Sankhyā : The Indian Journal of Statistics
1984, Volume **46**, Series B, Pt. 2, pp. 228–246.

HOUSEHOLD SURVEYS AS A DATA BASE FOR THE ANALYSIS OF OPTIMALITY AND DISEQUILIBRIUM

By ANGUS DEATON

Princeton University

SUMMARY. The usefulness of household budget surveys is explored with respect to examples involving optimality and disequilibrium. The first example shows how an integrated set of commodity demand and labour supply functions can be used both to implement optimal tax formulae and to test the hypothesis of labour market equilibrium. The second example concerns the appropriate empirical specification of demand functions when food is allocated through "Fair-price" shops, as for example, with sugar in India.

0. Introduction

Household surveys are a near universal data base, available in some form in virtually all countries, developed or undeveloped, market or centrally planned. Originally conceived in 19th century Europe as a vehicle for exposing and publicising the poverty and living conditions of the working class, they remain important today as data sources, not only for the identification and measurement of poverty, but also for more general welfare analysis, for the calculation of index numbers, for forecasting and planning, and for the evaluation and construction of policy. In the developing world, counting the poor is perhaps the most important official use of household surveys, and the results are widely used in the discussion and determination of both national and international aid policies. In developed countries, the central purpose of household surveys is often seen to be the provision of weights for the consumer price index. As inflation has become a central and dominant target of government policy, so has the construction of the consumer price indexes become an important political issue rather than an arcane statistical one. But from the point of view of academic research, the major development over the last decade has been the increasing availability for analytical work, not only of the published survey reports with their summary statistics and cross-tabulations, but of the individual household files. These data

Prepared for the IXth International Conference of Applied Econometrics VIth International Conference on the Econometric Modelling of the Socialist Economies, Budapest, 29th March—1st April 1982. I should like to thank the SSRC for financial support under grant HR 7637, "The Economics and Econometrics of Consumer Behaviour".

AMS (1980) *subject classification :* 90A12, 90A14.

Key words and phrases : Household surveys : Optimal tax ; Labour market disequilibrium : Fair-price shops : Demand functions.

sets are very large and typically very rich. The usual sample size is between 5,000 and 10,000 households and each household will typically provide several hundred pieces of information. Much can therefore be learnt from the analysis of such data and although a great deal of work has been done and is being done, I believe that these data sources are still very greatly under-used, a situation in sharp contrast to the amount of attention given by econometricians to the very short time-series data of macroeconomic aggregates.

The standard econometric exercise associated with household surveys is the estimation of Engel curves, usually in conjunction with the measurement of the effect on family demand patterns of demographic composition, particularly the number, age and sex of children. The classic study of this type is probably still the rather early one of Prais and Houthakker (1955), although only aggregated data were available at that time. More recently, household surveys have been widely used in the analysis of the labour supply of both primary and secondary workers and this literature has been a continuing source of outstanding applied work and econometric innovation; for a survey, see Heckman, Killingsworth and MaCurdy (1979). In this paper, I discuss three examples which are rather different from these standard ones and which will I hope illustrate both the topics of this conference and the wide range of applicability of even a single cross-sectional household survey. All three examples contain some element of disequilibrium analysis, in labour markets or in goods markets. Example 1, however, is concerned with optimality in the specific context of optimal taxation. I discuss the main empirical ingredients that contribute to the nature of an optimal tax solution and show how many of them can be investigated on household survey data. Example 2 turns explicitly to issues of disequilibrium in the labour market and I review some work designed to test whether or not a given sample of consumers chooses their hours of work voluntarily. Example 3 concerns disequilibrium in the goods market, specifically in those cases where the government interferes in the market with procurement and distribution schemes. I shall be particularly interested in how demand analysis should be conducted in such circumstances and in how government policy on ration levels and subsidy prices affects the free market demand for the goods involved.

1. The calculation of optimal taxation

The modern theory of optimal taxation, although tracing its roots back to Ramsey's (1927) famous contribution, only really developed rapidly after the publication of Mirrlees (1971) and Diamond and Mirrlees (1971). In

spite of this relative youth, the subject now seems to be fairly well understood, at least in outline. Since lack of the appropriate information (and the impossibility of obtaining it) preclude lump-sum non-distortionary taxation, the government must rely on distortionary taxes to finance public goods, its own administrative expenses, the deficits of nationalised industries, and any transfer payments thought to be desirable in the interests of equity. It has long been realised that the solution of such second best problems requires a careful and explicit formulation of the problem; standard economic intuition based on first-best solutions is frequently very misleading. The main elements of a formulation are usually as follows, see Mirrlees (1978) and Atkinson and Stiglitz (1980, lectures 11–13), for overviews :

(a) A social welfare function which summarises the government's preferences over alternative distributions of welfare between individuals.

(b) Private preferences of individuals which govern individual behaviour. They also determine individual welfare levels up to a monotone increasing transformation. Hence, the specification of individual preferences or utility functions determines not only how agents respond to government policy (e.g. tax changes) but what the welfare consequences of such policies will be.

(c) The objective determinants of the distribution of welfare independent of government actions, i.e. the endowments of individuals. In many of the models, individuals are identical except that they have differing 'ability' endowments, hence, in a market enviroment they have different wage rates.

