribution are given in the Appendix. It was considered that linear expenditure functions may provide an adequate fit to empirical data, given that possible heteroscedasticity was suitably modelled. Although Houthakker (1957, p.539) states, "Linear functions were used by Allen and Bowley, but it is now generally recognized that they do not provide an adequate fit", it is not clear upon what criteria the claim is based.

In addition to the five linear expenditure functions, we estimate the multiplicative model that is defined by

Model 2:

$$Y_t = e^{\gamma_0} x_{1t}^{\gamma_1} x_{2t}^{\gamma_2} e^{V_t}, \qquad t=1,2,\ldots,T, \qquad (2.4)$$

where the random errors, V_t, $t=1,2,\ldots,T$, are assumed to be identically distributed as normal random variables with mean zero and constant variance, η^2.

Model 2 is commonly used in econometric studies because it reduces to the classical (homoscedastic) linear model when logarithms are taken. It is noted that the expenditure function of Model 2 is related to that of Model 1 in that the variances of the expenditure observations are proportional to the square of their expectations.[2]

The maximum-likelihood estimates for the parameters of the six household expenditure functions are obtained and the values of the loglikelihood functions calculated so that a comparison can be made between the fit of the different models to the available sample data.

For the heteroscedastic model (2.1)-(2.3) with the variance parameters, p, equal to two, the hypotheses that the expenditure variables have lognormal or gamma distributions are tested by calculating the values of two random variables of somewhat complicated form.[3] The first random variable, $R_f/[\mathrm{Var}(R_f)]^{\frac{1}{2}}$, is involved when the null hypothesis, H_0: Y_t has lognormal distribution, represented by f, is tested against the alternative hypothesis, H_1: Y_t has gamma distribution, represented by g. The second random variable, $R_g/[\mathrm{Var}(R_g)]^{\frac{1}{2}}$, is involved when the null hypothesis, H_0: Y_t has gamma distribution, is tested against the alternative hypothes, H_1: Y_t has lognormal distribution. The two random variables have asymptotically standard normal distributions when the respective null hypotheses are true. For a given test procedure, the null hypothesis is rejected in favour of the alternative hypothesis if the value of the test statistic is too small (i.e., a negative number with absolute value that is too large). When both tests are applied, a particular distributional assumption, say, gamma, g, is clearly established, if $R_f/[\mathrm{Var}(R_f)]^{\frac{1}{2}}$ is too small, but $R_g/[\mathrm{Var}(R_g)]^{\frac{1}{2}}$ is not too small. A large positive value of a test statistic is taken to imply departure from the null hypothesis away from the alternative hypothesis [see Cox (1962, p.407)].

For the linear expenditure functions (2.1)-(2.3), the partial elasticities are estimated at the mean values and are compared with those for the expenditure function (2.4) with constant elasticities. The elasticity parameters for total expenditure and household size, denoted by e_1 and e_2, respectively, are defined by $e_1 = \bar{x}_1\beta_1/E(\bar{Y})$

and $e_2 = \bar{x}_2\beta_2/E(\bar{Y})$, where $E(\bar{Y}) \equiv \beta_0 + \bar{x}_1\beta_1 + \bar{x}_2\beta_2$ and $\bar{x}_1$ and $\bar{x}_2$ denote the average total expenditure and average household size for the survey households having a positive expenditure on the expenditure category involved. The standard errors of the maximum-likelihood estimators for the elasticities are estimated with use of the approximate formula for the variance of the ratio of two random variables [e.g., see Mood, Graybill and Boes (1974, p. 181)].

The expenditure functions (2.1) and (2.4) are defined for positive expenditure values. This is consistent with the distributional assumptions for Models 1b, 1c and 2. It is obviously only an approximation to state that a positive random variable has normal distribution. It is implicit that our modelling for a given expenditure category is for the population of households that purchase positive quantities of the items involved. When considering modelling expenditures for all households in a large population it is evident that the expenditure random variables involved are likely to be combinations of discrete and continuous random variables. For example, if expenditure on cigarettes is of interest, then there is a positive probability of expenditure being zero (namely, the proportion of the population which does not purchase cigarettes). In considering households which make purchases of these items, then a conditional distribution is defined.

