Estimation of Household Demand Systems with Theoretically Compatible Engel Curves and Unit Value Specifications

Ian Crawford
François Laisney
Ian Preston

THE INSTITUTE FOR FISCAL STUDIES
WP02/17

Estimation of Household Demand Systems with Theoretically Compatible Engel Curves and Unit Value Specifications

Ian Crawford[*], François Laisney[†]& Ian Preston[‡]

June 2002

Abstract

We develop a method for estimation of price reactions using unit value data which exploits the implicit links between quantity and unit value choices. This allows us to combine appealing Engel curve specifications with a model of unit value determination in a way which is consistent with demand theory, unlike methods hitherto prominent in the literature. The method is applied to Czech data.

Key Words: Consumer demand, unit values

JEL Classification: D11, D12

Acknowledgements:
We are grateful for comments and advice from two anonymous referees, as well as from Richard Blundell, Fiona Coulter, Angus Deaton, Philippe De Vreyer, Bas Donkers, Vera Kameničkova, Colin Lawson, Arthur Lewbel, Thierry Magnac, Edmond Malinvaud, Costas Meghir, Daniel Munich, Jean-Marc Robin, David Ulph, Jiřì Večernik, Frank Windmeijer, and participants at conferences in Istanbul, Leuven, Mannheim, Prague and various seminars. This research was undertaken with support from two programmes of the European Commission: PHARE-ACE (project "Tax Policy During Economic Transition") and HCM (project "Microeconometrics of Public Policy Issues"). All errors remain our own.

[*]Institute for Fiscal Studies.
Email: `ian_crawford@ifs.org.uk`

[†]BETA-THEME, Université Louis Pasteur, Strasbourg and ZEW, Mannheim.
Email: `laisney@zew.de`

[‡]University College London and Institute for Fiscal Studies.
Email: `i.preston@ucl.ac.uk`

Executive Summary

One of the main difficulties in the estimation of demand systems using household data concerns the precise estimation of price reactions. The reason is that, whereas data on households normally exhibit considerable variation in expenditures, this is not typically the case for prices. Very often information about geographical variation in prices or variation over time within the period covered by one cross-section is lacking, so that prices are assumed uniform over all households of the same cross-section.

Data sets which contain information, not only on expenditures, but also on *quantities* consumed for a set of goods, offer interesting possibilities: this allows the computation of individual *unit values* for the spending of each household on any of these goods. It might be thought possible to model demand for these goods treating these unit values as prices. These would appear much more attractive for estimation purposes than aggregate prices, which are just averages that no household actually pays. Yet, since the goods are invariably subject to some degree of aggregation, it is undoubtedly true that much of the variation in unit values will actually result from household choice regarding the nature of the goods purchased.

We develop a method for estimation of price reactions using unit value data which exploits the implicit links between quantity and unit value choices and which builds on methods previously proposed in the demand literature. This allows us to combine appealing Engel curve specifications with a model of unit value determination in a way which is consistent with demand theory, unlike methods hitherto prominent in the literature.

We illustrate the technique with an application to Czech Family Budget Survey data.

1 Introduction

One of the main difficulties in the estimation of demand systems using household data concerns the precise estimation of price reactions. While requiring care in the treatment of endogeneity, income effects are more easily estimated. Yet unless, say, one is prepared to make strong assumptions on functional form which result in a connection between price and income effects, but which, if wrong, produce important biases in the estimation of price elasticities (see e.g. Deaton and Muellbauer, 1980), price effects are difficult to capture. The reason is that, whereas data on households normally exhibit considerable variation in expenditures, this is not typically the case for prices. Very often information about geographical variation in prices or variation over time within the period covered by one cross-section is lacking, so that prices are assumed uniform over all households of the same cross-section. Indeed most studies based on the Family Expenditure Survey for instance, a long series of cross-sections of UK households, have relied solely on year-to-year variation of prices under that assumption (see e.g. Banks, Blundell and Lewbel, 1996). In the absence of such long series, researchers have often resorted to combining a small number of cross-sections with aggregate time series data, the idea being basically to identify the income effects from the cross-sectional data and the price effects from the aggregate data (examples of studies relying on that strategy are Stone, 1954, Jorgenson, Lau and Stoker, 1982, and Nichèle and Robin, 1995). Note that Lewbel (1989) proposes an original approach to the identification and estimation of demand systems on pure cross-section data (without any price variation). Considering two levels of aggregation, under the assumption of homothetic weak separability Lewbel's approach imposes no restriction on functional form at the higher level. We shall draw parallels between his approach and ours later.

Data sets which contain information, not only on expenditures, but also on *quantities* consumed for a set of goods, offer interesting possibilities: this allows the computation of individual *unit values* for the spending of each household on any of these goods. It might be thought possible to model demand for these goods treating these unit values as prices. These would appear much more attractive for estimation purposes than aggregate prices, which are just averages that no household actually pays. Yet, since the goods are invariably subject to some degree of aggregation, it is undoubtedly true that much of the variation in unit values will actually result from household choice regarding the nature of the goods purchased.

Deaton (1987, 1988, 1990, 1997) has developed a way of modelling price reactions jointly with choice of unit values in data of this type, under assumptions about fixity of underlying relative prices within spatially defined areas. The model suffers however from reliance on an approximation which is only compatible everywhere with demand theory for the theoretically unappealing loglinear demand specification. We develop an alternative, but related, approach which exploits the implicit links between quantity and unit value choices. This marks an advance on previous work in allowing in principle for appealing Engel curve specifications to be combined with a model of unit value determination in a way which is fully consistent with demand theory.

We illustrate the technique with an application to Czech Family Budget Survey data which has the feature that the geographical location of households is fairly precise. The preference specification used is a linear approximation of the Almost Ideal Demand (AID) system and the eight goods categories retained are six categories of food, plus clothing and footwear. In order to avoid unnecessary separability assumptions, the demand system is estimated conditionally on expenditures on several other good categories, on durable ownership and on labour market status. The results are encouraging, and our approach has subsequently been applied with success in the context of the estimation of heterogeneous labour demand functions by De Vreyer (2000).

Section 2 discusses relevant points from demand theory. Section 3 outlines a three stage estimation methodology. Section 4 describes the Czech data used and Section 5 presents illustrative results.

2 Demand and unit values

We start by developing an approach to modelling the determination of unit values. Our assumptions and notation follow those of Deaton's several papers on the subject. For the purpose of empirical investigation, goods are taken to be organised into m groups such as meat, fish, clothes and so on. Consumption within a group G is a vector of quantities $\mathbf{q}_G$ with unobserved prices $\mathbf{p}_G$. The consumer's total budget is X.

What we attempt to model is the determination of an observed group quantity index Q_G, defined by

$$Q_G \equiv \mathbf{1}'_G \mathbf{q}_G \ , \tag{1}$$

where $\mathbf{1}'_G$ is a vector of ones[1] and observed group spending $x_G \equiv \mathbf{p}'_G \mathbf{q}_G$ from which together we can calculate a unit value $V_G \equiv x_G / Q_G$.

Households reside within identifiable regional clusters and two central assumptions are made regarding the spatial variation in the unobserved prices.

[1]For simplicity, we assume, without loss of generality, that $\mathbf{q}_G$ is measured in the units chosen for aggregation by the data collector (like weight for meat or pairs for shoes).

Firstly, we assume that relative prices *within* each group are fixed everywhere, so that $\mathbf{p}_G = \pi_G \mathbf{p}_G^0$, where π_G is defined as a scalar (Paasche) linear homogeneous price *level* for the group (for instance, the price of meat), and $\mathbf{p}_G^0$ is a vector representing the fixed within-group relative price structure (for instance, the relative prices of different types and qualities of meat). This assumption will allow us to treat group G as a Hicks aggregate, so that $\mathbf{q}_G$ will be a function of total spending X and of the vector π of group price levels (generally, omission of a G subscript for a group variable will denote the vector of values for all groups).[2] As a consequence, Q_G, x_G and V_G, as well as related variables such as budget shares, $w_G \equiv x_G/X$, can also be written in this way. Secondly, the price vector π is assumed constant within the spatial clusters, though allowed to vary between clusters, so that we will use the notation π^c for clusters $c = 1, \ldots, C$.