(d) The instruments available to the government and the constraints upon their use. Linear or non-linear taxes on income may be allowed; uniform lump-sum benefits may be paid to all agents; taxes can be levied on commodities either at a uniform rate or in a way which discriminates against particular goods; the levels of provision of public goods can be determined, and it may even be possible to set non-linear tax rates for goods, for example if resale is impossible (e.g. electricity) or by imposing quotas. All of this activity is subject to the constraints set by the economy's productive technology, most usually analysed through the government's budget constraints.

Clearly (a) is determined by the policy maker's objectives, while (d) is a matter of defining the problem to be solved. The distribution of wage rates over individuals can be estimated from a household survey, provided the truncation of the distribution due to the non-observability of wage rates for those who do not work is explicitly recognised and taken into account, see e.g. Gronau (1974). However, the major contribution of household

survey data is in the determination of individual preferences and the measurement of behavioural responses. Four main research areas are of importance.

(a) *Testing for separability.* Atkinson and Stiglitz (1976) showed that in the standard optimal tax model with an optimal non-linear income tax, differential commodity taxation is unnecessary if preferences are weakly separable between leisure and goods. This result extends to any separable group of commodities, all of which should be optimally taxed at the same rate, see Mirrlees (1978). The empirical counterpart of goods being separable from leisure is that there should exist demand functions for goods in terms of total goods expenditure and prices alone and such conditions can readily be tested under various assumptions, see below. And when weak separability does not hold, other separability concepts, particularly implicit and quasi-separability, have direct consequences for the progressivity or regressivity of differential commodity taxation, Deaton (1981a). Separability conditions of various kinds also play a crucial role in the rules for optimally allocating public goods, see Lau, Sheshinski and Stiglitz (1978) and Jewitt (1982).

(b) *The shapes of Engel curves.* In Deaton (1979) I showed that the Atkinson and Stiglitz result on uniform commodity taxation extends to the case of an optimal linear income tax provided that the relationships between commodity demands and total commodity expenditures are linear. For non-linear Engel curves there will typically be scope for improving distribution by setting taxes which treat luxuries and necessities differently (note that this does not necessarily imply taxing luxuries at a higher rate). Determining the shape of Engel curves is, of course, the traditional province of household budget analysis.

(c) *Price responses.* Many optimal tax rules embody a balance between equity and efficiency. Goods tend to be more highly taxed if they are price inelastic and more highly taxed if they are more luxurious, two criteria which are frequently in conflict. In Deaton (1981a), I showed that many commodity tax rules can be characterized given knowledge of the dual of the Slutsky Substitution responses, the Antonelli matrix. This matrix gives the utility compensated derivatives with respect to quantity changes of the shadow prices of individual goods and, in commodity tax theory, it is the effects of changes in hours worked on the shadow prices of goods which are crucial. However, estimation of these effects in general requires the estimation of price elasticities and this requires more than one household survey, or even better, household panel data. Such data is much less widely available than the single cross-sections discussed in this paper.

(d) *Labour supply.* Income tax formulae depend critically on the effects on labour supply of changes in both wages and benefits. Substitution effects determine the distortion caused by the tax and limit the effectiveness of high marginal rates in generating revenue for redistribution or government expenditure. Income effects similarly limit the scope of redistribution through uniformly distributed benefits since while these benefit the poorest they may reduce labour supply of all workers and hence reduce income tax revenue. Estimating the size of these effects for different groups of workers has been a major focus of attention in the American literature cited above, not only using household surveys, but also using panel data and specially designed experiments.

In this paper, I should like to first discuss an attempt to test for separability and secondly to present some simple results on how labour supply responses determine an optimal income tax scheme. Fuller details are reported in Deaton (1982) and (1981b).

An extremely convenient form of preferences for cross-section analysis of both labour supply and commodity demands is Muellbauer's (1981) extension of the Gorman polar form. This expresses tastes through the cost function $c(u, w, p)$, the minimum cost of attaining utility u at wage rate w and commodity price vector p. The Muellbauer form is

$$c(u, w, p) = \{a(p)\}^{1-\delta} w^{\delta} u + b(p)w + d(p) \qquad \ldots \quad (1)$$

where δ is a parameter $0 < \delta < 1$ which is the marginal propensity to spend on leisure out of transfer income, and $a(p)$ and $d(p)$ are linearly homogeneous functions of p while $b(p)$ is homogeneous of degree zero. A utility maximizing consumer will have a value of costs equal to his or her total endowment $wT+\mu$, for time endowment T and transfer income μ. Commodity demands and income supply functions can be obtained from (1) in the usual way, yielding

$$p_i q_i = \gamma_i + \beta_i w + \alpha_i \mu \qquad \ldots \quad (2)$$

$$\mu + wl = \gamma_0 + \beta_0 w + \alpha_0 \mu \qquad \ldots \quad (3)$$

where q_i is the quantity purchased of good i, l is labour supply so that $\mu+wl$ is total income, and the p's, β's and γ's are related to the original $a(p)$, $b(p)$ and $d(p)$ functions, their partial derivatives, T and δ. Note in particular that $\gamma_0 = \Sigma\gamma_i$, $\beta_0 = \Sigma\beta_i$ and $\alpha_0 = \Sigma\alpha_i$ so that the budget constraint is satisfied, and that $\alpha_0 = (1-\delta)$ so that, as stated, δ represents the reduction in earned income wl consequent on a unit increase in transfer income μ. This parameter δ will represent the crucial disincentive measure when I come to look at optimal taxation.