This point is significant for the analysis of empirical data from a cross-sectional survey of households. It is theoretically inconsistent to include a large number of zero observations in the estimation of an expenditure function, for which an assumption of normality, lognormality or gamma distribution is made. It appears that Podder (1971) considered all sample households (a total of 5443) in the empirical analysis for all the nine categories of expenditure that he considered.[4]

III Empirical Results

The expenditure variables considered in our study, together with averages of the relevant variables, are listed in Table 1. The five variables involved are the first five considered by Podder (1971) and account for about 56% of the total expenditures of the sample households. It is noted that the number of observations for expenditure on food is 5437, which is six less than the number reported by Podder (1971). As reported in Bonyhady (1975), a check of the data tape obtained from Macquarie University revealed that there were 5442 sample households for which all expenditure data are present. Five of these households were found to be coded as having "zero food expenditure". Only the remaining 5437 households were included in our analyses on total expenditure on food. For the other expenditure variables considered, further households were coded as having "zero expenditures". Although these may be correct values for the expenditure categories involved (e.g., it is not unreasonable that about 25% of households of the population have zero purchases on cigarettes, tobacco and liquor),these households are excluded for our analysis of the appropriate conditional distributions. The average expenditure values given by Podder (1971) are obtained by multiplying the values in Table 1 by the ratio of our sample size to that used by Podder.

The maximum likelihood estimates for the parameters of the six expenditure models applied to the five expenditure variables are given in Table 2. The computer programmes used are those given in appendices of Bonyhady (1975); however, non-linear regression packages may be used to compute these maximum-likelihood estimates. The weighted least-squares estimates for the coefficients of the expectation model (2.2) are used to the case of the gamma distribution (Model 1b), because the weighted least-squares estimator is asymptotically efficient [see Amemiya (1973, p.930)]. The maximum likelihood estimate for the constant, σ^2, of the variance model (2.3) is obtained by iterative methods that involve the digamma function, ψ [see equation (2.25) of Amemiya (1973) and Table 6.1 of Abramowitz and Stegun (1965)]. The approximate maximum-likelihood estimates for the parameters of the heteroscedastic expenditure functions involving the lognormal and normal distributions (Models 1c, 1d and 1e) were obtained by performing iterations of the Newton-Raphson method until the changes in all coefficients were less than 0.0005. In general, few iterations were required for the normal distribution cases, but a relatively large number of iterations were required for the lognormal distribution. For example, for expenditure on food, the desired precision of estimation was achieved in four iterations for the normal case with p equal to two (Model 1d), fifteen iterations for the normal case with p unknown (Model 1e), and over 200 iterations for the lognormal case (Model 1c). A smaller number of iterations would be required if the terminating conditions were less severe than those used in this study.

The results obtained for the heteroscedastic model with the unknown power in the variance (Model 1e) indicate that the hypothesis of homoscedasticity of the expenditure observations is rejected for all five expenditure variables.[5] In addition, under the assumption of the normal distribution, the hypothesis that the variance of expenditure observations is proportional to the square of the expectations (i.e., H_0: $p = 2.0$) is also rejected, given the large data sets involved in this study, for all five expenditure variables. For the expenditure variables, other than cigarettes, tobacco and liquor, the estimates for the power parameter are significantly less than two.

A comparison of the six heteroscedastic models indicates that for none of the five expenditure variables is the value of the loglikelihood function for the multiplicative case (Model 2) the greatest. For the expenditure on food, the additive expenditure model with the assumption of normality and an unknown power in the variance (Model 1e) has the largest value of the loglikelihood function. For the other four expenditure variables, the model with the assumption of the gamma distribution with the variance power equal to two has largest value of the loglikelihood function. For all five expenditure variables, the model involving the gamma distribution has larger loglikelihood value than that for the lognormal distribution.

