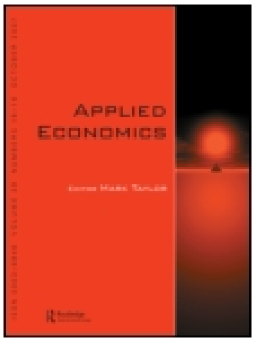

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: http://www.tandfonline.com/loi/raec20

Empirical modelling of the aggregation error in the representative consumer model

Camilo Sarmiento & Richard E. Just

To cite this article: Camilo Sarmiento & Richard E. Just (2005) Empirical modelling of the aggregation error in the representative consumer model, Applied Economics, 37:10, 1163-1175, DOI: 10.1080/00036840500123101

To link to this article: https://doi.org/10.1080/00036840500123101

Published online: 16 Aug 2006.

Submit your article to this journal

Article views: 29

View related articles

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=raec20

Applied Economics, 2005, **37**, 1163–1175

Routledge
Taylor & Francis Group

Empirical modelling of the aggregation error in the representative consumer model

Camilo Sarmiento[a,*] and Richard E. Just[b]

[a]*Pacific Institute for Research and Evaluation, 11710 Beltsville Drive, Suite 300, Calverton, MD 20705, USA*
[b]*University of Maryland, College Park, USA*

This paper examines different approaches to modelling the aggregation error associated with the representative consumer model. Each approach is based on an analytical framework intended for modelling aggregate time series data on quantities and prices with potential additional measures of income distribution. Simple functions that track aggregation error over time are found to perform better than more complex and theoretically sophisticated models. An explanation is given based on typical time series characteristics of economic data.

I. Introduction

Much progress has been made over the past 25 years on aggregation theory in consumption. Muellbauer (1975, 1976) extended the Gorman (1961) polar form to a non-linear function of income to obtain the price independent generalized linear (PIGL) and price independent generalized logarithmic (PIGLOG) functional forms. The almost ideal demand system (AIDS) of Deaton and Muellbauer (1980) implements Muellbauer's results to produce demands with a linear term in the logarithmic of income. Shortly after the article by Deaton and Muellbauer, in a remarkable and elegant contribution, Gorman (1981) derived the complete class of functional forms for non-linear aggregable demand models.[1] In applied demand analysis however, the most common assumption for modelling aggregate Marshallian demands is still the representative consumer model, which treats aggregate mean income as if modelling individual demands (e.g., De Boer *et al.*, 2000; Duffy, 2001; Karagiannis and Velentzas, 2004).

This paper models the aggregation error associated with the representative consumer model using simple additive functions to reduce bias and improve estimation and inference. To do so, the paper first considers integrability conditions in the representative consumer model and then shows that an additive function used to model aggregation error does not necessarily affect integrability properties. Therefore, a representative consumer model satisfying integrability conditions can incorporate an *ad hoc* intercept correction term to capture aggregation error. Two classes of models are then considered. The first class approximates aggregation error with a function that is conceptually motivated by the form of the representative consumer model and thus has cross-coefficient restrictions with it. Examples are parametric assumptions on the income distribution and Taylor series-based approximations using the model form. The second

*Corresponding author. E-mail: csarmiento@pire.org

[1] Lewbel (1990) shows that quadratic generalizations of the PIGL and PIGLOG systems attain Gorman's (1981) standard of generality and that these generalizations are analogous to the quadratic expenditure system developed by Howe *et al.* (1979) and perfected by van Daal and Merkies (1989).

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2005 Taylor & Francis Group Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00036840500123101

class alternatively approximates the aggregation error with *ad hoc* functions whose parameters are unconstrained by the representative consumer model (e.g., polynomial time trends). Each approximation is examined by simulation using a structure in which homotheticity fails. Specifically, the empirical analysis of each approximation is examined by simulation using the structure most commonly adopted in applied demand analysis, i.e., the AIDS.[2]

Comparison of different approaches to modelling aggregation error suggests some useful principles for empirical work. First, modelling aggregation error is useful for reducing biases. Second, and more surprisingly, simplicity is found to work best. For example, methods that attempt to refine approximations of the aggregation error at a given point in time based on conceptual manipulation of the particular representative consumer model are found to be too rigid in explaining variation of aggregation error across time; these approximations while conceptually accurate at a given point in time are less flexible in tracking aggregation error across time. The best models are found to be the simplest *ad hoc* forms that track aggregation error without coefficient restrictions. Interestingly, this principle for modelling aggregation error coincides with a forecasting principle whereby the simplest models often forecast best (Allen and Fildes, 2001).

II. Theoretical Basis of the Representative Consumer Model

Convex preferences and utility maximization imply that the ith consumer's demand for the kth good at time t with a price vector, $\mathbf{p}_t = (p_{1t}, p_{2t}, \ldots, p_{mt})$, and income, M_{it}, is represented as

$$x_{it}^k = f^k(\mathbf{p}_t, M_{it}) \tag{1}$$

The corresponding aggregate mean demand for the kth good at time t is

$$\underline{x}_t^k = \frac{\sum_i x_{it}^k}{n} = \frac{\sum_i f^k(\mathbf{p}_t, M_{it})}{n} \tag{2}$$

A restrictive form of Equation 2 commonly used in empirical work is the representative consumer model, which approximates aggregate mean demands as if modelling individual demands,

$$\underline{x}_t^k \cong f^k(\mathbf{p}_t, \underline{M}_t) \tag{3}$$

where the aggregate mean income is $\underline{M}_t = \Sigma_i M_{it}/n$.

From economic theory, the integrability conditions (symmetry, homogeneity and summability) applicable to Equation 1 are inherited by Equation 2 (Jorgenson, 1997). These properties provide useful structure for demand analysis based on economic theory (Gorman, 1981). For example, the property of symmetry guarantees uniqueness (path invariance) of common welfare measures (Silberberg, 1972). The following proposition establishes that integrability restrictions applicable to Equation 2 are also inherited by Equation 3.

Proposition 1. Assuming second-order smoothness conditions with standard consumer utility maximization, the integrability conditions of the unrestricted aggregate mean demand model in Equation 2 are inherited by the representative consumer model in Equation 3.

Proof: See the Appendix.[3]

The result in Proposition 1 provides a theoretical basis for the representative consumer model. A subtle implication is that an *ad hoc* intercept correction term that does not interact with integrability restrictions can be used to model aggregation error of the representative consumer model without destroying the appealing benefits derived from integrability. While applicability of integrability to the representative consumer model is clear from Proposition 1, the issue of estimating aggregation bias and which function to use for that purpose remains.