Equations (2) and (3) cannot be expected to hold exactly since there will be considerable variation in tastes from individual to individual. However, it is sensible to interpret the equations as conditional expectations of $p_i q_i$ and $\mu + wl$ on w and μ in the population sampled in the household survey. The parameters can then be estimated by multiple regression in the usual way. Table 1, reproduced from Deaton (1982), gives the estimates for a subsample of 1617 households from the British Family Expenditure Survey

TABLE 1. LINEAR MODEL WITHOUT SEPARABILITY

	Constant	*NCH*	*NEARN*	*OWN*	μ	w	R^2	*Ratio*
1. Food	4.58	1.56	1.37	0.07	0.10	2.86	0.278	30
	(8.8)	(16.9)	(5.0)	(0.3)	(8.4)	(6.6)		
2. Drink	0.32	0.03	0.88	−0.82	0.05	0.79	0.161	15
	(0.9)	(0.5)	(4.5)	(−4.8)	(6.6)	(2.6)		
3. Tobacco	1.76	0.05	0.45	−0.90	0.02	−0.55	0.164	−31
	(8.9)	(1.5)	(4.3)	(−9.8)	(4.2)	(−3.4)		
4. Clothing	0.14	0.54	0.72	0.15	0.09	1.97	0.099	21
	(0.2)	(4.4)	(2.0)	(0.5)	(6.3)	(3.4)		
5. Durable household	0.88	−0.15	0.16	1.11	0.04	2.64	0.011	70
goods	(0.6)	(−0.6)	(0.2)	(1.7)	(1.3)	(2.3)		
6. Other goods	0.90	0.19	0.11	0.26	0.05	1.94	0.076	36
	(2.0)	(2.4)	(0.5)	(1.2)	(5.6)	(5.3)		
7. Transport&Vehicles	1.56	−0.45	0.20	1.24	0.13	5.02	0.099	40
	(1.6)	(−2.5)	(0.4)	(2.7)	(5.8)	(6.0)		
8. Services	−1.85	−0.39	0.03	0.06	0.09	7.72	0.070	87
	(−1.6)	(−1.9)	(0.0)	(0.1)	(3.5)	(8.1)		
9. Composite	−5.40	−0.85	−3.94	−1.27	0.43	15.9	0.100	37
	(−2.1)	(−1.9)	(−3.0)	(−1.1)	(7.8)	(7.6)		
10. Total Income	2.89	0.53	−0.02	−0.11	1.00	38.3		38
	(4.4)	(4.6)	(−0.1)	(−0.14)	(70.5)	(70.9)		
	$W = 28.6$		$W_{3,8} = 8.1$		$2 \log L = -88647$			

of 1973. Households are restricted to those with male heads in employment who are aged between 18 and 64 and whose normal wage was between £0·85 and £3·00 per hour. The selection is more fully discussed in the reference given, see also Atkinson and Stern (1981). The additional variables in the regression, NCH, NEARN and OWN represent household characteristics variables and are the number of children, the number of earners and a (0, 1) dummy indicating whether or not the household owned its own home.

These regressions are seriously inadequate in many respects. The fits are not good, even for this sort of work, and there are indications that the linear functional form, although convenient, is not very accurate. Note also that the δ parameter is zero, indicating no disincentive effect of transfer income on labour supply, while at the same time the coefficient on w in the income equation indicates a backward sloping labour supply curve for the majority of workers. The conjunction of these two findings are inconsistent with standard theory, and with the model (1) above. However, before such findings could be taken seriously much more work needs to be done and similar early findings in the U.S. were later found to be artefacts of mis-measurement and poor economic and econometric specification, see again Heckman, Killingsworth and MaCurdy (1979).

Even so, these results can still be used to illustrate the testing for separability for the commodity demands. Now the goods form a separable group from leisure if and only if the compensated cross-price derivatives $s_{i0} = \partial q_i/\partial w$ with utility held constant take the form—see e.g. Deaton and Muellbauer (1980, 128-9)

$$s_{i0} \propto \partial q_i/\partial \mu \qquad \ldots \quad (4)$$

where the constant of proportionality is independent of i. With preferences (1), this holds if and only if $b(p)$ is independent of p so that in equations (2) and (3)

$$\beta_i = \alpha_i \tau, \quad i = 1, \ldots, n, \qquad \ldots \quad (5)$$

where τ, independent of i, is equal to $T-b$. The ratios of β_i and α_i are given in the last column of Table 1; under separability these ratios should be identical and should be equal to the number of hours per week which the individual has available for work and play, excluding subsistence leisure. Inspection would suggest that separability is not strikingly supported on this evidence. A formal test can be constructed on the Wald principle by testing the eight restrictions

$$\beta_i \alpha_9 - \beta_9 \alpha_i = 0, \quad i = 1, \ldots, 8. \qquad \ldots \quad (6)$$

The test statistic is 28·6 which, if separability is correct, is asymptotically a random drawing from a χ^2_8 distribution; hence the formal results support the visual impression. Note however that most of the problems are associated with goods 3 and 8; if these are excluded, the Wald statistic for the other goods falls to 8·1 which is within the acceptable range. There must also be some suspicion that the wage coefficient in the tobacco equation is biased by the omission both of other conditioning factors (occupation, education, etc.) and

of any correction for the censoring of the data at zero. But taken at face value, the results suggest that food, drink, clothing, durable household goods, other goods, transport and vehicles and the composite residual might all be subject to a uniform tax. The extent of discrimination against (or in favour) of tobacco and services is not usefully calculated without a better labour supply equation. Indeed all these results might well turn out quite differently once better models are estimated and better techniques are used.