That the assumption of the gamma distribution fits the expenditure data better than that of lognormality is confirmed by the tests of separate families of hypotheses which are reported in Table 3. For the test of the null hypothesis, H_0: expenditure observations have lognormal distribution, against the alternative hypothesis, H_1: expenditure observations have gamma distribution, the values of the appropriate test statistic are negative, and large in absolute value, for all five expenditure variables considered. These results indicate that there is a departure from lognormality <u>in the direction</u> of the gamma distribution. For the test of the null hypothesis, H_0: expenditure observations have gamma distribution, against the alternative hypothesis, H_1: expenditure observations have lognormal distribution, the values of the relevant test statistic are positive and large for four of the five expenditure variables. These results indicate that for the variables, excluding expenditure on cigarettes, tobacco and liquor, there is a departure from the gamma distribution <u>away from</u> lognormality. The results for expenditure on cigarettes, tobacco and liquor indicate that its distribution requires specifications between those for the lognormal and gamma distributions.[6]

The partial elasticities estimated for the different expenditure variables and the six expenditure models show reasonable consistency as to the sign. Only for expenditure on clothing are the household-size elasticities significantly negative for the heteroscedastic additive models, but positive for the multiplicative model. It appears that most of the partial elasticities for the additive heteroscedastic model having gamma distribution are significantly different from those for the multiplicative model. This is particularly so for the elasticities with respect to total expenditure. For example, for expenditure on clothing the estimated elasticities are 1.0721 and 1.2030 for Model 1b and Model 2, respectively. Relative to their estimated standard errors, the difference between the estimates is large.

A comparison of the results presented in Table 2 and those obtained by Podder (1971, p.391) reveals that the estimated partial elasticities for the multiplicative expenditure model are quite different. In fact, the values given by Podder, and their estimated standard errors, are larger than the corresponding values in Table 2. This is due to the "zero observations" used by Podder and the differences in the results obtained increase as the proportion of zero observations increases. In the case of expenditure on cigarettes, tobacco and liquor, for which about 25% of sample households had zero expenditure, the elasticity with respect to total expenditure was estimated by Podder (1971) to be 1.573, with an estimated standard error of 0.061. In our study the elasticity is estimated to be 0.6821 and its estimated standard error is 0.0265.

IV Conclusions

The class of heteroscedastic regression models considered in this paper permits the testing of the hypothesis that the variances of expenditure observations are proportional to the square of their expectations. Although, under the assumption of normality of expenditure observations, this hypothesis is rejected given the extensive sample data used in our study, it appears that such an hypothesis is a reasonable approximation for the variables considered.

Despite some claims to the contrary, additive expenditure functions, for which heteroscedasticity is appropriately modelled, may fit better than the corresponding multiplicative functions. The assumption of the gamma distribution is demonstrated to be more appropriate than the lognormal for the five expenditure variables considered in this analysis. This is particularly advantageous because the maximum-likelihood estimates for the coefficients of the expectations of the additive expenditure model are approximated by weighted least-squares regression in the case of the gamma distribution (Model 1b) and so may be estimated using a standard ordinary least-squares regression package.

The heteroscedastic regression models considered in this paper are worthy of consideration for statistical analyses of economic data. Alternative heteroscedastic models have been suggested in the literature. For example, Rutemiller and Bowers (1968) have suggested a model in which the standard deviations of observations on the dependent variable are proportional to a different linear combination of values of independent variables than is specified by their expectations. Harvey (1976) has suggested an heteroscedastic model in which the variances of observations of the dependent variable are a multiplicative function of values of variables that might be different from those involved in their expectations. Harvey (1976, p.465) suggests that "From the point of view of estimation, the multiplicative heteroscedasticity model...appears to be rather more attractive than the 'additive' model in which either the variance or the standard deviation ... is assumed to be related to a linear combination of known variables." It should be noted, however, that distributional assumptions for the observations on the dependent variable are likely to have implications on the form of the model to account for heteroscedasticity. As stated by Amemiya (1973, p.928) the heteroscedastic model (2.3) in which $p = 2$ naturally arises when the observations of the dependent variable follow either a lognormal or a gamma distribution.[7]

V. Appendix

Consider the statistical model defined by

$$Y_t = x_t\beta + U_t, \qquad t=1,2,\ldots,T, \tag{5.1}$$

where Y_t, $t=1,2,\ldots,T$ are independent normal random variables with means

$$E(Y_t) = x_t\beta \quad , \qquad t=1,2,\ldots,T, \tag{5.2}$$

and variances

$$\mathrm{Var}(Y_t) = \sigma^2(x_t\beta)^p, \quad t=1,2,\ldots,T, \tag{5.3}$$

where x_t, $t=1,2,\ldots,T$, are $(1\times k)$ vectors of known constants such that the matrix $\sum_{t=1}^{T} x_t'x_t$ has rank k;

β is a $(k \times 1)$ vector of unknown parameters;

σ^2 is an unknown positive parameter; and

p is an unknown parameter.