Deaton's approach is to estimate the within-cluster relationships whereby Q_G (or x_G or w_G) and V_G are determined as functions of X, and then to use cross-equation restrictions to recover structural price parameters from the between-cluster correlations between the residuals. We argue that this method is theoretically convincing only for a particular and unattractive demand specification, specifically loglinear demands. The strength of the cross-equation restrictions between the Q_G and V_G equations rules out mutually consistent specifications for the two if adopting specifications with more attractive Engel curves. We suggest an alternative approach, which combines the advantage of more general theoretical consistency and greater computational convenience.

[2] The assumption appears very strong, but Lewbel (1996) shows that this type of aggregation will be possible under the much weaker assumption of stochastic independence between π_G and the vector of relative prices $\mathbf{p}_G/\pi_G$. This will be the case, at least approximately, if the relative prices are stationary over time, whereas π_G is not. We will come back to this assumption when discussing the stochastic structure of the econometric specification.

Assuming weak separability of preferences in the partition corresponding to groups $1, \dots, G, \dots, m$, and using homogeneity, we can write

$$\begin{aligned} \mathbf{q}_G &= \mathbf{f}_G(x_G, \mathbf{p}_G) \\ &= \mathbf{f}_G(V_G Q_G/\pi_G, \mathbf{p}_G^0). \end{aligned} \tag{2}$$

Suppressing $\mathbf{p}_G^0$ (given that it is fixed) and noting that V_G/π_G then depends only on $\mathbf{q}_G$, we can write

$$V_G = \pi_G h_G(V_G Q_G/\pi_G) \tag{3}$$

which defines an implicit relationship between V_G/π_G and Q_G. The fact that the unit values are independent of the component items given π_G and Q_G does not require any assumption that goods are consumed in fixed proportions — indeed that assumption would have the much stronger consequence that V_G would be constant. This equation, which makes clear the cross-equation restrictions on the functional forms of quantity and unit value equations, is central to our treatment. It implies the equation

$$\frac{\partial \ln V_G/\partial \ln \pi_H - 1_{[G=H]}}{\partial \ln V_G/\partial \ln X} = \frac{\partial \ln Q_G/\partial \ln \pi_H}{\partial \ln Q_G/\partial \ln X}\ , \tag{4}$$

used by Deaton to derive price elasticities at the second stage of his estimation procedure, and can therefore be seen as implicitly underlying his results although not explicitly stated. This follows from

$$\begin{aligned} \frac{\partial \ln V_G}{\partial \ln \pi_H} - 1_{[G=H]} &= e_G^h \left(\frac{\partial \ln V_G}{\partial \ln \pi_H} + \frac{\partial \ln Q_G}{\partial \ln \pi_H} - 1_{[G=H]} \right), \\ \frac{\partial \ln V_G}{\partial \ln X} &= e_G^h \left(\frac{\partial \ln V_G}{\partial \ln X} + \frac{\partial \ln Q_G}{\partial \ln X} \right), \end{aligned}$$

where e_G^h denotes the elasticity of $h_G(.)$ with respect to its (scalar) argument, and is a consequence of the separability assumptions made, and not simply of definitions.

If both the quantity and unit value relationships are specified to be double logarithmic, then this specification is compatible with (3). Yet there are very strong restrictions on the coefficients: if $\ln Q_G = \alpha_G + \beta_G \ln X + \sum_H \gamma_{GH} \ln \pi_H$ and $\ln V_G - \ln \pi_G = a_G + b_G \ln X + \sum_H c_{GH} \ln \pi_H$, then by (4) $\beta_G / b_G = \gamma_{GH} / c_{GH}$ for all H. Even without these, the double logarithmic specification is very unsatisfactory: loglinear demands do not satisfy adding up and cannot capture zero demands. This is worrying since major difficulties arise if the method is applied with other functional forms. If, for instance, an AID type budget share equation — with w linear in $\ln X$ and $\ln \pi$ — is adopted while the log unit value is also specified linearly in the same variables, this is not compatible with (3) (except under extremely strong restrictions, see Appendix A for details). This specification has been adopted in Deaton (1990), Deaton and Grimard (1991) and Ayadi, Baccouche, Goaied and Matoussi (1995). In none of these papers is the incompatibility explicitly recognised, although Deaton (1997) suggests that it might be appropriate to use (4) at mean sample, assuming constancy of elasticities as a reasonable approximation to the truth.

From (3) it follows that, given a functional form ϕ for budget shares, one would need to estimate a consistent system

$$\begin{aligned} w_G &= \phi_G (X, \pi) \ , \\ \ln V_G &= \ln \pi_G + \ln h_G \left[\frac{X}{\pi_G} \phi_G (X, \pi) \right] \ . \end{aligned} \tag{5}$$

A simple linear specification in $\ln X$ and $\ln \pi$ for the share equation therefore requires a unit value equation which will be non-linear in these variables. The problem here is that, once the quantity or budget share relationship is specified, (3) imposes too many cross-equation restrictions to permit also an unrestricted dependence of unit values on X and π.

This paper aims to make an advance toward the important goal of developing a simple theory-consistent method capable of handling more plausible Engel curves. Our central suggestion is to specify the quantity or budget share relationship, $\phi_G(X, \pi)$, and then to derive a relationship between V_G and Q_G from an independent specification of (3) (since the form of h_G is unrestricted). In the case of homothetic within-group demands (2) and therefore constant within-group budget shares, h_G is constant and V_G is therefore proportional to π_G.[3] More generally, h_G will be increasing if more expensive goods within group G tend to be income elastic and decreasing if they tend to be income inelastic. In some cases it will be possible to write $V_G = \pi_G \psi_G(Q_G)$ for some function ψ_G. To be more specific, for example, a specification

$$\ln V_G = a_G + b_G \ln Q_G + \ln \pi_G, \tag{6}$$

which would arise from assuming $h_G\left(V_G Q_G/\pi_G\right) = A_G(V_G Q_G/\pi_G)^{B_G}$, where $b_G = A_G/(1 - A_G)$ and $a_G = (1 + b_G)\ln B_G$, is generally admissible on theoretical grounds. In addition to its theoretical consistency, the absence from the unit value equation of terms in π_H for $H \neq G$ and the known unitary elasticity with respect to π_G considerably simplify estimation of structural parameters, as will be apparent below.

3 Specification and estimation

Following Deaton, we will choose for our specification of the budget shares the *approximate* AID model with a log-linear approximation to the log price index. Clearly, using the full AID specification, or even some quadratic extension thereof (see e.g. Banks, Blundell and Lewbel, 1997), would yield a superior specification, but the implied non linearity would defeat the within-cluster estimation strategy adopted below. Thus, while we achieve theoretical consistency in the specification of the unit value equation, we only have an approximation thereof in the specification of the share equation — al-

[3] This is pointed out by De Vreyer (2000).

though arguably quite a good one (although Pashardes, 1993, documents problems with this approximation). The share equation for category G, demanded by household h in cluster c, is thus given by

$$w_G^h = \alpha_{0G} + \mathbf{Z}^h \alpha_G + \sum_H \gamma_{GH} \ln \pi_H^c + \beta_G \ln \breve{x}^h + u_G^h \tag{7}$$

where $\breve{x}^h$ is deflated expenditure, $\ln \breve{x}^h \equiv \ln X^h - \ln P^c \equiv \ln X^h - \sum_H \lambda_H \ln \pi_H^c$, P^c is a cluster price index for suitably chosen weights.[4] This leads to the equation

$$w_G^h = \alpha_{0G} + \mathbf{Z}^h \alpha_G + \sum_H \delta_{GH} \ln \pi_H^c + \beta_G \ln X^h + u_G^h, \tag{8}$$

with $\delta_{GH} = \gamma_{GH} - \beta_G \lambda_H$. Vector $\mathbf{Z}^h$ includes socio-demographic characteristics as well as further conditioning variables, mentioned in the introduction, and which will be described in the next section. Several of these are potentially endogenous and will be instrumented.