The Muellbauer cost function is also extremely analytically convenient in the context of an optimal linear income tax. If we allow the government to make an amount μ available to each individual and to tax labour income at a constant marginal rate τ, the optimal tax problem can be written

$$\max W = V(u^1, u^2, \ldots, u^N) \qquad \ldots \quad (7)$$

Subject to individual preferences

$$c(u^h, w^h(1-\tau), p) = w^h(1-\tau)T+\mu \qquad \ldots \quad (8)$$

and the government budget constraint

$$\sum_h \tau(w^h l^h) = R+N\mu, \qquad \ldots \quad (9)$$

where superscripts h denote the N individuals and R is revenue not redistributed. Under separability and linear Engel curves, social welfare would not be improved by allowing discriminating commodity taxation but even without these conditions, the problem describes the case where only an income tax is being considered. In Deaton (1981b) I show that if the government is endowed with a rank order social welfare function and if consumers have Muellbauer tastes, then the optimal tax rate τ can be calculated in the explicit form

$$\tau = \frac{1}{1-\delta} - \frac{\delta}{2\phi(1-\delta)^2}\left\{\left(1+\frac{4(1-\delta)\phi}{\delta}\right)^{1/2}-1\right\} \qquad \ldots \quad (10)$$

where $\phi = I/(1-\sigma-r)$. The parameters are defined as follows. Let $\bar{w}(T-b)$ be "potential per capita GNP"; $(T-b)$ is the maximum number of hours the consumer is capable of working if only subsistence leisure is taken, while $\bar{w}$ is the average wage. The quantity σ is consumers' subsistence expenditure on goods, $d(p)$ in (1) as a fraction of potential GDP, i.e.

$$\sigma = d(p)/\bar{w}(T-b). \qquad \ldots \quad (11)$$

The government revenue requirement R is expressed in similar form via

$$r = R/\bar{w}(T-b). \qquad \ldots \quad (12)$$

Since subsistence and government expenditure are prior claims on resources, the fraction $(1-\sigma-r)$ is effectively available for redistribution. I is a measure of pre-tax inequality as perceived by the government and is given by

$$I = (\bar{w}-w^*)/\bar{w} \qquad \ldots \quad (13)$$

for average wage $\bar{w}$ and $w_{min} \leqslant w^* \leqslant \bar{w}$. The position of w^* is determined by the degree of inequality aversion in the social welfare function. If the government is very averse to inequality (Rawlsian) $w^* = w_{min}$ and I takes its maximum value; if the government is blind to inequality, $w^* = \bar{w}$ and I is always zero. Since $(1-\sigma-r)$ is the fraction of potential GNP available for redistribution, the parameter $\phi = I/(1-\sigma-r)$ can be usefully thought of as the 'effective' degree of pre-tax inequality. (Note finally that (10) does not hold for $\delta = 1$ or 0 or $I = 0$: for $\delta = 1$, $\tau = \phi$; for $\delta = 0$, $\tau = 1$; for $I = 0$, $\tau = 0$; for $\sigma = 0$ and $I = 0$, the tax rate is indeterminate).

Figure 1 gives a complete characterization of τ in terms of its two determinants ϕ and δ. Increases in ϕ (i.e. increases in inequality, subsistence, or government revenue) always increase τ; increases in δ, the disincentive parameter, generally decrease τ except for high values of ϕ in which case τ falls and then rises with δ. In all cases, as the disincentives become large, τ tends to ϕ in which case we have the simple result that the optimal income tax rate is the effective degree of pre-tax inequality.

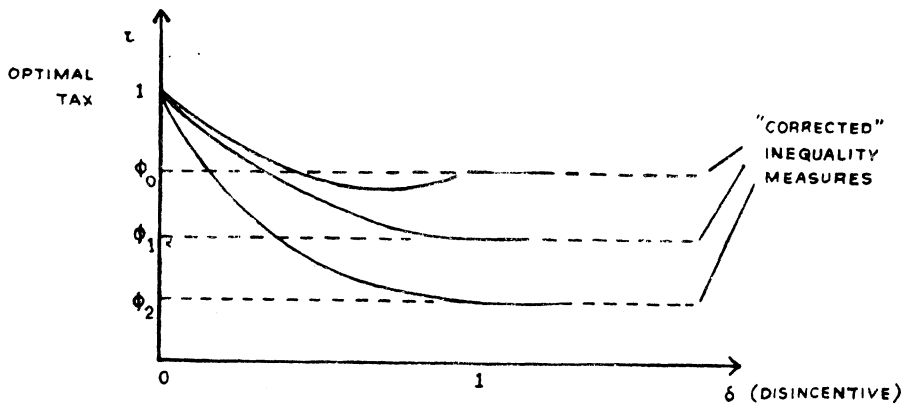

Figure 1

Application of the formula (10) requires (apart from a suspension of disbelief in the model's simplified assumptions) knowledge of only four parameters. Inequality reflects the government's preferences as much as the actual distribution of wages, and r is known, at least in principle, provided a plausible figure for potential GDP can be estimated. There is a large