If sample observations on the dependent variable are denoted by $y = (y_1, y_2,\ldots,y_T)'$, then the loglikelihood function is given by

$$L(\theta;y) = -\tfrac{1}{2}T\ell n 2\pi - \tfrac{1}{2}T\ell n\sigma^2 - \tfrac{1}{2}p\sum_{t=1}^{T}\ell n(x_t\beta) - \tfrac{1}{2}\sum_{t=1}^{T}\frac{(y_t-x_t\beta)^2}{\sigma^2(x_t\beta)^p} \tag{5.4}$$

where $\theta = (\beta', \sigma^2, p)'$.

The first-order partial derivatives of the loglikelihood function are

$$\frac{\partial L(\theta;y)}{\partial\beta} = -\tfrac{1}{2}p\sum_{t=1}^{T}\frac{x_t'}{x_t\beta} + \tfrac{1}{2}\sum_{t=1}^{T}\left[\frac{2(y_t-x_t\beta)}{\sigma^2(x_t\beta)^p} + \frac{p(y_t-x_t\beta)^2}{\sigma^2(x_t\beta)^{p+1}}\right]x_t' \tag{5.5}$$

$$\frac{\partial L(\theta;y)}{\partial\sigma^2} = -\tfrac{1}{2}T/\sigma^2 + \tfrac{1}{2}\sum_{t=1}^{T}\frac{(y_t-x_t\beta)^2}{\sigma^4(x_t\beta)^p} \tag{5.6}$$

$$\frac{\partial L(\theta;y)}{\partial p} = -\tfrac{1}{2}\sum_{t=1}^{T}\ell n(x_t\beta) + \tfrac{1}{2}\sum_{t=1}^{T}\frac{(y_t-x_t\beta)^2\ell n(x_t\beta)}{\sigma^2(x_t\beta)^p} \tag{5.7}$$

Given that the necessary regularity conditions are satisfied, the maximum-likelihood estimator for θ, denoted by $\hat{\theta}$, is such that the random variable, $\sqrt{T}(\hat{\theta}-\theta)$, converges in distribution, as T approaches infinity, to the $(k+2)$-variate normal with zero mean vector and covariance matrix

$$\text{Asy Cov}[\sqrt{T}(\hat{\theta}-\theta)] = \lim_{T\to\infty} \{E[-\frac{1}{T} \frac{\partial^2 L(\theta; Y)}{\partial\theta\,\partial\theta'}]\}^{-1}. \qquad (5.8)$$

This covariance matrix is given by

$$\text{Asy Cov}[\sqrt{T}(\hat{\theta}-\theta)] = \begin{bmatrix} \frac{1}{2}p^2 A_2 + A_p/\sigma^2 & \frac{1}{2}pA_2\beta/\sigma^2 & \frac{1}{2}pA_2^{*}\beta \\ \frac{1}{2}p\beta' A_2/\sigma^2 & \frac{1}{2}\sigma^{-4} & \frac{1}{2}C_1/\sigma^2 \\ \frac{1}{2}p\beta' A_2^{*} & \frac{1}{2}C_1/\sigma^2 & \frac{1}{2}C_2 \end{bmatrix} \qquad (5.9)$$

where $A_2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [x_t' x_t/(x_t\beta)^2]$

$$A_p = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [(x_t' x_t)/(x_t\beta)^p]$$

$$A_2^{*} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [\ln(x_t\beta)\, x_t' x_t/(x_t\beta)^2]$$

$$C_1 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \ln(x_t\beta)$$

and

$$C_2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [\ln(x_t\beta)]^2.$$

The estimate for the covariance matrix for the maximum-likelihood estimator, $\hat{\theta}$, is defined by the sample estimate for the matrix on the right-hand side of (5.9), evaluated at the maximum-likelihood estimate, $\hat{\theta}$, divided by the sample size, T. This matrix is positive definite and so is preferable to the estimate that is defined by -

$$\left[\frac{\partial^2 L(\hat{\theta},y)}{\partial\theta\,\partial\theta'}\right]^{-1}.$$