For (6) we assume that the unit value equation is of the form

$$\ln V_G^h = a_{0G} + \mathbf{Z}^h \mathbf{a}_G + \ln \pi_G^c + b_G \ln Q_G^h + v_G^h \ . \tag{9}$$

We assume independence between observations. This may appear unduly restrictive, as it rules out the presence of cluster effects. But firstly, we have to rule out the simultaneous appearance of cluster effects in both share and unit value equations, as this would preclude the identification of the price effects. Secondly, allowing cluster effects in the share equation only would not change anything in the sequel, provided these effects were independent

[4]While it would clearly be preferable to allow the weights of the price index to be cluster-specific, the following points can be made. Firstly, the existence of a cluster price index does not conflict with the recognition that unit values are household-specific. Secondly, the true underlying prices should be constant within clusters by the law of one price, even if unit values are not. Finally, it is only the weights which are constant across clusters and not the price index itself, and this represents already a substantial improvement on the typical treatment of prices as constant across the whole territory. In our empirical work a cluster will be defined as a spatial unit at a point in time

of π. This is where Lewbel's assumption is helpful again: allowing the (unobserved) relative prices $\mathbf{p}_G^0$ to vary across clusters — and thus become $\mathbf{p}_G^{0c}$ — introduces a cluster effect; assuming independence between $\mathbf{p}_G^{0c}$ and π^c makes this cluster effect innocuous.[5] And thirdly, postulating additive errors in these equations is questionable anyway, as equation (5) suggests.

The covariance matrix $\mathbf{\Omega}$ of the vectors $\left(\mathbf{u}^{h'}, \mathbf{v}^{h'}\right)'$ is assumed constant across observations and otherwise unrestricted. This homoscedasticity assumption is less plausible for log unit values than it is for budget shares, but we reckon that it would be difficult to relax it in the unit value model, as should become apparent below.

A first strategy might be to estimate (8) replacing prices with unit values while instrumenting the latter. An approach of this type has been adopted by Pitt (1983) and Strauss (1982). The implicit assumption of such an approach is that the vector of unit values $\mathbf{V}^h$ can be treated simply as an error-ridden observation of the price vector π^c, with a measurement error that is independent of π^c. In the context of our unit value model such an approach will be badly misspecified if b_G is not equal to zero in (9), in which case the ratios V_G^h/π_G^c will depend on the outlay X^h and the vector π^c of price indices. In such circumstances the parameters of (8) will not be properly recovered under the naive approach.

The estimation proceeds in three stages. In the first stage we estimate for each good a share equation and a log unit value equation using within-cluster estimation and instrumental variables in a 2SLS framework.[6] In the second

[5]Thanks to Philippe De Vreyer for having pointed this out.

[6]There are two reasons here for preferring 2SLS to the more efficient 3SLS procedure. First, 3SLS risks contaminating the estimates of the share equation by a misspecification of the unit value equation (or the reverse, but we have more confidence in the validity of the share specification). Second, 2SLS estimates of the share equations will automatically satisfy adding-up restrictions, whereas this does not necessarily hold for 3SLS estimates (see e.g. Bewley, 1986).

Further note that the within-cluster technique adopted will not only sweep away the stage we retrieve the price coefficients using between-cluster estimation while taking account of measurement errors on the unit values. The third stage imposes the symmetry restrictions through minimum distance estimation.

3.1 First stage

Averaging (8) over households within the cluster c yields

$$\overline{w_G}^c = \alpha_{0G} + \overline{\mathbf{Z}}^c \alpha_G + \sum_H \delta_{GH} \ln \pi_H^c + \beta_G \overline{\ln X}^c + \overline{u_G}^c. \tag{10}$$

The vector $\widehat{\alpha}_G$ and the scalar $\widehat{\beta}_G$ are recovered from within-cluster estimation, i.e. the estimating equation is obtained by subtracting (10) from (8). Similarly, forming cluster means from (9)

$$\overline{\ln V_G}^c = a_{0G} + \overline{\mathbf{Z}}^c \mathbf{a}_G + \ln \pi_G^c + b_G \overline{\ln Q_G}^c + \overline{v_G}^c \ , \tag{11}$$

$\widehat{\mathbf{a}}_G$ and $\widehat{b}_G$ result from within-cluster estimation.

Endogeneity issues are addressed by use of instrumental variables where appropriate, as discussed in the data section below. Several variables are instrumented by cluster means excluding the current observation. We justify this technique as follows. In the regression $y_i - \bar{y}^c = \left(X_i - \overline{X}^c\right)\beta + u_i - \bar{u}^c$, let $\check{X}_i = \frac{1}{n_c - 1} \sum_{\substack{j \in c \\ j \neq i}} X_j$ and consider the asymptotic covariance between $\check{X}_i$ and $u_i - \bar{u}^c$: we have

$$\mathrm{E}\left[\check{X}_i \left(u_i - \bar{u}^c\right)\right] = -\mathrm{E}\left[\check{X}_i \bar{u}^c\right] = -\frac{1}{n_c}\frac{1}{n_c - 1}\sum_{\substack{j \in c \\ j \neq i}} \mathrm{E}\left[X_j u_j\right] = -\frac{1}{n_c}\mathrm{E}\left[X u\right] \ ,$$

which goes to zero when the number of observations per cluster goes to infinity.

unobservable price indices from the share equations, but also any cluster-specific effect. At this stage, the independence between cluster effects and prices plays no role, but it becomes important in the next stage.

3.2 Second stage

Our initial approach had been to estimate parameters δ_{GH} in the second stage and to derive estimates of the parameters of interest λ_G and γ_{GH} for all G and H by minimum distance. The γ_{GH} must satisfy symmetry, $\gamma_{GH} = \gamma_{HG}$, and homogeneity, $\sum_H \gamma_{GH} = 0$, which also implies the adding-up restriction. The λ vector is subject to the restrictions $\lambda_G > 0$ and $\sum_H \lambda_H = 1$ (positive linear homogeneity of the price index). Besides these restrictions, the relationships between the parameters of interest (γ, λ) and the auxiliary parameters $\psi = (\delta, \beta)$ are the m^2 equations $\delta_{GH} = \gamma_{GH} - \beta_G \lambda_H$. Unfortunately, these restrictions are not sufficient to identify the parameters of interest, although their number might at first lure one into thinking they would. Indeed, if (γ, λ) satisfy all the restrictions, so will $(\dot{\gamma}, \dot{\lambda})$, with $\dot{\gamma}_{GH}(\kappa) = \gamma_{GH} + \kappa\beta_G\beta_H$ and $\dot{\lambda}_H(\kappa) = \lambda_H + \kappa\beta_H$, for all κ such that $\dot{\lambda}_H(\kappa) > 0$ for all H, given that $\sum_H \beta_H = 0$. Thus we arbitrarily set $\lambda = \bar{\mathbf{w}}$, the vector of average budget shares over all households, which provides identification of the γ coefficients. The second stage thus consists in estimating a matrix of γ coefficients conditional on $\lambda = \bar{\mathbf{w}}$, without imposing symmetry, the latter being imposed in a third stage by minimum distance.

Separating observables and unobservables in (10) and (11) yields

$$\begin{aligned} \eta_G^c &\equiv \overline{w_G}^c - \overline{\mathbf{Z}}^c \hat{\alpha}_G - \hat{\beta}_G \overline{\ln X}^c && (12) \\ &= \alpha_{0G} + \sum_H (\gamma_{GH} - \beta_G \lambda_H^c) \ln \pi_H^c + \overline{u_G}^c \equiv \eta_G^{*c} + \overline{u_G}^c && (13) \end{aligned}$$

and

$$\zeta_G^c \equiv \overline{\ln V_G}^c - \overline{\mathbf{Z}}^c \mathbf{a}_G - b_G \overline{\ln Q_G}^c = a_{0G} + \ln \pi_G^c + \overline{v_G}^c \equiv \zeta_G^{*c} + \overline{v_G}^c \; . \quad (14)$$

Only between-cluster information needs to be considered here, since no information on the price responses remains to be exploited within clusters.