III. Specifying Functions to Model Aggregation Error

Turning now to representation of aggregation error, which is a source of bias in representative consumer models, this section considers alternative functions that can be used to approximate aggregation error.

Consider the zero-order Taylor series approximation of Equation 2 around mean income, which yields the representative consumer model defined by Equation 3. The error of the approximation is

$$h_t^k = \frac{\sum_i f^k(\mathbf{p}_t, M_{it})}{n} - f^k(\mathbf{p}_t, \underline{M}_t)$$

An immediate possibility for approximating h_t^k is to use the second and perhaps several additional terms of the Taylor series expansion of Equation 2 around mean income, i.e.,

$$h_t^k = \frac{\sum_j H_j^k(\mathbf{p}_t, \underline{M}_t) \sum_i (M_{it} - \underline{M}_t)^j}{nj!} \tag{4}$$

[2] See Anderson and Blundell (1983), Duffy (2001), Alston *et al.* (2002), De Mello *et al.* (2002).
[3] This proposition can be extended to obtain an analogous result for the case of heterogeneous preferences (see the Appendix).

where $\mathrm{H}_j^k = (\partial^j f^k(\mathbf{p}_t, M_{it})/\partial M_{it}{}^j | M_{it} = \underline{M}_t)$. Note that H_t^k in Equation 4 and $f^k(p_t, \underline{M}_t)$ will generally have common coefficients thus requiring cross coefficient restrictions in estimation.

Another approach for approximating the difference in Equations 2 and 3 is to assume a specific parametric distribution of income. Then h_t^k can be represented as

$$h_t^k \cong \int_0^\infty f^k(\mathbf{p}_t, M_t)\mathrm{d}G(M_t, \theta) - f^k(\mathbf{p}_t, \underline{M}_t) \qquad (5)$$

where $G(M_t, \theta)$ with parameter vector θ represents a specific income distribution such as the lognormal (Lambert, 1993). This approximation of h_t^k also generally has coefficients in common with $f^k(p_t, \underline{M}_t)$ and thus requires cross-coefficient restrictions for estimation.

A third approach is to approximate h_t^k using quantile data on the income distribution. That is, the income distribution is approximated by representing each consumer's income by the mean income of the consumer's income quantile. Thus h_t^k is approximated by

$$h_t^k \cong \sum\nolimits_j s_j f^k(\mathbf{p}_t, \underline{M}_{jt}) - f^k(\mathbf{p}_t, \underline{M}_t) \qquad (6)$$

where $\underline{M}_{jt} = (\Sigma_{i \in Sj}\, M_{it})$ and $i \in S_j$ denotes the set of consumers included in the jth income quantile. This model is motivated empirically by the fact that most public data on the income distribution appears in quantile form and may not facilitate exact calculation of other summary statistics of the income distribution. In this case as well, cross-coefficient restrictions are required between $f^k(\mathbf{p}_t, \underline{M}_t)$ and h_t^k unless h_t^k is invariant to prices and income.

Each of the above approximations of error (based on a Taylor expansion, an assumed income distribution, or a quantile representation of the income distribution) can be interpreted as an intercept correction term where cross coefficient restrictions are applicable in the representative consumer model. Alternatively, an *ad hoc* proxy without coefficient restrictions can be used to approximate the aggregation error. For example, polynomial trends can be used to track variation of the aggregation error h_t^k across time,

$$h_t^k \cong \sum\nolimits_s a_{ks} t^s \qquad (7)$$

If the polynomial trend sufficiently captures omitted information, then the representative consumer model may be able to draw adequate inferences about aggregate economic behaviour. With this approach, h_t^k can be modelled by a distribution of correction factors over time such that cross-coefficient restrictions with $f^k(\mathbf{p}_t, \underline{M}_t)$ are not necessary for integrability. In addition to approximating h_t^k, a desirable property of this aggregation error correction term is low correlation with $f^k(\mathbf{p}_t, \underline{M}_t)$, which permits better econometric identification.

A potential shortcoming of Equation 7 is the presence of trends in most price indexes used in demand analysis. Therefore, the possibility of representing h_t^k with a term normalized by income level (which is therefore less likely to be correlated with price and income levels) is also considered, e.g.,

$$h_t^k \cong c\left[\frac{\Sigma(M_{it} - \underline{M}_t)^2}{n\underline{M}_t^2}\right] \qquad (8)$$

Note that symmetry and homogeneity impose no constraints on c. An intuitive motivation of Equation 8 is that the coefficient of variation of the income distribution is likely to be correlated with the aggregation error (e.g., the second-order term of the Taylor series expansion) but less likely to be correlated with $f^k(\mathbf{p}_t, \underline{M}_t)$ than a polynomial trend. The simulations in this paper show that the effectiveness of Equations 7 and 8 varies with the data generating process (DGP) of the included variables in the model.

This section suggests that approaches for representing aggregation error fall into two classes. In the first class, aggregation error is approximated with a function that generates cross-coefficient restrictions. This is the case when specific parametric assumptions are imposed on the income distribution and when Taylor approximation terms are used to represent aggregation error. In the second class, the aggregation error is approximated by *ad hoc* correction terms (e.g., polynomial trends) that do not require cross-coefficient restrictions with $f^k(\mathbf{p}_t, \underline{M}_t)$ for integrability (Proposition 1). The remainder of this paper evaluates these alternative approaches for modelling aggregation error by simulation using a popular structure for which homotheticity fails, the almost ideal demand system (Deaton and Muellbauer, 1980).

The almost ideal (AI) demand system (Deaton and Muellbauer, 1980) is probably the most commonly used structure in empirical work, and contains the simplest exact aggregation conditions with non-linear Engle curves. Indeed, as commonly recognized, the aggregation error of the AI model from imposing the representative consumer model is a subset of the error present in more complex exact aggregation structures (Karagiannis and Velentzas, 2004). Although the use of simple structures to illustrate broader implications is subject to criticism, such approaches are often used to illustrate salient points (e.g., Pope and Just, 1996, Clements and Hendry, 1998, Arndt *et al.*, 1999).