literature on using household surveys to estimate poverty lines and subsistence needs, see e.g. Sen (1981, p. 24) for a partial bibliography of the basic needs approach. The remaining parameters are δ and $(T-b)$, if potential GDP is to be estimated, and both can be estimated from equations such as (3) on a single household survey. But this is more easily said than done. As we saw in Table 1, δ is estimated to be zero (an optimal tax for the U.K. of 1 !) and the implicit value of $(T-b)$ is 39 which gives potential GDP marginally lower than actual GDP ! Of course, as discussed above, more sophisticated work may well yield better estimates and ones more in accord with the theory underlying the model. Indeed, time-series estimates of similar models for the U.S., e.g. Abbott and Ashenfelter (1976) and Phlips (1978) give estimates of δ of 0·12, although the figures for $(T-b)$ are still much too low. However there is a general problem in this type of work to which I will return at several points in this paper. Matching theory with empirical evidence in order to implement the former is always likely to be difficult. The idealized model concepts of the theory are often hard to match to the actual concepts of measurement and the attempt to do so tends to throw up a host of econometric difficulties. While household surveys contain virtually everything that is necessary for the analysis, making the correct inferences about the parameters of interest, or ultimately even rejecting the theory, is a research programme of challenging proportions on which only the first few steps have been taken.

2. Testing for labour market disequilibrium

One of the possible sources of difficulty in labour supply models such as that in the previous example is the doubtful nature of the assumption that workers are indeed free to choose their hours at a fixed, predetermined wage rate. Many economists regard it as obvious that at least some workers (the unemployed) do not choose the amount of leisure they consume, and would extend the argument to many employees who work an institutionally determined week. Indeed whether or not unemployment is voluntary has become an important issue in much recent macroeconomic debate. However, it is noticeable in Table 1 that the coefficient on wages in the income regression is 38·3 which is below but not far from hours worked over the sample (42·8). This is exactly what one would expect if hours were exogenously determined but wages were measured with error. These sort of considerations led Ashenfelter (1980) to develop a model based on the separable version of (3) which allowed some consumers to be quantity constrained and others to be making free choices. By aggregating over consumers, Ashenfelter develops a model of average demand and hours which depends not only on prices and

wages but also on the level of unemployment. Applied to U.S. time-series data, there is considerable evidence supporting the existence of some quantity restrictions on hours.

In a single cross-section, the theory of rationing can be used to develop demand functions with and without quantity constraints, the comparison of which can in principle be used to test for whether or not the constraints are binding. To see how this can be set up, let $u = \nu(q_0, q)$ be the utility function defined on leisure q_0 and a vector of goods q. The unconstrained consumer maximizes $\nu(q_0, q)$ subject to the usual budget constraint $wq_0+p.q = wT+\mu$ giving demand functions for goods and leisure

$$q_i = g_i(wT+\mu, w, p) \qquad \ldots \quad (14)$$

$$q_0 = g_0(wT+\mu, w, p) \qquad \ldots \quad (15)$$

in the usual way. These are the general equations corresponding to (2) and (3) above. With a constraint in the labour market, there is the additional constraint $q_0 = \bar{q}_0$ say, so that the maximization problem may be written in the form

$$\underset{q}{\text{maximize}}\ u = \nu(\bar{q}_0, q) \qquad \ldots \quad (16)$$

$$\text{subject to } p.q = w(T-\bar{q}_0)+\mu = y \qquad \ldots \quad (17)$$

where y is total income. The solution for the commodity demands may now be written

$$q_1^* = g_i^*(y, p, \bar{q}_0). \qquad \ldots \quad (18)$$

Note that w does not appear (18) in except as part of y; $\bar{q}_0$ on the other hand is both part of y and has an independent role in the function. This arises because of the potential effects of $\bar{q}_0$ on the marginal rates of substitution between the elements of q in the utility function. If such effects are absent, that is if the goods are weakly separable from leisure, $\bar{q}_0$ is absent from the demands (18). This offers an alternative methodology to that discussed in Example 1 for testing weak separability between goods and leisure, a methodology which is appropriate under the maintained hypothesis of quantity constrained hours. (We are currently conducting experiments along these lines and finding very little evidence for the presence of $\bar{q}_0$ in the demands.)

Discrimination between (14) and (15) on the one hand and (18) on the other can proceed in a variety of ways. Perhaps the neatest is to realise that if $\bar{q}_0$ in (18) is set at the value which would have been chosen in any case, then the commodity demands must also be those which would have been freely chosen. In other words, substitution of (15) for $\bar{q}_0$ in (18) yields (14). [It is

also true that elimination of w between (14) and (15) yield (18).] These arguments show that (18) *is true whether there is a quantity restriction or not.* The difference between the two situations lies only in the determination of q_0; under free choice, it is given by (15), under quantity restrictions by some exogenous process ancillary to the model of individual behaviour. With a convenient choice of functional form, see Deaton and Muellbauer (1981) for the case in which (18) is linear, (18) can be directly estimated and the parameters will be consistently estimated under the hypothesis that $\bar{q}_0$ is determined outside the model, i.e. that it is constrained. Under free choice, however, the estimates will not in general be consistent because q_0 and q_i are simultaneously determined by (14) and (15). However, when q_0 is given by (15), w provides an instrument which can be used to give consistent estimates of (18) and a Wu (1973) or Hausman (1978) test can be applied straightforwardly to test for the presence of bias and hence for whether or not hours are freely chosen.