The true relationship between η_G^{*c} and the vector ζ^{*c} with components ζ_H^{*c} is

$$\eta_G^{*c} = \rho_G + \omega_G^t + \sum_H \left(\gamma_{GH} - \beta_G \lambda_H\right) \zeta_H^{*c},$$

where $\omega_G^t = \sigma_G^t - \sum_H \left(\gamma_{GH} - \beta_G \lambda_H\right) s_H^t$ and $\rho_G \equiv \alpha_{0G} - \sum\limits_H \left(\gamma_{GH} - \beta_G \lambda_H\right) a_{0H}$, and thus

$$\eta_G^{*c} + \sum_H \beta_G \lambda_H \zeta_H^{*c} = \rho_G + \omega_G^t + \sum_H \gamma_{GH} \zeta_H^{*c}.$$

This suggests the regression of $\eta_G^c + \sum\limits_H \beta_G \lambda_H \zeta_H^c$ on ζ^c, time dummies and a constant. Measurement error bias is caused by the correlation between the vectors ζ^c, $\overline{\mathbf{v}}^c$ and possibly $\overline{\mathbf{u}}^c$, but is easily corrected because the variance of $(\overline{u}_G^c, \overline{\mathbf{v}}^c\ )$ can be estimated as

$$\hat{V}\begin{pmatrix} \overline{u}_G^c \\ \overline{\mathbf{v}}^c \end{pmatrix} = \frac{1}{n_c} \begin{bmatrix} \hat{\Omega}_{u_G} & \hat{\mathbf{\Omega}}_{u_G \mathbf{v}} \\ \hat{\mathbf{\Omega}}_{\mathbf{v} u_G} & \hat{\mathbf{\Omega}}_{\mathbf{v}} \end{bmatrix},$$

where each term of $\hat{\mathbf{\Omega}}$ is obtained from the first stage residuals. This is the place where the difficulty of relaxing the homoscedasticity assumption becomes manifest. Under the assumptions that the λ_H are known constants and that β is known, a consistent estimator of the vector γ_G, after demeaning the η and ζ variables and scaling them by $\sqrt{n_c}$, is given by[7]

$$\hat{\gamma}_G = [\sum_{c=1}^{C} n_c \zeta^c \zeta^{c\prime} - \hat{\mathbf{\Omega}}_v]^{-1} [\sum_{c=1}^{C} n_c \left(\eta_G^c \zeta^c + \beta_G \zeta^c \zeta^{c\prime} \lambda\right) - \hat{\mathbf{\Omega}}_{\nu u_G} - \hat{\mathbf{\Omega}}_{\mathbf{v}} \lambda]. \quad (15)$$

3.3 Third stage

We estimate symmetry-restricted parameters γ by minimum distance estimation conditional on λ. Following the efficiency arguments of Kodde, Palm and Pfann (1990, Theorem 5) we minimise only over γ rather than over γ and β. Given the linearity of the restrictions, the computations boil down to GLS estimation in the parameter space. This requires an estimate of the

[7]Asymptotics here concern the thought experiment where both the number of observations in each cluster n_c and the number of clusters C go to infinity.

variance of the unrestricted estimator, and a convenient way to obtain this is to recognise that the procedure of the first two stages falls into the framework of sequential GMM outlined by Newey (1984), as already pointed out by Deaton (1990). However an attractive alternative lies in bootstrapping and this is the path we follow. In producing the bootstrap samples we keep the size of clusters fixed, and the number of bootstrap samples is set equal to the number of observations.

4 Empirical illustration

We provide an illustration using data from the Czech Family Budget Surveys for 1991 and 1992. Households included in the sample were asked to maintain an expenditure diary for a full twelve months, recording both quantities and expenditures for certain goods. The length of the recording period has the advantage of virtually eliminating infrequency of purchase as an explanation for zero records on most main expenditure items. However the burden imposed on participants must have been arduous and we discarded 480 households who did not take part over the full year. The data is a panel, but the household identifier is not retained between years, necessitating considerable effort to recover and use the panel structure (at best with imprecision). Households whose circumstances change in any major way are dropped from the sample — an unfortunate feature which again diminishes the usefulness of the panel aspect and which must also affect the cross-sectional sampling properties.

We concentrate in this paper on a subsample of married couples. The wife's labour force participation is used as a conditioning variable and instrumented. The sample size obtained in pooling the two years is 4668 households. Given that the number of identifiable geographical clusters is

179, we have an average of 26 households per cluster, with a minimum of 7 and a maximum of 60.[8]

Eight categories of goods were selected for demand estimation. The choice was constrained by the need to have both quantities and expenditures available. Since we have no reason to believe that the availability of quantity information — related to the survey design — is directly connected to the structure of preferences, it is not attractive to assume that the latter are separable in the corresponding partition. Rather, following Browning and Meghir (1991), we condition the budget shares (though not the unit values) of the included goods on the expenditures on the excluded goods.[9] Furthermore we will condition the budget shares for the modelled goods on durable ownership and on a variable describing the labour market status of the household. In the words of Browning and Meghir:

> "The conditional demand system will be correctly specified whether or not [labour market status] is chosen optimally. Additionally we do not need to model explicitly the budget constraint for the conditioning goods. This is particularly significant for labour supply and for durables [...] Conditional demand functions are an economical way of relaxing separability and still maintaining the focus on the goods of interest."

Homogeneity with respect to the prices of the excluded goods will be ensured by expressing the conditioning expenditures in relative terms with respect to one of them, namely housing expenditure.[10] The problem of zero conditioning expenditures is taken care of by the introduction of dummies. Under weak separability, these conditioning variables should play no role in

[8] The reason why we have an uneven number of geographical clusters over the two time periods is that one of the clusters had too few observations in one of the periods to be retained in the analysis.

[9] Detailed lists of the goods in both categories, including also the exact composition of each aggregate, are given in Appendix B.

[10] Thanks to Arthur Lewbel for pointing this out.

the demand equations. The compatibility between this conditional approach and the unit value model described above is ensured by the fact that the conditional cost function is amenable to Hicks aggregation.

For some commodities, the survey includes "in kind" quantities and expenditures as well as purchased goods. We treat all quantities together (purchased or not), evaluating in kind quantities on the basis of the unit values as an approximation.

Furthermore we condition the budget shares and unit values for the modelled goods on durable ownership and on a variable describing the labour market status of the household. Other variables used include a wide range of socio-demographic and housing characteristics.[11] We treat as endogenous the logs of total expenditure X and of quantity Q_G, the conditioning expenditures and durable ownership variables and labour market status of the wife. Instruments include the log of income (which should be correlated with $\ln X$ and $\ln Q_G$), wife's age and education, and age of the youngest child (which should all be correlated with the wife's participation) and cluster means of the conditioning expenditures and durable ownership excluding the current observation.

5 Estimates

In Tables 1 and 2, we report the first stage results for all goods along with their asymptotic standard errors.[12] These are the outcome of within-cluster 2SLS regressions of the type explained above: the estimating equations are

[11] Descriptive statistics on the variables used are given in Appendix B.

[12] Insofar as we have not, for the compelling reasons outlined above, attempted to identify households present in both years, we should recognise the implications for the validity of the standard errors. In particular, the inferences drawn are based on inconsistent estimates of the variance of the estimated coefficients. One way out of this difficulty would be to report results separately for 1991 and 1992, which ought also to be of interest in their own right.

obtained by subtracting (10) from (8), and (11) from (9). Note that in the equations presented for unit values the price effects have been swept away, so that only the choice component embodied in the unit values is reflected here.

From the share equations (Table 1) we see that several of the conditioning goods are significant, implying decisive rejection of separability of preferences in the partition modelled goods / other goods.

The woman's participation has a significant impact on four budget shares: on meat and starches a negative impact, and a positive one on clothing and on footwear, implying also clear rejection of the separability of preferences in the partition leisure/goods. It has a positive and significant impact on all unit values, except for the two categories dairy and vegetables/fruit, and the combined effect on quantity and on unit value of meat and starches as opposed to clothing and footwear has a neat interpretation. Several other variables have contrasted effects on quantity and unit value (see e.g. the effects of education of the household head on the quantity and unit value of meat purchased), but a complete enumeration would be tedious, and the reader will be able to browse through the results without our guidance.

Turning to the impact of expenditure on budget shares, for meat, alcohol and clothing the coefficient of $\ln X$ is significantly positive. Budget shares for vegetables/fruit and for dairy products are significantly negatively affected by total spending. A test for the absence of the term $(\ln X)^2$ rejects the null for all goods. While this may suggest that a QUAID specification would be more appropriate, this would not be straightforward to implement, since the squared term should include the log of real expenditure, and thus involve a price index in a non-linear fashion, thus defeating within-cluster estimation.

Finally, note that expenditure on tobacco correlates positively with the budget shares on meat and alcohol, negatively with those on dairy, starches and vegetables/fruit, while expenditure on hygiene and health go the other way round. All these qualitative results conform with intuition.

From the unit values equations (Table 2) we see that clear evidence of a relationship between unit value and quantity appears for two goods, dairy products and starches, where the effect is negative. There is weaker evidence of a positive relationship in the case of vegetables and fruit. The fact that we obtain significant effects of quantity on unit value implies that the straightforward instrumenting approach of Pitt (1983) and Strauss (1982) would, as suggested, be inconsistent.

