Economic Modelling 28 (2011) 1747–1755

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Exploring the effects of aggregation error in the estimation of consumer demand elasticities

Frank T. Denton, Dean C. Mountain *

McMaster University, Canada

ARTICLE INFO

Article history:
Accepted 3 March 2011

JEL classification:
D12
C43

Keywords:
Aggregation error
Consumer demand elasticities
AIDS/QUAIDS models
Income distribution

ABSTRACT

Errors introduced by using aggregate data in estimating a consumer demand model have long been a concern. We study the effects of such errors on elasticity estimates derived from AIDS and QUAIDS models. Based on a survey of published articles, a generic parameterization of the income distribution, and the range of Gini coefficients reported for 28 OECD countries, we generate and analyze a large number of "observations" on the differences between elasticities calculated at the aggregate level and those calculated at the micro level. We suggest a procedure for evaluating the likely range of aggregation error when a model is estimated with aggregate data.

1. Introduction

In earlier days there was frequently no alternative to the use of aggregate time series data in estimating consumer demand models (Stone, 1954, to cite an early pioneering paper). Underlying the models was the notion of a representative consuming unit that maximized utility but aggregation blurred the relationship between micro theory and econometric practice. The likelihood of "aggregation bias" was well known but there was not much one could do about it. Later, as survey data for individual households became increasingly available (and increasingly rich in content), opportunities opened up for estimating micro-theoretic models using actual micro data. Nevertheless, it remains true today that micro data are not always available in particular contexts, or appropriate for particular research objectives. Survey data may be available in one country but not another, or available for broad categories of goods but not at a detailed level that may be required (food in total but not types of food, for example); a survey may fail entirely to provide certain variables of importance for a particular purpose; trends and dynamics may be of interest, thus necessitating the use of time series available only at the aggregate level. Whatever the reasons it is still the case that aggregate data are often used in estimating consumer demand models, and hence that aggregation bias remains on the list of concerns (we report on a survey of 21 articles containing estimated models; 15 of the articles used aggregate data). Other things equal (and sampling variability aside), elasticities calculated at the aggregate level will generally differ from those calculated at the micro level, even if the same model is used in both cases. The differences, how to calculate them, and what to do about them, are the subjects addressed in this paper.

We restrict our attention to two widely used models, Deaton and Muellbauer's (1980) "almost ideal demand system" (AIDS) and the quadratic extension of it (QUAIDS) proposed by Banks et al. (1997). Aggregation of an AIDS micro model over households requires the introduction of an "aggregation parameter" that depends on the distribution of household total expenditure — on the "income distribution," as we shall call it for convenience, with slight inaccuracy; aggregation of a QUAIDS model requires two such parameters. We consider expenditure elasticities and own-price elasticities in the paper and there is, for each, a micro form and a corresponding macro form. This allows us to do a search for articles with AIDS/QUAIDS models that provide either micro or macro elasticities, calculate the corresponding macro or micro elasticities (under alternative assumptions about income distribution), and thus create a data set reflective of the types and magnitudes of aggregation effects actually found in the empirical literature. Along the way we introduce some procedures for characterizing the income distribution in a generic form and (using data for OECD countries) establish a range of distributions according to degree of inequality. On that basis we are then able to arrive at what we think is a reasonable range for the aggregation parameters and study the effects on elasticities over that range.

2. The models at the micro level

Assume K households, indexed by k, and I commodities, indexed by i (or by j if a supplementary index is required). Households face

* Corresponding author. Tel.: +1 905 525 9140x23988.
E-mail address: mountain@mcmaster.ca (D.C. Mountain).

0264-9993/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.econmod.2011.03.010

prices $p_1, \ldots, p_I$. Household k spends x_{ik} on commodity i and x_k on all commodities combined. Its expenditure share for commodity i is therefore $w_{ik} = x_{ik}/x_k$. AIDS and QUAIDS models are both of interest but AIDS is nested within QUAIDS , and so we focus on the latter. The QUAIDS model at the micro level is defined as follows:

$$w_{ik} = \alpha_i + \sum_{j=1}^{I} \gamma_{ij}\, lnp_j + \beta_i\, ln(x_k/q) + \lambda_i(\, ln(x_k/q))^2 / Q \qquad (\forall i, k) \tag{1}$$

$$q = \exp\left\{ \sum_{i=1}^{I} \alpha_i\, lnp_i + 1/2 \sum_{i=1}^{I} \sum_{j=1}^{I} \gamma_{ij}(\, lnp_i)(\, lnp_j) \right\} \tag{2}$$

$$Q = \exp\left\{ \sum_{i=1}^{I} \beta_i\, lnp_i \right\} \tag{3}$$

The corresponding AIDS model is obtained by setting $\lambda_i = 0$ ($\forall i$), and consequently dropping Eq. (3). Eq. (1) might (and typically would) have additional linear terms representing household demographic characteristics, region of residence, etc., in which case the intercepts would be household-specific (α_{ik} rather than α_i). However, that would have no fundamental bearing on the theoretical analysis, and we ignore it, for the moment.