In Deaton (1982) I reported tests on the data used in Table 1 which were similar in spirit to those proposed above. However, I found that these data were unable to discriminate between the two models, each being equally well supported (or rejected). In the light of the previous discussion there are two reasons why such tests might fail, both of which apply in the current situation. The first is that the instrument has little or no explanatory power for q_0. One could of course interpret this as an implicit rejection of equation (15) and thus of the flexible hours model. However, if labour supply is genuinely inelastic, $\bar{q}_0$, although freely chosen, is essentially exogenous in (18), and no bias will occur. The second problem can occur even when (15) fits well and wages have a significant effect on hours. To illustrate, rewrite (14) and (15) in the linear form (2) and (3) as

$$z = Ax + aw + u \qquad \ldots \quad (19)$$

$$y = b'x + \beta w + u_0 \qquad \ldots \quad (20)$$

where x is the vector of exogenous variables in the regression, z is the vector of commodity expenditures, y is income $(w(T-q_0)+\mu)$, u is a vector of errors representing individual taste effects and (20) is the sum of the rows of (19) so that, in particular, $u_0 = \Sigma u_i$. Assume that the u_i have zero mean and constant covariance matrix Ω, typical element ω_{ij}, i.e.

$$\Omega = E(uu'). \qquad \ldots \quad (21)$$

The analogue of (18) is obtained by solving (20) for w and substituting in (19) to give

$$z = (A - \beta^{-1}ab')x + \beta^{-1}aw + (u - \beta^{-1}au_0). \qquad \ldots \quad (22)$$

The bias and inconsistency arise because w is correlated with u_0 and hence with the compound error. However w is uncorrelated with the error if

$$\frac{a_i}{\Sigma a_j} = \frac{\sum_k \omega_{ik}}{\sum_j \sum_k \omega_{jk}} \qquad \ldots \quad (23)$$

a condition which is by no means implausible if the effects of the shocks work in a similar way to wage changes. It is possible to construct stochastic preference models along the lines of Theil (1971) which yield this sort of result and in the particular data set underlying Table 1, the conditions (23) are approximately met.

Once again, it is clear that further work is necessary if this sort of testing is to be more successful. It may even be the case that the use of rationing theory is not a very productive approach given that it is only the behaviour of hours which discriminates between the models. If so, a more direct approach would be to model hours itself and attempt to discriminate between institutional determinants on the one hand and supply determinants on the other.

3. Disequilibrium in the goods market-fair price shops

In many countries of the world, particularly in poor countries, governments seek to influence the retail supplies of good by procurement and resale policies. These can take various forms and the one I am going to discuss here is that currently prevailing in India. Certain consumer necessities (rice, flour, sugar, kerosene, with some variation from state to state) are procured by the government from producers and resold in "fair-price" shops. Access to these shops is controlled by the issue of ration cards, so that each household (at least in urban areas) has a weekly entitlement depending on family size. A free market exists side by side with the fair price shops, usually legally (e.g. sugar) sometimes illegally (rice in Calcutta). Although the 'fair' price is always lower than the free market price, some goods, particularly rice, can be easily adulterated so that market equilibrium is attained through quality variation. For other goods, e.g. sugar, this does not take place, and there is a genuine element of subsidization, at least to those households who can obtain ration cards. In this example, I shall focus on sugar, although the model could easily be transplanted to other applications. I draw heavily on the paper by Chetty (1981), who also discusses the production side of the policy. Here I focus entirely on the implications for modelling individual household behaviour and on the derivation of free-market demand.

Figure 2 illustrates the situation for a household with fair-price shop quota z. This can be bought at price p_0 in the fair-price shop along CD until the quota OZ is attained. Subsequent purchases can be made in the free market at the higher price p_0 until the budget is exhausted at A. The budget line is thus the kinked line ACD. However, if the consumer can resell sugar in the free-market, the line CB is also attainable in which case the fair price ration is equivalent to a money subsidy of $(p_1-p_0)z$. Formally, if resale is allowed, the consumer has total resources x say, $p.q$ is spent on other goods, p_0z in the fair price shop, and a positive or negative amount $p_1(q_1-z)$ in the free market. The budget line AB is thus given by

$$p.q+p_1q_1 = x+(p_1-p_0)z \qquad \ldots \quad (24)$$

which makes the nature of the subsidy clear. If this is in fact the situation,

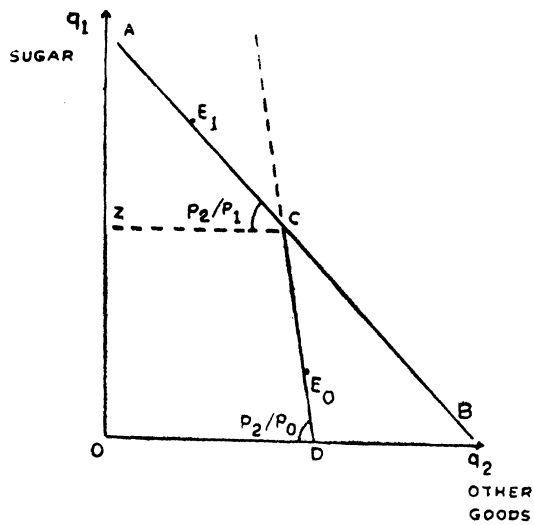

Figure 2 : The opportunity set for sugar purchases

it might reasonably be asked why the government would wish to operate such a scheme at all rather than simply give cash subsidies. One possibility is administrative; it may be very hard to set up a cash grant scheme with any chance of the money reaching the intended recipients. While grants in kind are inefficient and also are subject to administrative "loss" (perhaps 10% of the cereal crop in India vanishes in this way), the larger part does get through. A more convincing argument comes from taking uncertainty into account. If supplies fluctuate over time through harvest failure or other factors, and if it is recognised that administrative schemes are slow to respond (see Sen (1981, Chapters 6–9) for several examples), then a cash subsidy scheme is unlikely to protect consumers against fluctuations in prices. Fair-price

shops on the other hand, directly backed by buffer stocks, will continue to provide the ration quantities (although administrative "losses" may be larger) without any need for a policy change. Of course, only those with ration cards are so protected, and the government's actions may well exacerbate the price fluctuations in the free market so worsening the position of the uncovered consumers. Sen's documentation of the 1943 Bengal famine provides an excellent example; those affected were the rural population while the (pre-famine) urban population of Calcutta was largely protected by the procurement schemes then in operation.