IV. Simulation Models

The almost ideal (AI) demand system is typically estimated with budget shares (Deaton and Muellbauer, 1980) based on the individual consumer equation,

$$\frac{f^k(\mathbf{p}_t, M_{it})p_{kt}}{M_{it}} = \varphi_k + \sum_j \beta_{kj} \ln \mathrm{p}_{jt} + \gamma_k \ln M_{it} - \gamma_k \ln P_t \quad (9)$$

where $\ln P_t = \alpha_0 + \Sigma_k\, \varphi_k \ln p_{kt} + (1/2)\Sigma_j \Sigma_k\, \beta_{kj} \times \ln p_{kt} \ln p_{jt}$.[4] Aggregating Equation 9 over individuals yields the aggregate demand corresponding to Equation 2,

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_t} = \varphi_k + \sum_j \beta_{kj} \ln p_{jt} + \gamma_k \frac{\sum_i M_{it} \ln M_{it}}{\sum_i M_{it}} - \gamma_k \ln P_t \quad (10)$$

Using Equation 10 as a standard of comparison, seven models are defined for simulation of aggregation error. Models 1 and 2 are the true and representative consumer models, respectively, which provide standards of comparison. Models 3, 4, and 5 follow the first class of approaches to aggregation error approximation suggested in Equations 4, 5, and 6, and thus introduce terms with specific cross-coefficient relationships to other terms in the demand specification. Models 6 and 7 approximate aggregation error with the second class of approaches in Equations 7 and 8 where parameters are unrestricted.

Model 1. The Exact Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \frac{\sum_i M_{it} \ln M_{it}}{\sum_i M_{it}}$$

where $A(p_t) = \varphi_k + \Sigma_j\, \beta_{kj} \ln p_{jt} - \gamma_k \ln P_t$.

Model 2. The Representative Consumer Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \ln \underline{M}_t$$

Model 3. The Second-Order Aggregation Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \ln \underline{M}_t + \gamma_k \left[\frac{\left(\sum_i M_{it} - \underline{M}_t\right)^2}{2n\underline{M}_t^2}\right]^5$$

This model is motivated by a second-order Taylor approximation around mean income as in Equation 4.

Model 4. The Quantile Approximation Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \frac{\left(\sum_i s_j \underline{M}_{jt} \ln \underline{M}_{jt}\right)}{\underline{M}_t}$$

where $\underline{M}_{jt} = (\Sigma_{i\in S_j} M_{it})/n_j$ and S_j represents the set of all consumers included in the jth quantile.

Model 5. The Lognormal AI Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \ln(\underline{M}_t^2) - \frac{\gamma_k\left[\ln \underline{M}_t^2 + \sum_i (M_{it} - \underline{M}_t)^2/n\right]}{2} + \gamma_k \ln\left[1 + \frac{\sum_i (M_{it} - \underline{M}_t)^2}{n\underline{M}_t^2}\right]$$

This approach assumes the income distribution follows a specific parametric form – the lognormal distribution – which if applicable reduces the number of summary statistics of income distribution required.[6]

[4] The relevance of an exact price index rather than the Stone index in the AI model is discussed by Deaton and Muellbauer (1980) and Pashardes (1993). However, both an exact price index and an exact aggregate income index are required for integrability conditions to hold with aggregate data.

[5] In this model, $\Sigma_i f^k(\mathbf{p}_t, M_{it})/n$ is approximated by $f^k(\boldsymbol{p}_t, \underline{M}_t) + \partial f^k(\mathbf{p}_t, M_{it})/\partial M_{it}|_{\underline{M}}[\Sigma_i(M_{it} - \underline{M}_t)/n] + \partial^2 f^k(\mathbf{p}_t, M_{jt})/\partial M_{it}^2|_{\underline{M}}$ $[\Sigma_i(M_{it} - \underline{M}_t)^2/2n]$ where specifically for the AI model $\partial f^k(\boldsymbol{p}_t, M_{it})/\partial M_{it} = A(\boldsymbol{p}_t)/p_{kt} + \gamma_k \ln M_{it}/p_{kt} + \gamma_k/p_{kt}$ and $\partial^2 f^k(\mathbf{p}_t, M_{it})/\partial M_{it}{}^2 = \gamma_k(1/M_{it})(1/p_{kt})$. Model 3 follows upon noting that $\Sigma_i(M_{it} - \underline{M}_t)/n = 0$.

[6] The lognormal assumption is the most common for modelling income distributions in empirical work. Under lognormality (see Jorgenson, 1997, p. 327),

$$\frac{\sum_i M_{it} \ln M_{it}}{\sum_i M_{it}} = \underline{\ln M_t} + \frac{\sum_i \left(\ln M_{it} - \underline{\ln M_t}\right)^2}{n}$$

where $\ln \underline{M}_t = \Sigma_i \ln M_{it}/n$, in which one can further substitute

$$\underline{\ln M_t} = \ln(\underline{M}_t^2) - \ln\left[\frac{\underline{M}_t^2 + \sum_i (M_{it} - \underline{M}_t)^2/n}{2}\right]$$

and

$$\frac{\sum_i (\ln M_{it} - \underline{\ln M_t})^2}{n} = \ln\left[1 + \frac{\sum_i (M_{it} - \underline{M}_t)^2}{n\underline{M}_t^2}\right]$$

(see Greene, 1993, p. 60). Models 1 and 5 are thus equivalent under lognormality.

Model 6. The Flexible Second-Moment Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \ln \underline{M}_t - b_k\left[\frac{\sum_i (M_{it} - \underline{M}_t)^2}{2n\underline{M}_t^2}\right]$$

This model uses the term $b_k\ [\Sigma_i(M_{it} - M_t)^2/2n\underline{M}_i^2]$ in Equation 8 as a proxy for all higher-order Taylor terms. As such, it enters without coefficient restrictions, which allows the income variance to capture some of the effects of other terms beyond the second.

Model 7. A Time-Flexible Instrumentation Model.

$$\frac{\sum_i f^k(\mathbf{p}_t, M_{it})p_{kt}}{\sum_i M_{it}} = A(\mathbf{p}_t) + \gamma_k \ln \underline{M}_t - \gamma_k \sum_k a_k t^k$$

This approach models the aggregation error without cross coefficients restrictions as in Equation 7 using a flexible polynomial $\Sigma_k a_k t^k$ to correct for all omitted moments of the income distribution.[7]

V. Simulation: Refinements with Coefficient Restrictions

To evaluate the aggregation error and associated biases from each approximation, a simulation and representative disaggregated income and expenditure data from the Panel Survey of Income Dynamics (PSID) for 1968–1992 is used. Data are used on both the distribution of personal disposable income and the distribution of expenditures over time.[8] To assess the likely bias of a specific model compared to the exact model (Model 1), Tables 1 and 2 report the mean and standard deviation of the aggregation error. Table 1 uses the personal disposable income data whereas Table 2 uses food expenditure data.

Analysis of Tables 1 and 2 show that modelling the aggregation error with Models 3–5 successfully refine approximations of the function at a given point in time. But surprisingly, these approximations are unable to track the aggregation error across time better than the representative consumer model. Thus, they are ineffective in reducing aggregation bias with time series data. That is, approximations of the aggregation error at a given point in time appear to be too rigid to explain aggregation error across time.

