NBER WORKING PAPER SERIES

HETEROSKEDASTICITY-CONSISTENT ESTIMATION OF THE VARIANCE-COVARIANCE MATRIX FOR THE ALMOST IDEAL DEMAND SYSTEM

Melvyn A. Fuss

Working Paper No. 2401

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1987

Financial support from the Association of Canadian Distillers through a grant to the Institute for Policy Analysis' Policy and Economic Analysis Program (PEAP) is gratefully acknowledged. However any opinions expressed in this paper are to be attributed solely to the author since they may not represent the views of the Association. An earlier version of this note appeared as Appendix A to Fuss and Waverman. The research reported here is part of the NBER's research program in Productivity. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #2401
October 1987

Heteroskedasticity-Consistent Estimation of the Variance-Covariance Matrix for the Almost Ideal Demand System

ABSTRACT

In this note I demonstrate the previously overlooked fact that if the AIDS aggregate demand model is constructed as the aggregation of individual consumer demands, then the error structure for any individual equation is necessarily heteroskedastic unless the distribution of income is constant across aggregates. Maximum likelihood estimation which ignores this heteroskedasticity yields inconsistent estimates of the variance-covariance matrix and renders likelihood ratio tests of the restrictions of consumer demand theory inappropriate. A heterokedasticity-consistent estimator of the variance-covariance matrix is proposed by adopting the technique of White (1980) to the case at hand.

Melvyn A. Fuss
Department of Economics
University of Toronto
150 St. George Street
Toronto, Ontario
CANADA M5S 1A1

1. Introduction

The Almost Ideal Demand System (AIDS) has quickly become a popular functional form for estimation and testing of consumer demand theory since its introduction by Deaton and Muellbauer (1980).[1] AIDS provides a first order approximation to any arbitrary demand system, and can "almost" be made consistent with classical demand theory through linear resrictions on the parameters of the system.[2] Since most empirical applications of consumer demand theory employ aggregate data, the desirable aggregation properties of AIDS renders it a particularly useful functional form.[3]

In this note I demonstrate the previously overlooked fact that if the AIDS aggregate demand model is constructed as the aggregation of individual consumer demands, then the error structure for any individual equation is necessarily heteroskedastic unless the distribution of income is constant across aggregates. Maximum likelihood estimation which ignores this heteroskedasticity yields inconsistent estimates of the variance-covariance matrix and renders likelihood ratio tests of the restrictions of consumer demand theory inappropriate. A heterokedasticity-consistent estimator of the variance-covariance matrix is proposed by adopting the technique of White (1980) to the case at hand.

2. The AIDS Model and Aggregation

Muellbauer (1975, 1976) has shown that exact aggregation with respect to AIDS is possible if the demand system takes the form

$$S_{hki} = \alpha_k + \sum_j \gamma_{kj} \log p_{ji} + \beta_k \log [x_{hi}/k_{hi} \cdot P_i] + \varepsilon_{hki} \quad (1)$$

where S_{hki} is the expenditure share for the hth consumer of the kth commodity in the ith aggregate, p_{ji} is the price of the jth commodity (faced by all consumers in aggregate i), x_{hi} is an error term which is assumed to be iid across consumers and homoskedastic. P_i is a price index defined by

$$\log P_i = \alpha_o + \sum_k \alpha_k \log p_{ki} + \frac{1}{2} \sum_k \gamma_{kk}(\log p_{ki})^2$$

$$+ \sum_k \sum_{\substack{j \\ j<k}} \gamma_{kj} \log p_{ki} \log p_{ji} \quad (2)$$

and α_k, β_k, γ_{kj} are parameters to be estimated.

The term x_{hi}/k_{hi} in (1) is called an "income equivalent" and can be interpreted as the total expenditure by consumer h such that, if different consumers h with different preferences faced the same prices, they would choose the same expenditure shares of the commodities. This specification allows for a limited amount of taste variation among consumers without requiring linear Engel curves for exact aggregation to be valid.