It should be noted that the class of heteroscedastic models defined by (5.1) - (5.3) is such that the units in which the observations are scaled have a bearing on the suitability of the model. If one defines the random variable, Y_t^*, as a multiple of Y_t, say, $Y_t^* = c\, Y_t$, then the variance of Y_t^* is <u>not</u> proportional to the power, p, of the expectation of Y_t^* unless $p = 2$. Further, if the model (5.1) - (5.3) is defined for individual data, but only grouped data is available, then the variances of the group means is <u>not</u> proportional to a power of their expectations even if the power is known to be 2.0.

Footnotes

*The present addresses of the authors are: Department of Economic Statistics, University of New England, Armidale, N.S.W., 2351, and The Treasury, Canberra, A.C.T., 2600, respectively.

1. Transformation of observations to obtain homoscedastic observations remains the subject of research. Box and Cox (1964) consider the model in which transformed observations, $Y_t^{(\lambda)}$, are homoscedastic and normally distributed, where λ is an unknown parameter such that $Y_t^{(\lambda)} = (Y_t^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$ and $Y_t^{(\lambda)} = \ln Y_t$ if $\lambda = 0$. It can be shown that this power transformation obtains approximately homoscedastic observations if the original observations, Y_t, $t=1,2,\ldots,T$, have variance structure defined by $\sigma_t^2 = \sigma^2 \mu_t^{2(1-\lambda)}$, $t=1,2,\ldots,T$, where μ_t and σ_t^2 are the mean and variance of Y_t. (This is related to the variance model considered in this paper.) Box and Hill (1974) and Pritchard and Bacon (1977) have reported estimates of the power-transformation parameter, λ, between -0.8 and 0.72 for analyses of chemical engineering data.

2. For Model 2 it follows that

$$E(Y_t) = (\prod_{i=0}^{2} x_{it}^{\gamma_i}) e^{\frac{1}{2}\eta^2} \equiv \mu_t, \text{ where } x_{0t} = e \text{ for all } t, \text{ and}$$

$$Var(Y_t) = (\prod_{i=0}^{2} x_{it}^{\gamma_i})^2 (e^{2\eta^2} - e^{\eta^2}) \equiv \sigma^2(\mu_t)^2, \text{ where } \sigma^2 = e^{\eta^2} - 1.$$

3. See Amemiya (1973) for the form of these random variables.

4. Presumably, the "zero observations" were coded as 1.0 by Podder (1971) so that the logarithmic transformation could be performed before estimating the double-logarithmic expenditure functions on page 391.

5. This result follows from the fact that under suitable regularity conditions the random variable, $\frac{\hat{P}-p}{[\hat{Var}(\hat{P})]^{\frac{1}{2}}}$, converges in distribution to the standard normal random variable, where $\hat{P}$ denotes the maximum-likelihood estimator for the parameter, p, and $\hat{Var}(\hat{P})$ denotes the maximum-likelihood estimator for the variance of $\hat{P}$, (see the Appendix).

6. The lognormal distribution is more positively skewed and peaked than the gamma distribution. For the heteroscedastic model (2.1) - (2.3) with p=2, the coefficients of skewness and kurtosis, γ_1, and γ_2, respectively, have values $\gamma_1 = 2\sigma$ and $\gamma_2 = 6\sigma^2$ for the gamma distribution and $\gamma_1 = \sigma(3 + \sigma^2)$ and $\gamma_2 = \sigma^2[(2 + \sigma^2)^3 + 3(1 + \sigma^2) + 5]$ for the lognormal distribution (see Kendall and Stuart (1969, p.85 and p.169)).