It is interesting to note the influence of the durable ownership variables: possession of a freezer, for instance, which is significant in only two of the food share equations, meat (+) and starches (-), appears to have a significant influence in five of the food unit value equations, always entering with a negative sign. To a lesser extent a similar observation can be made for motor vehicle ownership — while it does not appear to affect budget shares, there is some evidence that it is associated with lower unit values for some goods. The most intuitive explanation for both effects is that households thus equipped have better opportunities for purchasing in large quantities and for either taking advantage of low price opportunities or searching for them.

Third stage estimates of the symmetry (and homogeneity, which follows given adding-up) restricted parameters γ_{GH} are given in Table 3. These are estimated using the measurement error correction procedure at the second stage.[13] The table also reports the minimised value of the criterion, which provides a χ^2 test of the restrictions. We obtain a rejection at any reasonable

[13] The underlying second stage estimates are omitted in order to save on space, but are available on request.

level of significance. One possible explanation could be that the restricted estimates also embody the restriction that the deflator of total expenditure is a Stone price index which varies across clusters, but with fixed weights. Given our focus on a sample of married couples, a more direct source of this rejection may lie in the misspecification of the unitary model of household preferences, as forcefully documented by Browning and Chiappori (1998), who reject symmetry for couples but not singles on Canadian data.

Marshallian elasticities based on the first and third stage estimates and computed at mean sample using the formulae

$$\begin{aligned} e_{GH} &= \left(\gamma_{GH} - \beta_G \bar{w}_H\right) / \bar{w}_G - 1_{[G=H]}, \\ e_G &= 1 + \beta_G / \bar{w}_G, \end{aligned}$$

are reported in Table 4. Expenditure elasticities for all goods are significantly different from zero. Dairy, vegetable and fruit and alcohol categories present elasticities which differ significantly from one, classifying alcohol as a luxury, dairy, vegetables and fruit as necessities. A surprise is perhaps the low budget elasticity of the vegetable and fruit category, which is at variance with results typically found for other countries. All but one of the (uncompensated) own price elasticities are negative and several are significantly different from zero. Several significant cross price effects are also observed.

The only point of reference we have in assessing these elasticities is the work of Ratinger (1995), based on monthly data on 300 households of employees from January 1990 to September 1992, apparently using published average expenditures on food.[14] Given the differences between the two studies, concerning data and methodologies, the comparison is difficult. For in-

[14] The paper gives no information on the nature of the price data used.

stance, depending on the degree of aggregation of goods, Ratinger reports expenditure elasticities between .98 and 1.14 for meat.

Table 1: **Engel curves**

	Meat		Dairy		Starches		Veg/Fruit		Sweet		Alcohol		Clothes		Shoes	
	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*
Household characteristics																
Wife's participation	**-3.98**	-2.8	-1.25	-1.3	**-2.08**	-2.7	1.23	1.6	.081	.1	-1.61	-1.3	**5.91**	3.8	**1.70**	2.5
Blue collar	**.819**	3.3	.170	1.0	.157	1.2	-.095	-.8	-.022	-.3	-.261	-1.2	-.495	-1.7	**-.273**	-2.3
Farmer	**1.62**	5.2	.173	.8	**.388**	2.3	**-.305**	-2.1	-.070	-.6	.249	1.1	**-1.54**	-4.3	**-.516**	-3.6
Age of head of household	**5.36**	6.7	.948	1.7	**1.38**	3.5	.055	.1	.237	.9	**1.84**	2.6	**-7.87**	-9.0	**-1.95**	-5.2
Age of hoh squared	**-.494**	-5.5	-.028	-.5	**-.091**	-2.0	.060	1.3	.007	.2	**-.255**	-3.3	**.638**	6.7	**.163**	4.0
Owner-occupier	**.726**	2.9	**.554**	3.2	**.338**	2.6	**-.519**	-4.3	-.159	-1.8	-.086	-.4	**-.872**	-3.1	.018	.2
No mod-cons	-.178	-.7	.202	1.1	.169	1.2	**.274**	2.0	-.039	-.4	-.149	-.7	-.359	-1.2	.081	.6
Number of. hh members	**-1.07**	-2.6	**1.84**	6.7	**1.00**	4.6	**.422**	2.0	.229	1.5	**-1.631**	-4.7	**-1.30**	-2.8	**.506**	2.7
Average age of children	.051	1.9	.020	1.1	**.063**	4.5	-.026	-1.9	-.019	-1.9	**-.090**	-4.1	.002	.1	-.001	-.1
Basic education - hoh	**.831**	3.8	**-.569**	-3.7	219	1.9	.060	.5	.099	1.3	.022	.1	**-.682**	-2.7	.019	.2
Advanced education - hoh	**-1.43**	-4.9	-.172	-.8	**-.386**	-2.6	-.012	-.1	-.098	-.9	.034	.1	**1.75**	5.0	**.316**	2.3
Rural	**.571**	2.0	-.037	-.2	.290	1.9	**-.486**	-3.6	.077	.8	**.571**	2.6	**-.800**	-2.6	-.187	-1.4
Space per person	**.025**	2.1	.013	1.6	.012	1.7	**-.013**	-2.2	**.011**	2.2	-.005	-.5	**-.035**	-2.5	-.008	-1.6
Durable ownership																
Gas supplied	.140	.6	**-.717**	-4.2	**-.299**	-2.3	-.162	-1.3	**-.222**	-2.4	.267	1.2	**.680**	2.4	**.312**	2.5
Number of leisure durables	.039	.7	**-.092**	-2.4	003	.1	-.017	-.6	.002	.1	**.152**	3.3	-.027	-.4	**-.060**	-2.3
Freezer	**.593**	3.0	-.092	-.7	**-.261**	-2.6	.043	.4	-.076	-1.1	.277	1.7	-.361	-1.6	-.124	-1.3
Phone	**.759**	3.4	**-.363**	2.4	-.202	-1.8	-.041	-.3	-.029	-.4	.013	.1	-.183	-.7	.045	.4
Car or motor bike	.101	.4	-.189	-1.1	-.202	-1.5	-.109	-.8	.052	.6	-.308	-1.4	**.662**	2.3	-.006	.0
Automatic washing machine	-.317	-1.5	**.313**	2.0	-.057	-.5	.006	.1	-.003	-.0	-.010	-.1	-.022	-.1	.089	.8
Food processor	-.191	-1.0	**.280**	2.2	**-.201**	-2.2	.099	1.1	.001	.0	**-.494**	3.3	.372	1.8	.133	1.5
Caravan and/or dacha	.093	.4	-.202	-1.2	**.267**	2.1	**-.281**	-2.2	-.156	-1.7	**.481**	2.2	-.167	-.6	-.035	-.3
Garage	.299	1.4	-.045	-.3	-.196	-1.8	-.009	-.1	-.073	-1.0	**-.422**	-2.4	.348	1.4	.097	1.0
Conditioning expenditures																
ln(Transport)	**-.693**	-3.9	**.339**	2.4	.008	.1	.193	1.9	.034	.5	-.209	-1.3	.323	1.5	006	.1
ln(Hygiene)	**-1.96**	-4.4	**.878**	2.9	**-.977**	-4.2	**1.06**	4.4	**.536**	3.4	**-1.02**	-2.7	.986	1.9	**.487**	2.4
ln(Food out)	-.042	-.2	**-.323**	-2.2	.046	.4	-.128	-1.2	-.018	-.2	**.491**	2.8	-.094	-.4	.068	.7
ln(Culture)	**-1.20**	-5.4	**-.344**	-2.2	-.065	-.6	-.003	-.0	.059	.8	**.419**	2.4	**.938**	3.8	**.197**	2.0
ln(Fuel)	.149	.8	.261	1.7	**.287**	2.2	-.021	-.2	.104	1.7	.036	.2	**-.697**	-2.8	-.118	-1.0
ln(Tobaccco)	**.246**	4.6	**-.223**	-5.9	**-.077**	-2.9	**-.150**	-5.4	-.023	-1.2	**.293**	6.5	-.064	-1.0	-.003	-.1
ln(Other food)	**-.636**	-2.0	**.543**	2.4	-.054	-.3	**1.09**	6.2	**.363**	3.0	-.120	-.4	**-1.05**	-2.7	-.129	-.8
ln(Textiles)	**-.900**	-7.6	**-.275**	-3.4	**-.260**	-4.3	.094	1.6	**-.132**	-3.2	**-.250**	-2.4	**1.47**	10.2	**.258**	4.2
ln(Medical)	**-.461**	-5.9	**.128**	2.5	**.158**	3.9	**.222**	5.8	**.108**	4.0	-.088	-1.4	-.028	-.3	-.038	-1.0
ln(Furniture)	**-.352**	-4.7	-.087	-1.6	-.075	-1.9	-.030	-.7	-.040	-1.5	.067	1.0	**.437**	4.8	**.080**	2.0
No food out	1.10	.7	**-2.39**	-2.2	1.03	1.0	-1.12	-1.4	.405	.7	2.15	1.7	-1.08	-.6	-.097	-.1
No Tobacco	.461	1.0	**-.735**	-2.3	**-.773**	-3.3	**-.543**	-2.1	.065	.4	.671	1.8	.793	1.6	.061	.3
No Medical	-.905	-1.2	-.040	-.1	.472	1.3	.047	.1	.483	1.8	-.050	-.1	.260	.3	-.268	-.8
ln(Total expenditure)	**5.31**	2.3	**-8.55**	-5.5	-2.28	-1.9	**-4.85**	-4.1	-1.36	-1.6	**4.29**	2.2	**8.02**	3.1	-.578	-.5