The lognormal assumption in Model 5 is more accurate at any point in time than either first- or second-order Taylor approximation (compare μ in Table 1). In spite of this accuracy, however, assuming this widely accepted parametric distribution does not capture the variation across time of higher moments (compare *SD* in Table 1). In fact, the approximation of the aggregation error across time from assuming a lognormal distribution (Model 5) is larger than from using the simple representative consumer model of Model 2. The analysis of Models 3 and 4 show similar results for both the distribution of income and the distribution of food expenditure cases.

While the results in Tables 1 and 2 are based on a comparison of regression variables, for further confirmation bias associated with alternative representations of the income distribution by means of simulation is evaluated. To simulate the biases from approximating Model 1, a set of prices is used that attempt to capture typical correlations of price indexes and income found in empirical work. Specifically, US price indexes are used for 1968–1992 on major expenditure groups including food, housing, clothing, transportation, medical care, and entertainment collected from the National Income and Product Accounts (NIPA). The aggregate income indexes of each model were calculated from the distribution of personal disposable income for 1968–1992 from the PSID.

Also biases for a system that is weakly separable with respect to food consumption are simulated using disaggregate US food price indexes for 1968–1992 as collected from NIPA. The aggregate food expenditures indexes for each model were calculated from the distribution of total food consumption expenditures for 1968–1992 from the PSID.

Biases are evaluated by imposing the coefficient restrictions of the AI model under plausible assumptions on certain parameters. That is, a complete share

[7] With heterogeneous tastes, the term $\Sigma_i \gamma_{ik} M_{it} \ln M_{it}/n$ in the AI model is alternatively represented by sample covariance as $(\Sigma_i\ \gamma_{ik}/n)(\Sigma_i\ M_{it}\ \ln M_{it}/n) + \mathrm{Cov}(\gamma_{ik},\ \Sigma_i\ M_{it}\ \ln M_{it})$. Thus, correlation between tastes and income in the AI model is equivalent to time varying intercepts. Thus, polynomial trends may reflect existence of heterogeneous tastes.

[8] For US consumption, the PSID only reports disaggregate data on food expenditures. For the distribution of total expenditures, the Consumer Expenditure Survey (CES) only has comparable data after 1985 (see DeJuan and Seater, 1999). Nevertheless, a broad branch of applications of the AI model considers food consumption, for which weak separability is consistent with some empirical evidence and is widely assumed.

Table 1. Indexes of aggregate personal disposable income from the PSID survey[a]

Year	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7 Order 1	Model 7 Order 2	Model 7 Order 3
1968	9.175	8.849	9.279	8.949	9.159	8.905	9.172	9.171	9.202
1969	9.177	8.872	9.279	8.982	9.170	8.907	9.153	9.152	9.168
1970	9.193	8.910	9.272	8.978	9.182	8.922	9.162	9.162	9.166
1971	9.215	8.933	9.290	9.033	9.202	8.940	9.185	9.185	9.180
1972	9.323	9.035	9.410	9.139	9.315	9.028	9.379	9.379	9.367
1973	9.267	9.000	9.313	9.080	9.243	8.984	9.245	9.245	9.230
1974	9.328	9.066	9.379	9.157	9.309	9.034	9.345	9.345	9.328
1975	9.309	9.040	9.368	9.135	9.292	9.018	9.286	9.286	9.269
1976	9.543	9.147	9.582	9.251	9.460	9.212	9.732	9.732	9.716
1977	9.394	9.123	9.415	9.198	9.353	9.090	9.413	9.413	9.400
1978	9.392	9.073	9.480	9.136	9.371	9.085	9.386	9.387	9.377
1979	9.350	9.078	9.498	9.181	9.383	9.046	9.280	9.280	9.275
1980	9.405	9.081	9.907	9.201	9.569	9.070	9.370	9.370	9.370
1981	9.397	9.089	9.592	9.203	9.437	9.082	9.332	9.332	9.336
1982	9.447	9.113	9.833	9.240	9.559	9.111	9.410	9.410	9.419
1983	9.604	9.256	9.751	9.409	9.600	9.255	9.703	9.703	9.716
1984	9.517	9.189	9.713	9.323	9.547	9.181	9.506	9.506	9.522
1985	9.543	9.223	9.685	9.270	9.550	9.205	9.536	9.536	9.553
1986	9.538	9.227	9.651	9.357	9.534	9.203	9.505	9.505	9.523
1987	9.584	9.253	9.788	9.397	9.617	9.235	9.576	9.576	9.592
1988	9.621	9.292	9.821	9.437	9.653	9.266	9.628	9.628	9.640
1989	9.698	9.310	9.910	9.525	9.704	9.329	9.761	9.761	9.767
1990	9.518	9.184	9.657	9.333	9.517	9.185	9.379	9.379	9.376
1991	9.702	9.300	10.106	9.488	9.780	9.320	9.725	9.725	9.710
1992	9.770	9.358	9.997	9.358	9.770	9.387	9.839	9.838	9.808
μ[b]	0.000	0.334	0.165	0.215	0.033	0.334	0.045	0.045	0.043
R	0.000	1.000	0.494	0.644	0.099	1.000	0.011	0.011	0.011
SD	0.000	0.045	0.120	0.051	0.050	0.037	0.023	0.023	0.008
W	0.000	1.000	2.640	1.133	1.092	0.824	0.504	0.516	0.176

Notes: [a] Column i reports z_i where $z_1=(\Sigma_i M_{it}\ln M_{it})/(\Sigma_i M_{it})$ where M_{it} is personal disposable income; $z_2=\ln\underline{M}_t$ where $\underline{M}_t$ is the mean personal disposable income; $z_3=\ln\underline{M}_t+\Sigma_i(M_{it}-\underline{M}_t)^2/2n\underline{M}_t^{\,2}$, $z_4=(\Sigma_j s_j\underline{M}_{jt}\ln\underline{M}_{jt})/\underline{M}_t$ where $\underline{M}_{jt}=(\Sigma_{i\in Sj}M_{it})/n_j$ are the quantile mean incomes; and $z_5=\underline{\ln M_t}+\Sigma_i(\ln M_{it}-\underline{\ln M_t})^2/n$ where $\underline{\ln M_t}=\ln(\underline{M}_t^{\,2})-\ln[\underline{M}_t^{\,2}+\Sigma_i(M_{it}-\underline{M}_t)^2/n]/2$ and $\Sigma_i(\ln M_{it}-\underline{\ln M_t})^2/n=\ln[1+\Sigma_i(M_{it}-\underline{M}_t)^2/n\underline{M}_t^{\,2}]$. The Model 6 column reports predictions from the regression of z_1 on z_2 and $z_6=[\Sigma_i\overline{(M_i-\underline{M})^2/2n\underline{M}^2}]$; and the Model 7, Order k, column reports predictions from the regression of z_1 on z_2 and $z_7=\sum_{j=1}^{k}a_j t^j$ where t is time.