7. If Y has lognormal distribution and is defined by Y = exp(X), where X has normal distribution with mean μ_X and variance σ_X^2, then the mean and variance of Y are given by $E(Y) = \exp(\mu_X + \frac{1}{2}\sigma_X^2) \equiv \mu_Y$ and

$Var(Y) = \sigma^2(\mu_Y)^2$, where $\sigma^2 = \exp(\sigma_X^2) - 1$ (c.f., Mood, Graybill and Boes (1974), p.117). Further, if Y has gamma distribution with parameters r and λ (see Mood, Graybill and Boes (1974), p.112), then $E(Y) = r/\lambda \equiv \mu_Y$ and $Var(Y) = r/\lambda^2 \equiv \sigma^2(\mu_Y)^2$, where $\sigma^2 = 1/r$.

The heteroscedastic model of this paper assumes that the means, μ_t, t=1,2,...,T, of the observations of the dependent variable are described by a reduced number of parameters and that the variances, σ_t^2, t=1,2,...,T, are defined in terms of the means by the same constant of proportionality, σ^2, i.e., $\sigma_t^2 = \sigma^2(\mu_t)^2$, t=1,2,...,T.

Table 1: Expenditure Variables and Average Data for the 1966-68 Macquarie University Survey

Expenditure Variable	Sample Size	Averages for Survey Data: Expenditure*	Total Exp.	House size
1. Food (incl. away-from-homes expenses)	5437	1177.30 (30.0%)	3922.02	3.49
2. Clothing, footwear and manchester goods	5246	323.56 (8.1%)	3998.41	3.54
3. Cigarettes, tobacco and liquor	4099	256.97 (6.1%)	4218.49	3.65
4. Utilities (fuel,gas, electricity & telephone)	5331	142.24 (3.6%)	3939.52	3.51
5. Transport (fares and motor vehicle running expenses)	4974	409.67 (10.0%)	4108.34	3.60

*The figure below the average expenditure for a given item is the percentage of the total expenditure for households with positive expenditure on the given item.