Notes: All coefficients $\times$ 100

Table 2: **Unit value equations**

	Meat		Dairy		Starches		Veg/Fruit		Sweet		Alcohol		Clothes		Shoes	
	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*	Coeff	*t*
Household characteristics																
Wife's participation	**4.65**	*4.2*	1.839	*0.7*	**6.72**	*3.8*	692	*.3*	**8.23**	*2.4*	**23.3**	*5.4*	**38.1**	*6.8*	**46.3**	*7.0*
Blue collar	-.718	*-1.6*	**-3.01**	*-2.8*	**-1.45**	*-2.0*	-.080	*-.1*	1.03	*.7*	-3.23	*-1.7*	**-4.61**	*-2.0*	-2.92	*-1.2*
Farmer	-.741	*-1.2*	**-7.26**	*-5.2*	**-2.46**	*-2.6*	**-6.16**	*-4.0*	.469	*.2*	**-7.79**	*-2.9*	-4.41	*-1.1*	-4.78	*-1.0*
Age of head of household	-2.81	*-1.6*	**-10.4**	*-2.8*	**-12.3**	*-5.3*	**-12.4**	*-3.0*	-8.34	*-1.5*	**-34.3**	*-4.4*	**-34.3**	*-2.7*	-12.8	*-.8*
Age of hoh squared	.189	*1.1*	.745	*1.9*	**1.20**	*5.0*	**.870**	*2.2*	.928	*1.7*	**3.24**	*4.6*	**3.50**	*3.4*	1.60	*1.3*
Owner-occupier	-.318	*-.5*	.994	*.9*	**-2.30**	*-3.4*	-2.13	*-1.9*	**-3.02**	*-2.0*	**-5.60**	*-3.0*	.639	*.2*	**-4.50**	*-1.6*
No mod-cons	**-1.27**	*-2.1*	-.181	*-.2*	.023	*.0*	-1.34	*-1.3*	-2.27	*-1.5*	**-3.46**	*-2.0*	1.62	*.7*	-.000	*-.0*
Number of hh members	-.421	*-1.2*	1.024	*1.3*	**1.79**	*3.6*	-1.94	*-1.8*	**3.58**	*2.4*	1.66	*.6*	-5.45	*-1.3*	**-12.2**	*-2.5*
Average age of children	-.059	*-1.3*	-.101	*-1.0*	-.103	*-1.5*	**-.500**	*-5.0*	-.128	*-.9*	-.345	*-1.9*	.189	*.9*	**1.44**	*6.3*
Basic education - hoh	-.746	*-1.7*	**-2.32**	*-2.1*	-.460	*-.7*	-.425	*-.4*	1.04	*.7*	**-4.96**	*-2.4*	-4.01	*-1.3*	-3.54	*-1.0*
Advanced education - hoh	**2.54**	*5.0*	**6.09**	*4.8*	**2.69**	*3.1*	2.28	*1.7*	-1.01	*-.5*	**6.44**	*2.6*	6.54	*1.7*	2.71	*.6*
Rural	-.222	*-.5*	-1.56	*-1.3*	**-2.35**	*-3.1*	**-2.85**	*-2.4*	-1.73	*-1.1*	-3.34	*-1.8*	-.879	*-.4*	-4.55	*-1.8*
Space per person	-.013	*-.5*	**-.215**	*-4.3*	**-.074**	*-2.3*	**-.112**	*-2.0*	.070	*1.0*	-.006	*-.1*	-.142	*-1.1*	-.021	*-.1*
Durable ownership																
Gas supplied	**1.35**	*2.9*	-.353	*-.3*	.558	*.8*	989	*1.0*	-.960	*-.7*	2.59	*1.4*	1.81	*.8*	1.76	*.7*
Number of leisure durables	-.009	*-.1*	.216	*.9*	**.333**	*2.2*	-.317	*-1.4*	.461	*1.5*	**.949**	*2.4*	.746	*1.6*	.046	*.1*
Freezer	**-.904**	*-2.4*	-.266	*-.3*	**-2.50**	*-4.4*	**-5.51**	*-6.8*	**-2.63**	*-2.1*	**-2.99**	*-2.0*	.680	*.3*	-.542	*-.2*
Phone	-.089	*-.2*	1.195	*1.3*	.627	*1.1*	156	*.2*	-1.21	*-1.0*	1.158	*.8*	-2.425	*-1.4*	-.825	*-.5*
Car or motor bike	-.836	*-1.9*	**-2.89**	*-2.7*	-1.15	*-1.7*	-1.73	*-1.7*	1.30	*.9*	-2.871	*-1.6*	**5.359**	*2.7*	-.014	*-.0*
Automatic washing machine	.500	*1.2*	**2.20**	*2.3*	.513	*.8*	997	*1.1*	.601	*.5*	2.513	*1.5*	-1.650	*-.8*	-1.03	*-.5*
Food processor	-.370	*-1.0*	-.967	*-1.2*	-.220	*-.4*	**-2.64**	*-3.5*	**3.04**	*2.9*	**2.75**	*2.2*	-.700	*-.5*	-.451	*-.3*
Caravan and/or dacha	**-1.06**	*-2.3*	-.877	*-.8*	.291	*.4*	-1.20	*-1.2*	.725	*.5*	**-6.36**	*-3.6*	-1.168	*-.6*	-.766	*-.3*
Garage	.818	*1.6*	-.014	*.0*	-.722	*-1.2*	**-2.11**	*-2.5*	.337	*.3*	-2.37	*-1.6*	2.860	*1.7*	.107	*.1*
ln(Quantity)	1.47	*.9*	**-9.68**	*-2.2*	**-6.92**	*-2.3*	12.4	*1.9*	-10.1	*-1.1*	-3.22	*-.2*	20.3	*.7*	5.89	*.2*

Notes: All coefficients × 100

Table 3: **Symmetry restricted estimates of γ**

	Meat	Dairy	Starches	Veg/Fruit	Sweet	Alcohol	Clothes	Shoes
Meat	1.764							
	3.016							
Dairy	**-9.882**	**15.905**						
	1.507	*1.403*						
Starches	**-3.061**	**-2.056**	**3.551**					
	1.292	*0.813*	*1.245*					
Veg/Fruit	**-10.320**	**4.471**	**-1.892**	**17.383**				
	1.332	*0.808*	*0.736*	*1.147*				
Sweet	**2.723**	-0.669	**1.577**	-1.013	0.102			
	0.941	*0.591*	*0.549*	*0.531*	*0.581*			
Alcohol	1.489	0.462	-0.545	**3.066**	**0.906**	**-2.634**		
	0.820	*0.537*	*0.432*	*0.473*	*0.324*	*0.861*		
Clothes	**10.564**	**-8.211**	1.134	**-8.530**	**-2.994**	**-1.916**	**12.879**	
	1.472	*0.936*	*0.775*	*0.844*	*0.551*	*0.819*	*1.836*	
Shoes	**6.724**	-0.021	**1.293**	**-3.164**	-0.631	**-0.827**	**-2.926**	-0.448
	0.986	*0.611*	*0.550*	*0.572*	*0.386*	*0.361*	*0.622*	*0.623*

Notes: All coefficients and standard errors $\times$ 100.
Standard errors below coefficients.
Wald test of symmetry restrictions $\chi^2_{28} = 351.29$ (Critical value at 5% is 41.34).