[b] Note that μ is the average absolute difference of the Model 1 column and the specified model column over 1968–1992; R is μ normalized by the average value of the representative consumer model (Model 2); SD is the standard deviation of variation across time of the difference; and W is SD normalized by the average value of the representative consumer model (Model 2).

equation system with cross-equation restrictions for seven goods is defined as follows.

$$
\begin{aligned}
\begin{bmatrix} \sum_i p_{1t}f^1(p_t,M_{it})/\sum_i M_{it} \\ \sum_i p_{2t}f^2(p_t,M_{it})/\sum_i M_{it} \\ \vdots \\ \sum_i p_{6t}f^6(p_t,M_{it})/\sum_i M_{it} \end{bmatrix}
&= \alpha_1 \begin{bmatrix} -\gamma_1 \\ -\gamma_2 \\ \vdots \\ -\gamma_6 \end{bmatrix}
+ \varphi_1 \begin{bmatrix} 1-\gamma_1\ln(p_{1t}/p_{7t}) \\ -\gamma_2\ln(p_{1t}/p_{7t}) \\ \vdots \\ -\gamma_6\ln(p_{1t}/p_{7t}) \end{bmatrix}
+ \cdots + \varphi_6 \begin{bmatrix} -\gamma_1\ln(p_{1t}/p_{7t}) \\ -\gamma_2\ln(p_{1t}/p_{7t}) \\ \vdots \\ 1-\gamma_6\ln(p_{1t}/p_{7t}) \end{bmatrix} \\
&\quad + B_{11} \begin{bmatrix} \ln(p_{1t}/p_{7t})-\gamma_1\ln(p_{1t}/p_{7t})^2 \\ -\gamma_2\ln(p_{1t}/p_{7t})^2 \\ \vdots \\ -\gamma_6\ln(p_{1t}/p_{7t})^2 \end{bmatrix} + \cdots \\
&\quad + \mathrm{B}_{12} \begin{bmatrix} \ln(p_{1t}/p_{7t})-\gamma_1\ln(p_{1t}/p_{7t})\ln(p_{2t}/p_{7t}) \\ -\gamma_2\ln(p_{1t}/p_{7t})\ln(p_{2t}/p_{7t}) \\ \vdots \\ -\gamma_6\ln(p_{1t}/p_{7t})\ln(p_{2t}/p_{7t}) \end{bmatrix}
\end{aligned}
\tag{11}
$$

Table 2. Indexes of aggregate food consumption from the PSID[a]

Year	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7 Order 1	Model 7 Order 2	Model 7 Order 3
1968	7.598	7.399	7.591	7.407	7.561	7.409	7.588	7.653	7.619
1969	7.327	7.141	7.272	7.144	7.258	7.257	7.288	7.332	7.319
1970	7.387	7.148	7.283	7.147	7.267	7.362	7.379	7.407	7.407
1971	7.393	7.132	7.282	7.131	7.263	7.357	7.386	7.399	7.410
1972	7.462	7.186	7.402	7.189	7.366	7.295	7.447	7.448	7.464
1973	7.506	7.265	7.407	7.264	7.390	7.469	7.431	7.423	7.441
1974	7.592	7.316	7.496	7.314	7.470	7.505	7.529	7.513	7.531
1975	7.490	7.238	7.380	7.236	7.363	7.464	7.383	7.358	7.375
1976	7.763	7.157	7.359	7.123	7.327	7.957	7.990	7.959	7.973
1977	7.683	7.148	7.353	7.153	7.333	7.692	7.693	7.688	7.699
1978	7.562	7.158	7.344	7.148	7.316	7.591	7.565	7.531	7.539
1979	7.508	7.076	7.274	7.063	7.243	7.536	7.519	7.482	7.485
1980	7.560	7.052	7.257	7.028	7.224	7.648	7.626	7.589	7.586
1981	7.523	7.015	7.213	6.990	7.182	7.628	7.568	7.532	7.524
1982	7.576	6.989	7.214	6.948	7.174	7.701	7.678	7.647	7.634
1983	7.576	6.983	7.230	6.944	7.184	7.657	7.663	7.637	7.620
1984	7.584	7.000	7.305	6.965	7.238	7.529	7.639	7.621	7.602
1985[a]	7.518	6.995	7.212	7.012	7.175	7.596	7.490	7.481	7.462
1988	7.477	6.973	7.181	6.944	7.147	7.556	7.408	7.410	7.394
1989	7.533	6.984	7.241	6.955	7.191	7.548	7.487	7.502	7.492
1990	7.555	6.965	7.303	6.930	7.223	7.432	7.529	7.558	7.558
1991	7.623	7.002	7.323	6.955	7.249	7.570	7.605	7.651	7.665
1992	7.605	7.022	7.272	6.953	7.225	7.670	7.526	7.591	7.623
μ[b]	0.000	0.433	0.222	0.442	0.258	0.080	0.054	0.047	0.043
R	0.000	1.000	0.513	1.021	0.596	0.185	0.124	0.109	0.100
SD	0.000	0.137	0.100	0.175	0.109	0.084	0.072	0.064	0.063
W	0.000	1.000	0.730	1.281	0.799	0.616	0.528	0.468	0.459

Notes: [a]See Table 1, footnote a. Note that the years 1986 and 1987 are excluded because of lack of availability of the distribution of total expenditure on food across individuals from the PSID for these years.
[b]See Table 1, footnote b.

$$+\cdots+\mathrm{B}_{66}\begin{bmatrix}-\gamma_1\ln(p_{1t}/p_{7t})^2\\ -\gamma_2\ln(p_{2t}/p_{7t})^2\\ \vdots\\ \ln(p_{6t}/p_{7t})-\gamma_6\ln(p_{6t}/p_{7t})^2\end{bmatrix}$$

$$+\cdots+\begin{bmatrix}\gamma_1\sum_i M_{it}\ln M_{it}/\sum_i M_{it}\\ \gamma_2\sum_i M_{it}\ln M_{it}/\sum_i M_{it}\\ \vdots\\ \gamma_6\sum_i M_{it}\ln M_{it}/\sum_i M_{it}\end{bmatrix}.$$

From Equation 11, a linear projection suffices to evaluate biases if $\gamma_1,\ldots,\gamma_6$ are given. Thus, to measure the biases the omitted aggregation error is projected onto the explanatory variables of Equation 11 using the approximate index of aggregate income, e.g., $\ln\underline{M}_t$ for Model 2. Table 3 is constructed under the assumption that $\gamma_k=-0.05$ for the income inelastic goods (food, housing, medical care) and $\gamma_k=0.05$ for the income elastic goods (clothing, transportation, and entertainment).[9] Table 4 is similarly constructed for a system with weak separability with respect to food consumption using disaggregated prices of food assuming $\gamma_i=-0.05$ for food purchased for off-premise consumption, purchased meals and beverages, and food purchased by employees; and $\gamma_i=0.05$ for tobacco products, alcoholic beverages purchased for off-premise consumption, and other alcoholic beverages.