Returning now to the case where resale is not allowed, we wish to be able to predict which consumers will buy less than the quota, which will buy exactly the quota, and which will buy in the free market. This case is of interest since, even with resale possibilities, the opportunities may not be very good especially given the time and trouble involved in negotiation over relatively small quantities. The direct analytical approach is to set up a utility function and to maximize it with respect to the kinked budget line ACD and to locate the equilibrium in one of the possible regimes, AC, e.g. at E_1, CD, e.g. at E_0, or at C itself. However it is simpler to proceed in another way which uses any prior information we might have on the nature of the demand function. Assume that for a consumer facing a *linear* budget constraint (e.g. AB) demand for sugar is given by the usual (Marshallian) demand function.

$$q_s = g(y, p_s, p) \qquad \ldots \quad (25)$$

where y is total outlay or income, p_s is the price of sugar and p is the vector of other prices. Consider the two possible situations corresponding to the budget lines AB and FD. Along FD, unlimited purchases at p_0 are permitted and sugar demand is

$$q_s^A = g(x, p_0, p). \qquad \ldots \quad (26)$$

Along AB, on the other hand, the subsidy $(p_1 - p_0)z$ is paid but the price of sugar is higher so that

$$q_s^B = g(x + (p_1 - p_0)z, p_1, p). \qquad \ldots \quad (27)$$

Given a specification of g (from a utility function or cost function) q_s^A and q_s^B can be calculated, and these can be used to determine demand along the kinked budget line ACD. Assume first that $q_s^A \leqslant z$, i.e. equilibrium on FD is along CD. Then by revealed preference, $q_s^B \leqslant z$ since points along AC to the left of C were attainable when q_s^A was chosen. Alternatively, since the budget set FCD is preferred to ACD, any choice which is optimal for FCD

and which remains feasible is optimal for ACD. Hence $q_s^A \leqslant z$ implies that q_s^A is optimal on ACD. Similarly $q_s^A \geqslant z$ implies that q_s^B is optimal on ACD. The remaining case is $q_s^A > z$ and $q_s^B < z$ which implies that C is optimal.

Hence, if household h has demand function g^h, total outlay x^h and ration level z^h, it buys less than its quota if

$$g^h(x^h, p_0, p) < z^h; \qquad \ldots \quad (28)$$

it buys from the free market if

$$g^h(x^h+(p_1-p_0)z^h, p_1, p) > z^h \qquad \ldots \quad (29)$$

and it buys exactly its quota if

$$g^h(x^h+(p_1-p_0)z^h, p_1, p) \leqslant z^h \leqslant g^h(x^h, p_0, p). \qquad \ldots \quad (30)$$

The 32nd round of the Indian National Sample Survey asked households not only how much of each commodity they purchased but also from which source, so that when these data become available in the next few years such models will be easily estimated. The simplest specification for g would be a linear one with taste differences represented by an additive unobservable error term. The resulting model is akin to the Tobit specification with censoring at z rather than at zero; only now there is a branch on either side of the censoring point. Figure 3 illustrates the Engel curve; price quantity curves are similar. The construction of the appropriate likelihood is straightforward.

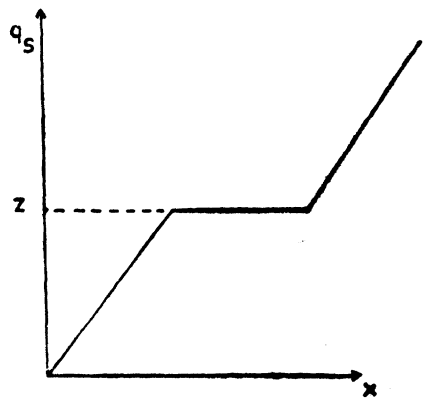

Figure 3 : Sugar demand as a function of total outlay

To analyse the free market demand it is natural to restrict attention to differences over households in income x, ignoring for the moment differences in tastes and quotas. In this case, provided sugar is a normal good, define x_T by

$$g(x_T+(p_1-p_0)z, p_1, p) = z \qquad \ldots \quad (31)$$

so that consumers with $x > x_T$ enter the free market. Hence, if $F(x; \theta)$ for parameters θ is the distribution function of x, percapita free market demand is given by

$$Q_F = \int_{x_T}^{\infty} \{g(x+(p_1-p_0)z, p_1, p)-z\}dF(x; \theta) \qquad \ldots \quad (32)$$

and this equation can be used to describe the relevant comparative statics. The particular points of interest here concern the operation not only of an intensive margin as individual free market purchasers buy more or less as prices, incomes and quotas change, but also of an extensive margin as new consumers enter or old consumers leave the free market. Consider, for example, $\partial Q_F/\partial z$, the effect on the free market demand of an increase in the ration. Differentiating (32) gives

$$\frac{\partial Q_F}{\partial z} = \int_{x_T}^{\infty} \left\{ \frac{\partial q_s}{\partial x} (p_1-p_0)-1 \right\} dF(x; \theta)$$

$$-\{g(x_T+(p_1-p_0)z, p_1, p_0)-z\}f(x_T, \theta)$$

which, by virtue of (31) is simply

$$\frac{\partial Q_F}{\partial z} = \int_{x_T}^{\infty} \left\{ \frac{\partial q_s}{\partial x} (p_1-p_0)-1 \right\} dF(x; \theta). \qquad \ldots \quad (33)$$