Table 2: Parameter Estimates for Household Expenditure Functions*

MODEL	Coefficients of Functions: Constant	Total Exp.	House Size	Variance Parameters: Constant**	Power	Log-Likelihood	Elasticities: Total Exp.	House Size
Food (T = 5437)								
Model 1a	226.1246	0.1304	125.8525	123396.40	0.0	-39583	0.4345	0.3735
	(12.29)	(0.0024)	(3.0217)				(0.0082)	(0.0091)
Model 1b	150.5580	0.1524	124.9054	0.0863	2.0	-38965	0.5045	0.3684
	(7.46)	(0.0027)	(3.1337)	(0.0016)			(0.0085)	(0.0089)
Model 1c	117.9993	0.1533	136.6095	0.1025	2.0	-39221	0.5025	0.3989
	(7.72)	(0.0029)	(3.3339)	(0.0021)			(0.0090)	(0.0094)
Model 1d	165.5522	0.1668	104.3291	0.0814	2.0	-38966	0.5524	0.3078
	(6.77)	(0.0025)	(2.7606)	(0.0017)			(0.0077)	(0.0080)
Model 1e	158.6535	0.1648	108.4739	0.4461	1.7565	-38952	0.5459	0.3201
	(7.46)	(0.0025)	(2.7994)	(0.1091)	(0.0348)		(0.0078)	(0.0081)
Model 2	2.6575	0.4782	0.3683	0.0971	2.0	-39078	0.4782	0.3683
	(0.0647)	(0.0085)	(0.0091)				(0.0085)	(0.0091)
Clothing (T = 5246)								
Model 1a	-40.2979	0.0858	5.8878	60608.45	0.0	-36327	1.0601	0.0645
	(9.00)	(0.0017)	(2.1555)				(0.0239)	(0.0236)
Model 1b	-3.8626	0.0839	-5.2746	0.5208	2.0	-34122	1.0721	-0.0597
	(3.18)	(0.0017)	(1.4911)	(0.0094)			(0.0150)	(0.0168)
Model 1c	-23.2452	0.0962	-3.8592	1.0407	2.0	-34501	1.1062	-0.0393
	(2.79)	(0.0021)	(1.8481)	(0.0284)			(0.0155)	(0.0188)
Model 1d	9.9097	0.0850	-9.7948	0.4806	2.0	-34954	1.0786	-0.1101
	(2.50)	(0.0013)	(1.0088)	(0.0131)			(0.0104)	(0.0112)
Model 1e	3.5063	0.0881	-10.3410	1.2350	1.8302	-34946	1.1039	-0.1149
	(2.62)	(0.0013)	(1.0536)	(0.1197)	(0.0164)		(0.0108)	(0.0115)
Model 2	-4.5827	1.2030	0.0823	1.0295	2.0	-34479	1.2030	0.0823
	(0.1894)	(0.0247)	(0.0257)				(0.0247)	(0.0257)
Cigarettes (T = 4099)								
Model 1a	95.5835	0.0454	-8.2521	49966.97	0.0	-27988	0.7453	-0.1173
	(9.84)	(0.0017)	(2.2058)				(0.0301)	(0.0314)
Model 1b	73.7524	0.0489	-5.9729	0.4819	2.0	-26081	0.7989	-0.0845
	(5.99)	(0.0017)	(1.5267)	(0.0099)			(0.0234)	(0.0215)
Model 1c	77.3565	0.0458	-1.8789	0.7814	2.0	-26127	0.7327	-0.0260
	(6.85)	(0.0019)	(1.7877)	(0.0227)			(0.0264)	(0.0247)
Model 1d	73.0768	0.0557	-12.5654	0.5675	2.0	-27164	0.8964	-0.1750
	(4.40)	(0.0013)	(1.0505)	(0.0183)			(0.0162)	(0.0143)
Model 1e	86.1703	0.0516	-12.0071	0.1092	2.2991	-27157	0.8371	-0.1688
	(4.15)	(0.0013)	(0.9787)	(0.0056)	(0.0070)		(0.0155)	(0.0135)
Model 2	-0.2996	0.6821	-0.0731	0.7757	2.0	-26115	0.6821	-0.0731
	(0.2073)	(0.0265)	(0.0265)				(0.0265)	(0.0265)

Table 2 (Cont.)

Utilities (T = 5331)								
Model 1a	68.9881	0.0135	5.7166	4659.11	0.0	-30077	0.3738	0.1412
	(2.43)	(0.0005)	(0.5966)				(0.0132)	(0.0148)
Model 1b	63.9235	0.0152	5.3164	0.2361	2.0	-29548	0.4203	0.1311
	(2.02)	(0.0006)	(0.6201)	(0.0044)			(0.0152)	(0.0152)
Model 1c	58.7529	0.0156	7.3713	0.3288	2.0	-29834	0.4206	0.1773
	(2.21)	(0.0006)	(0.7040)	(0.0073)			(0.0166)	(0.0168)
Model 1d	70.0446	0.0160	2.5904	0.2248	2.0	-29864	0.4439	0.0640
	(1.69)	(0.0005)	(0.4923)	(0.0052)			(0.0123)	(0.0121)
Model 1e	66.9603	0.0165	2.9996	1.0061	1.6957	-29858	0.4564	0.0739
	(1.81)	(0.0005)	(0.5158)	(0.1234)	(0.0246)		(0.0128)	(0.0127)
Model 2	1.5882	0.3743	0.1531	0.3252	2.0	-29807	0.3743	0.1531
	(0.1140)	(0.0150)	(0.0162)				(0.0150)	(0.0162)
Transport (T = 4974)								
Model 1a	39.9792	0.0858	4.7213	92011.00	0.0	-35482	0.8608	0.4153
	(11.83)	(0.0022)	(2.7358)				(0.0235)	(0.0241)
Model 1b	-17.5203	0.1022	3.6339	0.4484	2.0	-33578	1.0107	0.0315
	(4.36)	(0.0021)	(1.9819)	(0.0084)			(0.0156)	(0.0172)
Model 1c	-43.0863	0.1139	7.7932	0.8711	2.0	-34040	1.0332	0.0620
	(3.64)	(0.0026)	(2.5019)	(0.0235)			(0.0169)	(0.0200)
Model 1d	2.9993	0.1007	-1.3061	0.4086	2.0	-34183	1.0041	-0.0114
	(3.61)	(0.0016)	(1.3725)	(0.0110)			(0.0114)	(0.0120)
Model 1e	-8.7782	0.1042	-1.2366	1.1238	1.8284	-34175	1.0319	-0.0107
	(3.63)	(0.0017)	(1.4250)	(0.0774)	(0.0108)		(0.0116)	(0.0124)
Model 2	-3.3852	1.0872	0.1101	0.8793	2.0	-34057	1.0872	0.1101
	(0.1935)	(0.0249)	(0.0251)				(0.0249)	(0.0251)