The simulation results show that both the log-normal assumption and second-order Taylor series approximation, which promise better approximations on conceptual grounds, fail in some cases to capture as much aggregation bias as the simple representative consumer model. For example, the average absolute bias of coefficient estimates for Model 2 is lower than

[9] In the AI model, the income elasticity of the kth good is $1+(\gamma_k/s_k)$ where s_k is the share of the good. Therefore, if γ_k is -0.09 and s_k is 0.1, then the income elasticity is 0.01.

Table 3. Biases from approximations of the distribution of personal disposable income[a]

Biases	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7 Order 1	Model 7 Order 2	Model 7 Order 3
α_1	37.928	187.946	85.039	63.263	19.571	23.741	−44.446	−54.992
φ_1	2.512	15.310	6.670	5.303	1.305	2.401	−2.599	−3.323
φ_2	2.516	15.273	6.672	5.286	1.300	2.399	−2.606	−3.325
φ_3	2.506	15.334	6.676	5.297	1.292	2.390	−2.598	−3.319
φ_4	−1.274	−3.458	−1.818	−1.028	−0.660	0.023	1.844	2.173
φ_5	−1.287	−3.764	−1.857	−1.012	−0.656	0.017	1.831	2.166
φ_6	−1.289	−3.466	−1.827	−1.019	−0.658	0.017	1.833	2.172
β_{11}	−0.415	−0.636	−0.455	−0.180	−0.242	−0.383	0.155	0.299
β_{12}	−0.357	−0.935	−0.606	−0.258	−0.194	−0.316	0.289	0.339
β_{13}	−0.036	−0.478	−0.178	−0.139	0.002	−0.031	0.148	0.176
β_{14}	0.015	0.078	0.045	−0.048	−0.073	−0.042	−0.026	−0.008
β_{15}	0.142	0.167	0.198	−0.220	−0.063	0.067	0.114	0.092
β_{16}	−0.032	0.077	−0.030	0.160	0.113	0.051	−0.021	−0.034
β_{22}	−0.168	−1.076	−0.243	−0.257	−0.134	−0.209	0.277	0.285
β_{23}	−0.041	−0.470	−0.185	−0.137	0.001	−0.033	0.150	0.178
β_{24}	0.091	0.505	0.381	0.168	0.010	0.030	−0.021	−0.052
β_{25}	−0.015	0.618	0.025	−0.076	−0.062	0.008	0.140	0.131
β_{26}	−0.144	−0.433	−0.495	−0.099	0.009	−0.045	−0.011	0.027
β_{33}	−0.448	−2.407	−1.090	−0.875	−0.261	−0.434	0.314	0.420
β_{34}	0.134	0.638	0.294	0.244	0.088	0.134	−0.055	−0.080
β_{35}	0.137	0.627	0.301	0.247	0.089	0.134	−0.058	−0.084
β_{36}	0.139	0.653	0.310	0.245	0.086	0.134	−0.054	−0.081
β_{44}	−0.173	−0.252	−0.031	0.015	−0.037	−0.015	0.135	0.209
β_{45}	−0.268	−0.513	−0.393	−0.246	−0.137	−0.035	0.352	0.421
β_{46}	−0.179	−0.802	−0.589	−0.328	−0.131	−0.047	0.277	0.289
β_{55}	−0.153	−1.046	−0.342	0.111	0.086	−0.038	−0.103	−0.058
β_{56}	0.021	−0.078	0.173	−0.323	−0.225	0.034	0.423	0.402
β_{66}	−0.155	0.017	0.197	0.159	−0.001	0.003	0.098	0.176
γ_I	−0.029	−0.626	−0.235	−0.224	−0.160	−0.103	0.072	0.094
μ_1[b]	1.814	8.886	4.047	2.999	0.953	1.149	2.105	2.600
μ_2	0.524	2.491	1.154	0.847	0.288	0.342	0.593	0.729

Notes: [a]For evaluating the biases, each of the price indexes are transformed so as to use Equation 3 as the AI model representation. This table assumes $\gamma_i = -0.05$ for food, household operations, and medical care and $\gamma_i = 0.05$ for clothing, transportation, and recreation. Using definitions from Table 1, footnote a, the column for model i reports $(x'x)^{-1}x'w_i$ where the matrix x includes the regression variables associated with the structure in Equation 3 from 1968 to 1992 for given $(\gamma_1, \ldots, \gamma_6)$ considering the NIPA price indexes of food, household operations, medical care, clothing, transportation, and recreation equations where the price index of all other goods is used as the numeraire; $w_2 = z_1 - z_2$, $w_3 = z_1 - z_3$, $w_4 = z_1 - z_4$, and $w_5 = z_1 - z_5$. For Model 6 the x matrix is augmented by z_6 and $w_6 = z_1 - z_2$; and for Model 7, Order k, the x matrix is augmented by z_7 and $w_7 = z_1 - z_2$.
[b]Note that $\mu_1 = \Sigma_j|\delta_{jq}|/m$ where δ_{jq} is the bias in the jth coefficient in column q; and μ_2 is the sum of the biases excluding the constant term α_1.

Models 3, 4 or 5 in Table 3. In Table 4, which is based on separability of food expenditures, including second-order terms or the exact lognormal parametric form improves somewhat upon the representative consumer model (compare Models 3 and 5 to Model 2).

While the simulation results of this section are subject to specific assumptions regarding the elasticities, a range of simulation results for other elasticity assumptions have been compiled with $\gamma_k = -0.07$ ($\gamma_k = -0.03$) for the income inelastic goods (food, housing, medical care) and $\gamma_k = 0.03$ ($\gamma_k = 0.07$) for the income elastic goods (clothing, transportation, and entertainment) and similarly for the food expenditure data.[10] These variations are not sufficient to cause substantive changes in the ranking of models with respect to income aggregation bias.

Using the personal disposable income data, Model 2 is best among Models 2–5 in every case

[10] Detailed tables such as Tables 3 and 5 for cases with $\gamma_k = -0.07$ ($\gamma_k = -0.03$) for the income inelastic goods (food, housing, medical care) and $\gamma_k = 0.03$ ($\gamma_k = 0.07$) for the income elastic goods (clothing, transportation, and entertainment) are available upon request.