Aggregating (1) over all consumers in the ith aggregate yields

$$\bar{S}_{ki} = \alpha_k + \sum_j \gamma_{kj} \log p_{ji} + \beta_k[\log (\bar{x}_i/k_i) - \log P_i] + \bar{\varepsilon}_{ki} \quad (3)$$

where

$$\bar{S}_{ki} = \frac{\sum_h p_{ki} \cdot A_{hki}}{\sum_h x_{hi}} = \frac{\sum_h x_{hi} \cdot S_{hki}}{\sum_h x_{hi}} \quad (4)$$

for A_{hki} = the quantity demanded of the kth commodity,

$$\bar{\varepsilon}_{ki} = \sum_h x_{hi} \cdot \varepsilon_{hki} / \sum_h x_{hi} \qquad (5)$$

and, following Deaton and Muellbauer (1980), we define the aggregate index k_i by

$$\log(\bar{x}_i/k_i) \equiv \sum_h x_{hi} \log(x_{hi}/k_{hi}) \qquad (6)$$

where $\bar{x_i}$ is the average expenditure per consumer in aggregate i. The aggregate k_i is viewed as a distributional entropy index, measuring the distribution of consumer characteristics withing aggregate i. Aside from the problem of determining k_i, (3) is applicable to aggregate data and will, in principle, yield estimates of the parameters of the individual consumer's demand system. In most applications of AIDS where aggregation has been considered explicitly, k_i has been assumed constant over i or uncorrelated with $\bar{x}_i$ or p_{ji} due to lack of appropriate data (Deaton and Muellbauer (1980), Veall and Zimmermann (1986)).[4]

Assuming that ε_{hki} is homoskedastic implies that $\bar{\varepsilon}_{ki}$ is heteroskedastic, since

$$V(\bar{\varepsilon}_{ki}) = V\left[\sum_h \frac{x_{hi}}{\sum_h x_{hi}} \varepsilon_{hki} \right] = \sum_h \left[\frac{x_{hi}}{X_i} \right]^2 \sigma_k^2 \qquad (7)$$

where $V(\varepsilon_{hki}) = \sigma_k^2$ and $X_i = \sum_h x_{hi}$. The degree of heteroskedasticity depends on the unknown expenditure of alcohol distribution in the i-th aggregate. Note that $\bar{\varepsilon}_{ki}$ is homoskedastic in general only if x_{hi}/X_i is independent of i, which will occur if the distribution of income is constant across aggregates.

The heteroskedaticity cannot be modelled since it is inherently unobservable.[5] However by adapting a procedure due to White (1980), it is possible to calculate a heteroskedaticity-consistent variance-covariance matrix for the parameter estimates.

3. A Heteroskedasticity-Consistent Variance-Covariance Matrix for AIDS

We begin by using a fundamental statistical relation: the asymptotic variance-covariance matrix of a maximum likelihood estimator is equal to the inverse of the variance-covariance matrix of the gradient of the likelihood function (Kendall and Stuart (1967), Vol. II, p. 9, Berndt, Hall, Hall and Hausman (1974), p. 657).

Suppose we write the system of equations (1) in the form

$$\underline{y} = f(\underline{z}, \underline{\beta}) + \underline{\bar{\varepsilon}} \qquad (8)$$

where $\underline{y}$ is the vector of left-hand side variables, $\underline{z}$ is the vector of right-hand side variables, and $\underline{\bar{\varepsilon}}$ is the stacked vector of error terms $\bar{\varepsilon}_{ki}$ stacked first over the aggregates and then over equations. From (7),

$$\text{var}\,(\underline{\bar{\varepsilon}}_{ki}) = E(\underline{\bar{\varepsilon}}_{ki})^2 = \{ \sum_h \left[\frac{x_{hi}}{X_i} \right]^2 \} \cdot \sigma_k^2 = \underline{\lambda}_i \cdot \sigma_k^2 \qquad (9)$$

where $\underline{\lambda}_i$ is a vector of data for aggregate i, each element at time t corresponding to $\sum_h \left[\frac{x_{hit}}{X_{it}} \right]^2$.