Table 4: **Marshallian elasticities**

				Price					*Total*
	Meat	Dairy	Starches	Veg/Fruit	Sweet	Alcohol	Clothes	Shoes	*budget*
Meat	**-0.968**	**-0.428**	**-0.145**	**-0.424**	**0.095**	0.041	**0.381**	**0.254**	**1.195**
	0.118	*0.066*	*0.053*	*0.054*	*0.038*	*0.035*	*0.064*	*0.040*	*0.116*
Dairy	**-0.463**	-0.062	-0.064	**0.309**	-0.009	**0.070**	**-0.403**	0.026	**0.596**
	0.100	*0.086*	*0.050*	*0.056*	*0.036*	*0.034*	*0.070*	*0.037*	*0.113*
Starches	-0.231	-0.151	**-0.721**	**-0.158**	**0.159**	-0.033	0.141	**0.131**	**0.864**
	0.121	*0.081*	*0.122*	*0.069*	*0.052*	*0.042*	*0.079*	*0.053*	*0.134*
Veg/Fruit	**-1.003**	**0.591**	-0.146	**1.067**	-0.080	**0.382**	**-0.848**	**-0.320**	**0.356**
	0.148	*0.095*	*0.082*	*0.130*	*0.059*	*0.055*	*0.102*	*0.065*	*0.133*
Sweet	**0.566**	-0.077	**0.321**	-0.167	**-1.078**	**0.188**	**-0.510**	-0.103	**0.860**
	0.174	*0.116*	*0.103*	*0.099*	*0.117*	*0.063*	*0.112*	*0.073*	*0.165*
Alcohol	0.052	-0.042	**-0.130**	**0.340**	0.081	**-1.407**	**-0.331**	**-0.132**	**1.570**
	0.119	*0.083*	*0.061*	*0.065*	*0.044*	*0.132*	*0.117*	*0.049*	*0.276*
Clothes	**0.414**	**-0.469**	0.010	**-0.444**	**-0.168**	**-0.125**	**-0.275**	**-0.162**	**1.219**
	0.078	*0.055*	*0.040*	*0.045*	*0.028*	*0.042*	*0.096*	*0.032*	*0.154*
Shoes	**1.218**	0.015	**0.241**	**-0.553**	-0.106	**-0.138**	**-0.501**	**-1.018**	**0.842**
	0.187	*0.118*	*0.100*	*0.107*	*0.070*	*0.068*	*0.124*	*0.125*	*0.229*

Note: Standard errors below coefficients
Bold entries correspond to rejection of H_0: e=0; for expenditure elasticities, rejection of H_0: e=1 occurs for dairy, veg/fruit and alcohol.

6 Conclusion

We have presented a new approach to the estimation of demand systems on the basis of unit values and have argued that its advantage over Deaton's approach, which also treats unit values as consumer choice variables, is improved consistency with demand theory, while its advantage over the naive treatment of unit values as erre-ridden measurements of prices is improved statisticla consistency. Monte Carlo experiments designed to compare its performance with alternative methods are presented by Lahatte et al. (1998). They do suggest that our theory-consistent specification for the log unit value equation outperforms Deaton's first order Taylor expansion, but that both specifications may perform poorly when data are generated by a more flexible form.

Another advantage of our approach over alternatives is its relative computational simplicity, the main difference residing in the second stage where we can treat goods separately whereas a system estimation is necessary in Deaton's approach. This simplification may allow us to consider more complicated settings, where for instance spatial patterns of consumption are of primary interest. In this respect, it is interesting to note that in her work on spatial aspects of consumption, and while aware of Deaton's work, Case (1991) chooses to treat unit values as error-ridden measurements of prices rather than to model them as the outcome consumer choices. Combining, on one side, a proper treatment of the fact that unit values are outcomes of choice and, on the other side, the spatial patterns of demand, would seem a rewarding endeavour.

Interestingly our proposal resembles Lewbel's (1989) approach to the identification and estimation of demand systems in the absence of price variation in several respects. Lewbel makes the same homothetic weak separability assumption and develops a method allowing price indices, constructed from budget share estimation at the lower level even without price variation, to be used to estimate unrestricted upper level demand equations. An appealing extension would thus combine both approaches. A possibility would apply our within-cluster techniques using more disaggregated unit value information at the lower level[15], allowing a freeing up of the demand specification at the higher level in the fashion proposed by Lewbel. Finally it would also obviously be interesting to introduce these techniques in a collective model of household consumption as the rejection of symmetry suggests.

[15]Table 7 indicates the finest level of disaggregation at which unit value information is available in the Czech data

Appendix A. Implications of the same log-linear specification for shares and log unit values

Suppose the share equations are derived from the AID functional form,

$$w_G = \alpha_G + \sum_H \delta_{GH} \ln \pi_H + \beta_G \ln X + u_G \ ,$$

and that the log unit values have a similar form

$$\ln V_G = A_G + \sum_H D_{GH} \ln \pi_H + B_G \ln X + U_G \ ,$$

as in Deaton (1990). Then

$$\begin{array}{lcll}
\partial \ln Q_G / \partial \ln \pi_H & = & \partial \left(\ln w_G - \ln V_G \right) / \partial \ln \pi_H & = & \delta_{GH}/w_G - D_{GH} \ , \\
\partial \ln Q_G / \partial \ln X & = & \partial \left(\ln w_G + \ln X - \ln V_G \right) / \partial \ln X & = & \beta_G / w_G + 1 - B_G \ , \\
\partial \ln \xi_G / \partial \ln \pi_H & = & \partial \left(\ln V_G - \ln \pi_G \right) / \partial \ln \pi_H & = & D_{GH} - 1_{[G=H]} \ , \\
\partial \ln \xi_G / \partial \ln X & = & \partial \left(\ln V_G - \ln \pi_G \right) / \partial \ln X & = & B_G \ .
\end{array}$$

Hence, by (4), considering first the case where $G \neq H$:

$$\frac{\delta_{GH} - w_G^h D_{GH}}{\beta_G + w_G^h \left(1 - B_G\right)} = \frac{D_{GH}}{B_G} \ ,$$

where both denominators are assumed different from 0. This implies

$$D_{GH} w_G^h = B_G \delta_{GH} - \beta_G D_{GH} \ .$$

For this to hold for all w_G^h and all G and $H \neq G$ requires $D_{GH} = \delta_{GH} = 0$, since $B_G \neq 0$. Thus in both equations only the own price is included. But turning to the case $G = H$, we see that the restriction is even more severe, because then for all G

$$\frac{\delta_{GG} - w_G D_{GG}}{\beta_G + w_G \left(1 - B_G\right)} = \frac{D_{GG} - 1}{B_G} \ ,$$

which implies

$$\begin{array}{rl}
B_G = & 1 - D_{GG} \ , \\
\beta_G = & -\delta_{GG} \ ,
\end{array}$$

so that in the end there is only one free slope parameter in each equation.

Appendix B. Data

Tables 5 and 6 presents descriptive statistics for the variables used, Table 7 gives the precise definition of goods.