Table 4. Biases from approximations of the distribution of food expenditures[a]

Biases	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7 Order 1	Model 7 Order 2	Model 7 Order 3
α_1	473.369	366.547	567.537	404.657	206.235	−22.166	−10.818	−40.542
φ_1	27.335	20.918	44.562	22.832	13.575	0.169	0.726	−0.839
φ_2	27.112	20.721	44.336	22.608	13.457	0.098	0.654	−0.907
φ_3	27.268	20.806	44.510	22.719	13.472	0.047	0.618	−0.928
φ_4	−19.936	−15.826	−12.083	−17.674	−7.189	2.216	1.657	3.036
φ_5	−20.048	−15.943	−12.199	−17.737	−7.181	2.296	1.732	3.127
φ_6	−19.894	−15.652	−12.106	−17.545	−6.981	2.525	1.950	3.362
β_{11}	−1.276	−0.959	−2.881	−1.092	−0.415	−0.146	−0.213	0.025
β_{12}	−1.131	−0.787	−2.857	−0.885	−0.387	0.183	0.111	0.303
β_{13}	−1.060	−0.839	−2.713	−0.900	−0.511	−0.191	−0.219	0.031
β_{14}	1.426	0.976	1.623	1.059	0.808	−0.702	−0.708	−0.647
β_{15}	0.498	0.261	0.610	0.273	0.250	−0.415	−0.404	−0.336
β_{16}	1.221	1.028	1.248	1.059	0.816	−0.164	−0.173	−0.077
β_{22}	−1.702	−1.273	−3.504	−1.425	−0.754	0.323	0.256	0.435
β_{23}	−1.686	−1.345	−3.446	−1.462	−0.913	0.037	0.013	0.244
β_{24}	1.815	1.354	1.829	1.525	0.822	0.316	0.335	0.326
β_{25}	1.020	0.716	1.051	0.814	0.420	0.198	0.223	0.243
β_{26}	1.810	1.475	1.751	1.599	0.987	0.395	0.399	0.425
β_{33}	−5.666	−4.127	−7.568	−4.649	−2.619	−0.330	−0.474	−0.472
β_{34}	1.896	1.379	1.908	1.556	0.809	0.187	0.244	0.322
β_{35}	0.565	0.223	0.766	0.347	0.082	−0.149	−0.087	0.047
β_{36}	1.834	1.343	1.808	1.510	0.790	−0.027	0.028	0.117
β_{44}	−0.884	0.102	−0.449	−0.245	0.762	1.312	1.175	1.514
β_{45}	−1.809	−1.111	−1.167	−1.399	−0.260	0.373	0.292	0.557
β_{46}	−1.253	−1.065	−0.452	−1.143	−0.560	−1.126	−1.168	−1.011
β_{55}	0.263	0.425	−0.039	0.296	0.338	0.766	0.758	0.754
β_{56}	−2.021	−1.629	−1.235	−1.804	−0.863	−0.542	−0.580	−0.392
β_{66}	−1.379	−0.699	−0.819	−0.913	−0.177	0.057	−0.017	0.234
γ_i	−0.435	−0.303	−0.601	−0.299	−1.712	−0.146	−0.147	−0.129
μ_1[b]	22.331	17.236	26.816	19.035	9.798	1.297	0.903	2.117
μ_2	6.223	4.760	7.504	5.263	2.783	0.551	0.549	0.744

Notes: [a] For evaluating the biases, each of the price indexes are transformed so as to use Equation 3 as the AI model representation. This table assumes $\gamma_i = -0.05$ for food purchased for off-premise consumption, purchased meals and beverages, and food purchased by employees; and $\gamma_i = 0.05$ for tobacco products, alcoholic beverages purchased for off-premise consumption, and other alcoholic beverages. For descriptions of column content, see Table 3, footnote a.
[b] See Table 3, footnote b.

followed by Models 5, 4, and 3, in that order. Using the food expenditure data, Model 3 is best in every case followed by Model 5 and then either Models 4 and 2 or Models 2 and 4 in order. Therefore, modelling the aggregation error, h_t^k, with a well-established conceptual basis, but with coefficients that are restricted by the representative consumer model fail to systematically reduce aggregation biases. The next section considers biases from introducing instruments for the aggregation error as in Model 6 and 7, which avoids cross-coefficient restrictions.

VI. Simulation: Refinements without Coefficient Restrictions

Table 1 shows that Model 6 is apparently a more flexible approximation than either the lognormal or second-order Taylor approximations. Even though Models 3 and 5 represent the exact index of income at a particular point in time more accurately than Model 6, Model 6 captures variation over time more effectively than Models 2–5. The simulation reported in Table 3 also shows that Model 6 dramatically reduces aggregation bias compared to Models 2–5. This result holds using data on both the distribution of income and the distribution of food expenditures.

Interestingly, Model 6 adds the same general measure of the income distribution, the variance normalized by mean income, as in Models 3 and 5 but in a more flexible framework. Thus, the second moment of the income distribution without the rigidity of cross coefficient restrictions is able to act as an instrument for higher-order terms as well and thus reduce aggregation bias in the representative consumer model.

From Model 7, aggregation bias can also be reduced by using information other than moments of the income distribution. Table 1 shows that variation in the aggregation error with Model 7 based on a third-order (first-order) polynomial yields one-third (two-thirds) of the variation of the aggregation error compared to Model 6 (compare *SD* for the two models). Comparing with other models Table 1 reveals that the use of the representative consumer model with a linear time trend reduces variation of the aggregation bias to half that of the simple representative consumer model. Overall, these results provide strong evidence that flexible intertemporal representations of aggregation bias are able to capture aggregation error more effectively than rigid methods based on conceptually-accurate, time-specific approximations.

Next consider the performance of trend polynomials in reducing average absolute bias. Examining the results in Table 1 suggests that Model 7 is superior to Model 6. However, for reducing biases, Model 7 has the shortcoming when applied to typical data that polynomial trends are closely related to trends in price indexes. Table 3 reveals the reduced benefits of using overly flexible polynomial trends, which may end up tracking the included income distribution and price variables and thus reduce benefits from tracking the aggregation error. In Table 3, Model 6 is more effective in reducing average absolute aggregation bias than the high-order polynomials in Model 7. That is, the effectiveness of Model 7 decreases with increased flexibility (inclusion of higher-order polynomial terms).

For the demand system under weak separability with respect to food consumption, the trend polynomials of Model 7 reduce aggregation bias better than Model 6 because disaggregate food prices exhibit less trend. This property may be particularly useful for estimating demand systems for food involving disaggregate food prices. Specifically, Table 4 shows that Model 7 reduces aggregation bias considerably more than Model 6 based on the food expenditure data whether a first-, second-, or third-order polynomial is used.