Similarly,

$$\text{cov}(\underline{\bar{\varepsilon}}_{ki}\underline{\bar{\varepsilon}}_{\ell i}) = \underline{\lambda}_i \, \sigma_{k\ell} \qquad (10)$$

Hence

$$V(\underline{\bar{\varepsilon}}) = E(\underline{\bar{\varepsilon}}\ \underline{\bar{\varepsilon}}') = \begin{bmatrix} \sigma_{11} & \cdots\cdots & \sigma_{1K} \\ \sigma_{21} & & \\ \vdots & & \\ \sigma_{2K} & \cdots\cdots & \sigma_{KK} \end{bmatrix} \bullet \underline{\lambda} = \Sigma \bullet \underline{\lambda} \qquad (11)$$

with $\underline{\lambda} = \begin{bmatrix} \underline{\lambda}_1 \\ \underline{\lambda}_2 \\ \vdots \\ \underline{\lambda}_I \end{bmatrix}$, where I is the number of aggregates in the sample.

The gradient of the likelihood function for the system of equations (8) when the heteroskedasticity of $\underline{\bar{\varepsilon}}$ is ignored can be written in the form (Berndt, Hall, Hall and Hausman (1974))

$$q = \left(\frac{\partial f}{\partial \underline{\beta}} \right)' (\Sigma^{-1} \bullet I)\ \underline{\bar{\varepsilon}} \qquad (12)$$

The variance-covariance matrix of the gradient is given by

$$\begin{aligned} E(qq') &= \left(\frac{\partial f}{\partial \underline{\beta}} \right)' (\Sigma^{-1} \bullet I)\ [E(\underline{\bar{\varepsilon}}\ \ \underline{\bar{\varepsilon}}')](\Sigma^{-1} \bullet I)\left(\frac{\partial f}{\partial \underline{\beta}} \right) \\ &= \left(\frac{\partial f}{\partial \underline{\beta}} \right)' (\Sigma^{-1} \bullet I)\ [\Sigma \bullet \underline{\lambda}](\Sigma^{-1} \bullet I)\left(\frac{\partial f}{\partial \underline{\beta}} \right) \\ &= \left(\frac{\partial f}{\partial \underline{\beta}} \right)' (\Sigma^{-1} \bullet \underline{\lambda})\left(\frac{\partial f}{\partial \underline{\beta}} \right) \qquad (13) \end{aligned}$$

The asymptotic variance-covariance matrix of $\hat{\underline{\beta}}$, the estimate of $\underline{\beta}$ calculated by the maximum likelihood (ML) method but assuming $\underline{\lambda} = I$ when forming the likelihood function, is given by

$$V(\hat{\underline{\beta}}) = \left\{\left(\frac{\partial f}{\partial \underline{\beta}} \right)' (\Sigma^{-1} \bullet \underline{\lambda}) \left(\frac{\partial f}{\partial \underline{\beta}} \right)\right\}^{-1} \qquad (14)$$

In order to find a consistent estimate of $V(\hat{\underline{\beta}})$ we first consider the squared residual $\bar{e}_{ki}^2$ from ML estimation when heteroskedasicity is ignored. This residual will provide the key to obtaining a consistent estimate of $V(\hat{\underline{\beta}})$.

Normalizing λ_{11} (the first observation in the first aggregate) to be unity, an estimation of λ_{it}, i, t not both = 1, can be obtained as

$$\hat{\lambda}_{it}^{k} = \frac{\bar{e}_{kit}^2}{\bar{e}_{k11}^2} \qquad k = 1, \ldots K \qquad (15)$$

There is one estimate of λ_{it} from each equation in the system.[6] A natural single estimate of λ_{it} is

$$\hat{\lambda}_{it} = \frac{1}{K} \sum_{k} \hat{\lambda}_{it}^{k} \qquad (16)$$

From White's (1980) extension of his Theorem 1 to the non-linear case (p. 821), it is clear that utilizing $\hat{\lambda}_{it}$ along with consistent estimates of $\frac{\partial f}{\partial \underline{\beta}}$ and Σ will yield a consistent estimate of $V(\hat{\underline{\beta}})$.