Table 5: **Descriptive statistics (Modelled goods)**

	Budget shares		ln Quantity		ln Unit value	
Good	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Meat	0.250	0.069	5.107	0.425	3.957	0.141
Dairy	0.185	0.045	6.723	0.425	2.052	0.312
Starches	0.116	0.035	6.064	0.437	2.229	0.196
Veg/Fruit	0.081	0.032	5.060	0.507	2.839	0.247
Sweet	0.060	0.022	4.963	0.649	2.651	0.391
Alcohol	0.083	0.052	4.791	1.018	2.980	0.473
Clothes	0.172	0.082	4.022	0.606	4.564	0.457
Shoes	0.053	0.031	1.916	0.760	5.423	0.525

Table 6: **Descriptive statistics (Explanatory variables)**

Variable	Mean	Std Dev
Household characteristics		
Wife's participation	0.734	0.442
Blue collar	0.401	0.490
Farmer	0.234	0.424
Age of head of household	4.361	1.193
Age of hoh squared	0.442	11.415
Owner-occupier	0.394	0.489
No mod-cons	0.169	0.374
Number of. hh members	3.252	1.024
Average age of children	6.210	6.013
Basic education - hoh	0.472	0.499
Advanced education - hoh	0.174	0.379
Rural	0.325	0.469
Space per person	26.392	11.535
ln(Total expenditure)	10.491	0.304
ln(Income)	11.485	0.302
Durable ownership		
Gas supplied	0.525	0.499
Number of leisure durables	4.298	1.849
Freezer	0.568	0.495
Phone	0.351	0.477
Car or motor bike	0.803	0.398
Automatic washing machine	0.775	0.418
Food processor	0.444	0.497
Caravan and/or dacha	0.192	0.394
Garage	0.465	0.499
Conditioning expenditures		
ln(Transport)	7.734	0.674
ln(Hygiene)	8.037	0.487
ln(Food out)	7.817	1.714
ln(Culture)	8.247	0.832
ln(Fuel)	5.316	3.567
ln(Tobaccco)	4.558	3.216
ln(Other food)	8.314	0.513
ln(Textiles)	7.113	1.041
ln(Medical)	5.027	1.828
ln(Furniture)	8.041	1.380
No food out	0.026	0.160
No Tobacco	0.202	0.401
No Medical	0.045	0.207

Table 7: **Description of goods**

Modelled goods	
Meat	pork, beef, other meats and offal, poultry, canned meat, other meat products, fresh and canned fish.
	in kind: pork, other meat and meat products, poultry.
Dairy	butter, margarine, vegetable oil, eggs, bacon and lard, milk, cheese, other milk products.
	in kind: eggs, bacon and lard, milk.
Starches	potatoes, bread, bakery products, wheat flour, other cereal products, rice, pulses.
	in kind: potatoes.
Veg/Fruit	fresh vegetables, frozen vegetable products, fresh fruit, tropical fruit, frozen and dried fruit.
	in kind: fresh vegetables, fresh fruit
Sweet	sugar, chocolate, syrup and concentrates, non-alcoholic beverages.
Alcohol	beer, wine, other alcoholic drinks.
Clothes	cloth or fabric, stockings/socks, knitwear for adults and for children, knit clothes for adults and for children, ready to wear clothing for adults and for children.
Shoes	men's, women's, and children's shoes.
Conditioning goods	
Transport and communication	commuting to work, other public transport, telephone etc.
Hygiene	soaps, detergents, cosmetics, toiletries, laundry, home help, dry cleaning, hairdresser, cosmetics.
Meals out	company canteens, school canteens, restaurants, other catering.
Culture and recreation	books, magazines, toys, culture articles (durable), (non durable), sports equipment, tourist accommodation, flowers, other personal services, education, culture, sports, entertainment, creche, kindergarten, recreation inside Czech Republic, recreation abroad, other services.
Energy	fuel all types, electricity, gas, central heating and other municipal services.
Tobacco	tobacco products.
Other food	confectionery, coffee, tea, ready to cook foods, powdered food, other food.
	in kind: other foods and beverages, free catering
Other textile	textiles, furs, haberdashery, leatherware, tailoring services.
Housing	rent, maintenance, cooperative flat payment.
Medical	medicines and health care goods, medical treatment.
Furniture and equipment	furniture, soft furnishings, glass, porcelain, pots and pans, refrigerators, freezers, electric razors, hairdryers, washing machines, dryers.

References

[1] Ayadi, Mohamed, Rafiq Baccouche, Mohamed Goaied and Mohamed S. Matoussi, 1995. Spatial Variations of Prices and Household Demand Analysis in Tunisia, *paper presented at the 7th World Congress of the Econometric Society, Tokyo.*

[2] Ayadi, Mohamed, Jaya Krishnakumar and Mohamed S. Matoussi, 2002. Pooling Surveys in the Estimation of Income and Price Elasticities: An Application to Tunisian Households, forthcoming, *Empirical Economics.*

[3] Banks, James, Richard Blundell and Arthur Lewbel, 1996. Tax Reform and Welfare Measurement: Do We Need Demand System Estimation?, *Economic Journal*, 106, 1227-1241.

[4] Banks, James, Richard Blundell and Arthur Lewbel, 1997. Quadratic Engel Curves and Consumer Demand, *Review of Economic and Statistics*, 4, 527-539.

[5] Bewley, Ronald, 1986. *Allocation Models: Specification, Estimation and Applications.* Cambridge, Mass.: Ballinger.

[6] Browning, Martin J. and Costas D. H. Meghir, 1991. The Effects of Male and Female Labour Supply on Commodity Demands, *Econometrica*, 59, 925-952.

[7] Browning, Martin J. and Pierre-André Chiappori, 1998. Efficient Intra-Hoursehold Allocations: A General Characterization and Empirical Tests, *Econometrica*, 66, 1241-1278.

[8] Case, Anne, 1991. Spatial Patterns in Household Demands, *Econometrica*, 59(4), 953-65.

[9] Deaton, Angus S., 1987. Estimation of Own and Cross-Price Elasticities from Household Survey Data, *Journal of Econometrics*, 36, 7-3.

[10] Deaton, Angus S., 1988. Quality, Quantity and Spatial Variation of Price, *American Economic Review*, 78, 418-43.

[11] Deaton, Angus S., 1990. Price Elasticities from Survey Data: Extension and Indonesian Results, *Journal of Econometrics*, 44, 281-309.

[12] Deaton, Angus S., 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore: John Hopkins University Press for the World Bank.

[13] Deaton, Angus S. and Franque Grimard, 1991. Demand Analysis for Tax Reform in Pakistan, *Discussion Paper No 151*, Research Programme in Development Studies, Centre for International Studies, Woodrow Wilson School, Princeton.

[14] Deaton, Angus S. and John N. J. Muellbauer, 1980. *Economics and Consumer Behavior*. Cambridge: Cambridge University Press.

[15] De Vreyer, Philippe, 2000, Coûts unitaires et estimation d'un système de demande de travail: théorie et application au cas de Taiwan", *Annales d'Économie et de Statistique*, 57, 211-238.

[16] Kodde, David A., Franz C. Palm and Gerard A. Pfann, 1990. Asymptotic Least-Squares Estimation Efficiency Considerations and Applications, *Journal of Applied Econometrics*, 5, 229-243.

[17] Jorgenson, Dale W., Lawrence J. Lau and Thomas M. Stoker, 1982. The Transcendental Logarithmic Model of Aggregate Consumer Behavior, in Basmann, Robert L. and George F. Rhodes (eds.) *Advances in Econometrics*, Vol.1, 97-238. Greenwich: JAI Press.

[18] Lahatte, Agénor, François Laisney, Ruth Miquel and Ian Preston, 1998. Demand Systems with Unit Values: a Comparison of Two Specifications, *Economics Letters*, 58(3), 281-290.

[19] Lewbel, Arthur, 1989. Identification and Estimation of Equivalence Scales under Weak Separability, *Review of Economic Studies*, 56, 311-316.

[20] Lewbel, Arthur, 1993. Stochastic Hicksian Aggregation with an Application to Grouping Goods Without Separable Utility, *Annales d'Économie et de Statistique*, 29, 17-42.

[21] Lewbel, Arthur, 1996. Aggregation Without Separability: A Generalized Composite Commodity Theorem, *American Economic Review*, 86, 524-543.

[22] Newey, Whitney K., 1984. A Method of Moments Interpretation of Sequential Estimators, *Economics Letters*, 14, 201-206.

[23] Nichèle, Véronique and Jean-Marc Robin, 1995. Simulation of Indirect Tax Reforms Using Pooled Micro and Macro French Data, *Journal of Public Economics*, 56(2), 225-44.

[24] Pashardes, Panos, 1993. Bias in the Estimation of the Almost Ideal Demand System with the Stone Index Approximation, *Economic Journal*, 103, 908-916.

[25] Pitt, Mark M., 1983. Food Preferences and Nutrition in Rural Bangladesh, *Review of Economics and Statistics*, 65, 105-14.

[26] Stone, J. Richard N., 1954. *The Measurement of Consumers' Expenditures and Behaviour in the United Kingdom*, Cambridge: Cambridge University Press.

[27] Strauss, John, 1982. Determinants of Food Consumption in Rural Sierra Leone, *Journal of Development Economics*, 11, 327-53.