There are different approaches to estimation. One is to simplify things by using approximations to q and Q, rather than the strict specifications of Eqs. (2) and (3): Deaton and Muellbauer (1980) used Stone's index to approximate q in estimating their AIDS model; Matsuda (2006) conducted experiments using the Stone index for q and Tornqvist, Laspeyres, and Paasche indexes as alternatives for Q, in estimating a QUAIDS model. A second approach is to retain the original specifications and use an iterative method: initial parameter values are chosen so as to obtain initial values of q and Q; Eq. (1) are then estimated, conditional on the initial values of those variables, thus obtaining new parameter values, and hence new calculated values of q and Q; and so the process goes until some convergence criterion is satisfied. This method was employed by Banks et al. (1997), Blundell and Robin (1999), and Denton et al. (1999) (an additional level of iteration was included in the latter paper to allow for serial correlation in the error term, following Beach and MacKinnon, 1979). A third approach is to substitute Eqs. (2) and (3) into Eq. (1) and estimate the combined system of equations by some appropriate constrained nonlinear method. The resulting system can be quite large and complex and we are not aware of any published study in which this approach was actually used. We consider it further below, from the point of view of identification in the context of estimation with aggregate data. Whatever the approach taken the model would likely be estimated under theoretical restrictions on its parameters (homogeneity, symmetry), using a Zellner-type estimator.

3. The models at the aggregate level

Consumer demand models can also be estimated with macro data. That this may introduce aggregation error is a longstanding worry, assuming one wishes to interpret the estimates as applying to the underlying micro model (Gorman, 1953; Stoker, 1984, 1986; Blundell and Stoker, 2005).

The macro version of QUAIDS consistent with Eq. (1) is obtained as follows. Let X_i be aggregate expenditure on commodity i, X overall aggregate expenditure, $W_i = X_i/X$ the aggregate expenditure share, and $\bar{x} = X/K$ mean expenditure per household. Eq. (1) can then be rewritten as

$$\begin{aligned} w_{ik} = \alpha_i + \sum_{j=1}^{I} \gamma_{ij}\, lnp_j + \beta_i(\, ln(x_k/\bar{x}) + \, ln(\bar{x}/q)) \\ + \lambda_i(\, ln(x_k/\bar{x}) + \, ln(\bar{x}/q))^2 / Q \end{aligned} \tag{4}$$

and for a household for which $x = \bar{x}$, as

$$w_{ik} = \alpha_i + \sum_{j=1}^{I} \gamma_{ij}\, lnp_j + \beta_i\, ln(\bar{x}/q) + \lambda_i(\, ln(\bar{x}/q))^2 / Q \tag{5}$$

The aggregate share equation corresponding to the micro share Eq. (5) is obtained by multiplying both sides of Eq. (4) by x_k/X and summing over k:

$$W_i = \alpha_i^* + \sum_{j=1}^{I} \gamma_{ij}\, lnp_j + \beta_i^*\, ln(\bar{x}/q) + \lambda_i(\, ln(\bar{x}/q))^2 / Q \tag{6}$$

where:

$$a_i^* = \alpha_i + \beta_i g + \lambda_i h / Q; \qquad \beta_i^* = \beta_i + 2\lambda_i g / Q;$$
$$g = \Sigma(x_k/X)\, ln(x_k/\bar{x}); \qquad h = \Sigma(x_k/X)(\, ln(x_k/\bar{x}))^2$$

Eqs. (2) and (3) still hold at the macro level, with α_i and β_i replaced by α_i^* and β_i^*.

Two new parameters, g and h, now appear in the macro equation. In theory, both are identifiable. To see this, consider the QUAIDS model with two commodities ($i = 1,2$). Dropping the equation for the second commodity to avoid singularity of the system in estimation, substituting Eqs. (2) and (3) into Eq. (6), and imposing homogeneity and symmetry restrictions, we obtain

$$\begin{aligned} W_1 = (\alpha_1 + \beta_1 g) + \beta_1\, ln\bar{x}^* + (\gamma_{11} - \beta_1\alpha_1)\, lnp_1^* \\ -(\beta_1\gamma_{11}/2)(\, lnp_1^*)^2 + \lambda_1 h \exp\{-\beta_1\, lnp_1^*\} \\ + 2\lambda_1 g\, ln(\bar{x}/q)\exp\{-\beta_1\, lnp_1^*\} \\ + \lambda_1(\, ln(\bar{x}/q))^2 \exp\{-\beta_1\, lnp_1^*\}\exp\{-\beta_1\, lnp_1^*\} \end{aligned} \tag{7}$$

where $p_1^* = p_1/p_2$, $\bar{x}^* = \bar{x}/p_2$, and q involves only the parameters α_1 and γ_{11}. A path to the determination of g and h is the following. β_1 is estimated directly when Eq. (7) is fitted to the data and γ_{11} can then be calculated immediately. Given β_1 and γ_{11}, α_1 can be calculated, and then g. Given g, λ_1 can be calculated, and given λ_1, h can be calculated (in a larger system g and h would be restricted to being the same in all equations and there would be other paths to their determination). One would like to exploit this identification property but unfortunately it is almost certain to be too weak to be useful, a fact confirmed by some experimentation with actual and artificial data. In the absence of other information the parameter estimates are too sensitive to small sampling errors to make them acceptable. This remains true even if the quadratic term is dropped, thus eliminating h and converting the model to the AIDS form. As a practical matter it is just not possible to extract reliably the $\beta_i g$ component from the combined intercept term $\alpha_i + \beta_i g$ and subsequent calculations of elasticities are almost certain to be unreliable. We therefore take a different approach, one that is likely to produce results at least within a reliable range.

4. Micro and macro elasticities

We are interested in the effects of aggregation on calculated elasticities. To simplify what follows (without loss of generality) we normalize prices and incomes (as in Denton and Mountain, 2001, 2004) so that $\ln p_i = 0$ for all i (hence $q = Q = 1$) and $ln\bar{x} = 0$. The elasticities are invariant to the normalization, which amounts simply to a particular choice of measurement units. Note too that it has no effect on g and h; they are invariant to the scaling of income — to what Lewbel (1990, 1992) terms "mean scaling" (as an aside, the mean scaling property also contributes to the justification for assuming g and h to be constant when the income distribution changes and a model is estimated with aggregate time series data, just as the other