The simplest way of completing this estimation problem is to use a standard systems estimation computer program to find the maximum likelihood estimator whose variance-covariance matrix is a consistent estimate of the inverse of (13). Let $\underline{\theta}$ be a transformation of both sides of equation (8) which results in the gradient

$$\hat{q} = \left(\frac{\partial f}{\partial \underline{\beta}} \right)' (I \otimes \underline{\theta}')(\Sigma^{-1} \otimes I)\, \bar{\underline{u}} \qquad (17)$$

where $\bar{\underline{u}}$ is a transformed error vector with variance-covariance matrix assumed by the computer program to be $\Sigma \otimes I$. Then the output of the computer program will contain the following estimate of the variance-covariance matrix of the gradient:

$$\hat{E}(\hat{q}\hat{q}') = \left(\frac{\partial \hat{f}}{\partial \underline{\beta}} \right)' (I \otimes \underline{\theta}')(\hat{\Sigma}^{-1} \otimes I)(\hat{\Sigma} \otimes I)(\hat{\Sigma}^{-1} \otimes I)(I \otimes \underline{\theta}) \left(\frac{\partial \hat{f}}{\partial \underline{\beta}} \right)$$

$$= \left(\frac{\partial \hat{f}}{\partial \underline{\beta}} \right)' [\hat{\Sigma}^{-1} \otimes \underline{\theta}'\underline{\theta}] \left(\frac{\partial \hat{f}}{\partial \underline{\beta}} \right) \qquad (18)$$

Comparing (18) with (13) it is apparent that setting $\underline{\theta}'\underline{\theta} = \underline{\lambda}$ will solve our problem, since both Σ and $\frac{\partial f}{\partial \underline{\beta}}$ will be estimated consistently.[7]

FOOTNOTES

1. See for example, Anderson and Blundell (1983), and Veall and Zimmermann (1986).

2. Concavity of the expenditure function in prices cannot be imposed through linear restrictions.

3. AIDS satisfies Muellbauer's (1975, 1976) exact aggregation conditions (see below) and the UC and NT properties specified by Freixas and Mas-Colell (1987) as required for the Weak Axiom of Revealed Preference to hold in the aggregate.

4. An exception is Fuss and Waverman (1987) who explicitly model the possible variation in k_i.

5. In the unlikely case that income is distributed uniformly within each aggregate $x_{hi}/X_i = \frac{1}{H_i}$, where H_i is the number of consumers in the ith aggregate, which is often observable.

6. Note that the covariance term (10) does not provide additional information. Denote $\hat{\lambda}_{it}^{k\ell}$ to be the estimate of λ_{it} obtained from using (10). Then it can easily be shown that $(\hat{\lambda}_{it}^{k\ell})^2 = \hat{\lambda}_{it}^{k} \cdot \hat{\lambda}_{it}^{\ell}$, $k \neq \ell$.

7. Note that setting $\underline{\theta}'\underline{\theta} = \underline{\lambda}$ is equivalent to multiplying both sides of the system of equations(8) for each data point by $\sqrt{\lambda_{it}}$.

REFERENCES

Anderson, G. and R. Blundell (1983), "Testing Restrictions in a Flexible Dynamic Demand System: An Application to Consumers' Expenditure in Canada", Review of Economic Studies, July, pp. 397-410.

Berndt, E., B. Hall, R. Hall and J. Hausman (1974), "Estimation and Inference in Non-Linear Structural Models", Annals of Economic and Social Measurement, 3, pp. 653-666.

Deaton, A. and J. Muellbauer (1980), "An Almost Ideal Demand System", American Economic Review, 70, pp. 312-326.

Fuss M. and L. Waverman (1987), "The Demand for Alcoholic Beverages in Canada: An Application of the Almost Ideal Demand System", Institute for Policy Analysis Working Paper No. 8709, University of Toronto, Toronto, Canada, June.

Kendall, M. and A. Stuart (1967), The Advanced Theory of Statistics, vol. II, Griffen, London, England.

Muellbauer, J. (1975), "Aggregation, Income Distribution and Consumer Demand", Review of Economic Studies, 62, pp. 525-543.

Muellbauer, J. (1976), "Community Preferences and the Representative Consumer", Econometrica, 44, pp.979-999.

Veall, M. and K. Zimmermann (1986), "A Monthly Dynamic Consumer Expenditure System for Germany with Different Kinds of Households" The Review of Economics and Statistics, May, pp. 256-264.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", Econometrica, May, pp. 817-838.