Since, at the extensive margin, consumers buy nothing in the free market, only the intensive margin is of importance. Chetty (1981), using a linear approximation, estimates this effect to be $-0{\cdot}84$, a number which is not obviously out of line with the theory. Similar derivatives can be given for price and income effects with changes in p_0, the fair-shop price, operating only via income effects. Note finally that this formulation suggests that free market demand is unlikely to be well approximated by a linear function, at least for all of its range, given the far from uniform distribution of income.

4. Conclusions

Although this paper is clearly very short on convincing empirical results, I hope that my three examples have illustrated the very great potential that exists for using household survey data for the analysis of important and interesting questions in economics. The problems which remain are the universal ones of applied econometrics, of adapting theoretical concepts and economic methodology to the appropriate interpretation of any body of data. But the data sources here are very rich and they are widely available. Their analysis remains a great challenge.

References

ABBOT, M. and ASHENFELTER, O. (1976): Labour supply, commodity demand, and the allocation of time. *Review of Economic Studies*, **43**, 389–411.

ASHENFELTER, O. (1980): Unemployment as disequilibrium in a model of aggregate labour supply. *Econometrica*, **48**, 547–64.

ATKINSON, A. B. and STERN N. (1981): On labour supply and commodity demands in Deaton, A. S. (ed), *Essays in the Theory and Measurement of Demand*, New York: Cambridge University Press.

ATKINSON, A. B. and STIGLITZ, J- E. (1976): The design of tax strcuture : direct versus indirect taxation. *Journal of Public Economics*, **6**, 55–75.

——— (1980): *Lectures on Public Economics*, McGraw-Hill.

CHETTY, V. K. (1981): Economics of price and distribtion controls—1 : the sugar Industry, Discussion Paper No. 8101, Indian Statistical Institute, New Delhi.

DEATON, A. S. (1979): Optimally uniform commodity taxes. *Economics Letters*, **2**, 357–361.

——— (1981): Optimal taxes and the structure of preferences, *Econometrica*, **49**, 1245–60.

——— (1982a): Model selection procedures, or does the consumption function exist ? in P. Corsi and G. Chow (eds), *The Reliability of Macroeconomics Models*, forthcoming. Wiley.

——— (1982b): An explicit solution to an optimal tax problem. Discussion Paper No. 81/112, Bristol University, Department of Economics.

DEATON, A. S. and MUELLBAUER J. (1980): *Economics and Consumer Behavior*, New York: Cambridge University Press.

——— (1981): Functional forms for labor supply and commodity demands with and without quantity restrictions. *Econometrica*, **49**, 1521–32.

DIAMOND, P. A. and MIRRLESS J. A. (1971): Optimal taxation and public production : I and II. *American Economic Review*. **61**, 8–27 and 261–78.

GRONAU, R. (1974) Wage comparisons—A selectivity bias, *Journal of Political Economy*. **82**, 1119–1143.

HAUSMAN, J. A. (1978): Specification tests in econometrics, *Econometrica*, **46**, 1251-72.

HECKMAN, J. J., KILLINGSWORTH M. R. and MACURDY (1979): Recent theoretical and empirical studies in labor supply: A partial survey, Mimeo.

JEWITT, I. (1982): Consumer preferences and the problem of public good supply, Discussion paper, University of Bristol, Department of Economics.

LAU, L. J., SHESHINSKI E. and STIGLITZ J. E. (1978): Efficiency in the optimum supply of public goods. *Econometrica*, **46**, 269–284.

MIRRLESS, J. A. (1971): An exploration in the theory of optimum income taxation. *Review of Economic Studies*, **38**, 175–208.

——— (1978) The theory of optimal taxation, in Arrow, K. J. and M. Intriligator (eds), *Handbook of Mathematical Economics*, North Holland, forthcoming.

MUELLBAUER, J. (1981): Linear aggregation in neoclassical labour supply. *Review of Economic Studies*, **48**, forthcoming.

PHLIPS, L. (1978): The demand for leisure and money. *Econometrica*, **46**, 1025–1043.

PRAIS, S. J. and HOUTHAKKER H. S. (1955): *The Analysis of Family Budgets*, Cambridge: Cambridge University Press. (Second edition 1971).

RAMSEY, F. (1927): A contribution to the theory of taxation. *Economic Journal*, **37**, 47–61.

SEN, A. K. (1981): *Poverty and Famines: An Essay on Entitlement and Deprivation*, Oxford University Press.

THEIL, H. (1971): An economic theory of the second moments of disturbances of behavioural equations. *American Economic Review*, **61**, 190–194.

WU, D-M. (1973): Alternative tests of independence between stochastic regressors and disturbances, *Econometrica*, **41**, 733–750.

Paper received: November, 1982.