* Estimated standard deviations for the estimators are given below the corresponding parameter estimates.

** For Model 2 the variance estimate given is for the parameter $\ell n(1 + \eta^2)$.

Table 3: Hypothesis Tests Involving the Gamma and Lognormal Distributions

Expenditure Variable, Y_t	Values of Test Statistics* $H_0: Y_t \sim$ lognormal distribution $R_f/[Var(R_f)]^{\frac{1}{2}}$	$H_0: Y_t \sim$ gamma distribution $R_g/[Var(R_g)]^{\frac{1}{2}}$
1. Food	-30.98	5.62
2. Clothing	-21.89	5.04
3. Cigarettes	-9.80	-6.12
4. Utilities	-23.79	10.99
5. Transport	-26.20	11.57

*The test statistics involved converge in distribution to the standard normal random variable when the respective null hypotheses are true, [see Amemiya (1973, p.932)].

References

Abramowitz, M. and Stegun, I.A. (1965) Handbook of Mathematical Functions, Dover, New York.

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.

Bartlett, M.S. (1947) "The Use of Transformations", Biometrics 3, 39-52.

Bonyhady, B.P. (1975) Estimation of Linear Models with Variances a Power of the Expectations: With Application to a Consumer Expenditure Relation, Unpublished B.App.Ec. thesis, University of New England, Armidale, N.S.W.

Box, G.E.P. and Cox, D.R. (1964) "An Analysis of Transformations", Journal of the Royal Statistical Society, Series B, 26, 211-252.

Box, G.E.P. and Hill, W.J. (1974) "Correcting Inhomogeneity of Variance with Power Transformation Weighting", Technometrics 16, 385-389.

Cox, D.R. (1962) "Further Results on Tests of Separate Families of Hypotheses", Journal of the Royal Statistical Society, Series B, 24, 406-424.

Harvey, A.C. (1976) "Estimating Regression Models with Multiplicative Heteroscedasticity", Econometrica 44, 461-465.

Houthakker, H.S. (1957) "An International Comparison of Household Expenditure Patterns, Commemorating the Centenary of Engel's Law", Econometrica 25, 532-551.

Jorgenson, E. (1965) Income-Expenditure Relations of Danish Wage and Salary Earners, The Statistical Department, Copenhagen.

Kempthorne, O. (1952) The Design and Analysis of Experiments, Wiley, New York.

Kendall, M.G. and Stuart, A. (1969) The Advanced Theory of Statistics, Vol.1, Third Edition, Griffin, London.

Mood, A.M. Graybill, F.A. and Boes, D.C. (1974) Introduction to the Theory of Statistics, 3rd Edition, McGraw-Hill, New York.

Podder, N. (1971) "Patterns of Household Consumption Expenditures in Australia", Economic Record 47, 379-398.

Prais, S.P. and Houthakker, H.S. (1955) The Analysis of Family Budgets, Cambridge University Press, London.

Pritchard, D.J. and Bacon, D.W. (1977) "Accounting for Heteroscedasticity in Experimental Design", Technometrics 19, 109-115.

Rutemiller, H.C. and Bowers, D.A. (1968) "Estimation in a Heteroscedastic Regression Model", Journal of the American Statistical Association 63, 552-557.

Theil, H. (1951) "Estimates and Their Sampling Variance of Parameters of Certain Heteroscedastic Disturbances", Review of the International Statistical Institute 19, 141-147.