In every case, both Models 6 and 7 appear to perform far better than Models 2–5, which have a clearer conceptual basis in the abstract case where variation over time is ignored. Overall, the results suggest that the coefficient-of-variation is a slightly better instrument than a polynomial time trend for the more aggregated system, while polynomial trends work better in more disaggregated food demand systems where price indexes exhibit less trend. The performance the trend models further suggests that empirical studies that include deterministic trends may have inadvertently corrected for some aggregation bias.[11]

VII. Conclusions

Surprisingly, the results of this paper show that the representative consumer model tends to capture aggregation error better than many rigid alternatives that are conceptually motivated. Examples are higher-order Taylor approximations and quantile representations of the income distribution. This may explain some of the success of the representative consumer model in spite of its well-known deficiencies. While models with cross-coefficient restrictions in terms representing aggregation error refine approximations of the function at a given point in time, these approximations are unable to track aggregation error across time better than the representative consumer model.

Alternatively, modelling the aggregation error using an intercept correction term with unrestricted coefficients is shown to be more effective in reducing biases. For example, adding an *ad hoc* polynomial trend component captures aggregation error substantially better than methods that include conceptually motivated second-order or quantile variation in the income distribution. However, too much flexibility in tracking the aggregation error fails to reduce the bias compared to a simpler time trend. Also, due to typical trends in economic time series, an instrument representing income variation normalized by mean income (the coefficient of variation) provides a better approach for eliminating aggregation bias than including only a variance term.

Which alternative works best depends on the smoothness of the income data and strength of price trends over time. If demand for food is estimated assuming separability with respect to food expenditures, then the distribution of total food

[11] Again, the simulation results reported here are based on specific assumptions of elasticities but other simulation results too voluminous to report here verify considerable robustness. When increasing or decreasing γ_k from –0.03 to -0.07 for the income inelastic goods and from 0.03 to 0.07 for the income elastic goods, the ranking of models by aggregation bias was almost unaffected. For personal income data, {Model 6 or Model 7, Order 1} where best followed by {Model 7, Order 2, or Model 2}; Model 7, Order 2; {Model 7, Order 3, or Model 5}; Model 4; and Model 3, in that order (brackets reflect the only cases where orders were reversed depending on elasticities). For food expenditure data, Model 7 was best in every case although the dominant order was either 2 or 3, followed by Model 6, Model 3, Model 5, and {Model 4 or Model 2}.

expenditures is smoother over time than when using personal disposable income. Also, food prices tend to have weaker trends over time than prices in other sectors. As a result, polynomial components are more effective in capturing aggregation bias for estimating food demands conditioned on aggregate food expenditures, whereas an instrument based on the coefficient of variation of income is more effective for studies of other demands conditioned on personal disposable income. (Notably, however, data for the latter instrument may be difficult to find.) In any case, the results obtained here suggest some important principles that now become hypotheses for further research with other data. The most important principles suggested here are that modelling aggregation error is useful for reducing bias and that simplicity works best.

References

Allen, G. P. and Fildes, R. (2001) Principles of forecasting: a handbook for researchers and practitioners, in *Econometric Forecasting* (Ed.) J. Scott Armstrong, Kluwer Academic Publishers, Norwell, MA, pp. 271–327.

Alston, J. M., Chalfant, J. A. and Piggott, N. E. (2002) Estimating and testing the compensated double-log demand model, *Applied Economics*, **34**, 1177–86.

Anderson, G. J. and Blundell, R. W. (1983) Testing restrictions in a flexible dynamic demand system: an application to consumers' expenditure in Canada, *Review of Economics Studies*, **50**, 397–410.

Arndt, C., Liu, S. and Preckel, P. V. (1999) On dual approaches to demand systems estimation in the presence of binding quantity constraints, *Applied Economics*, **31**, 999–1008.

Clements, M. P. and Hendry, D. F. (1998) *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

Deaton, A. S. and Muellbauer, J. (1980) An almost ideal demand system, *American Economic Review*, **70**, 312–26.

De Boer, P. M. C., Martinez, C. and Harkema, R. (2000) Trade liberalization and the allocation over domestic and foreign supplies: a case study for Spanish manufacturing, *Applied Economics*, **32**, 789–99.

De Mello, M., Pack, A. and Sinclair, M. T. (2002) A system of equations model of UK tourism demand in neighbouring countries, *Applied Economics*, **34**, 509–21.

DeJuan, J. P. and Seater, J. J. (1999) The permanent income hypothesis: evidence from the consumer expenditure survey, *Journal of Monetary Economics*, **43**, 351–76.

Duffy, M. (2001) Advertising in consumer allocation models: choice of functional form, *Applied Economics*, **33**, 437–56.

Gorman, W. M. (1961) On the class of preferences fields, *Metroeconomica*, **13**, 53–6.

Gorman, W. M. (1981) Some Engel curves, in *Essays in Theory and Measurement of Consumer Behavior* (Ed.) A. Deaton, Cambridge University Press, Cambridge, pp. 7–29.

Greene, W. (1993) *Econometric Analysis*, Macmillan Publishing Company, New York.

Howe, H., Pollak, R. A. and Wales, T. J. (1979) Theory and time series estimation of the quadratic expenditure function, *Econometrica*, **47**, 1231–47.

Jorgenson, D. W. (1997) *Welfare Volume 1: Aggregate Consumer Behavior*, MIT Press, Cambridge MA.

Karagiannis, G. and Velentzas, K. (2004) Decomposition analysis of consumers' demand changes: an application to Greek consumption data, *Applied Economics*, **36**, 467–504.

Lambert, P. J. (1993) *The Distribution and Redistribution of Income: A Mathematical Analysis*, Manchester University Press, Manchester.

Lewbel, A. (1990) Full rank demand systems, *International Economic Review*, **31**, 289–300

Muellbauer, J. (1975) Aggregation, income distribution, and consumer demand, *Review of Economic Studies*, **62**, 525–43.

Muellbauer, J. (1976) Community preference and the representative consumer, *Econometrica*, **44**, 979–99.

Pashardes, P. (1993) Bias in estimates of the almost ideal demand system with the Stone index approximation, *Economic Journal*, **103**, 908–15.

Pope, R. D. and Just, R. E. (1996) Empirical implementation of *exante* cost functions, *Journal of Econometrics*, **72**, 231–99.

Silberberg, E. (1972) Duality and the many consumers' surpluses, *American Economic Review*, **62**, 942–56.

van Daal, J. and Merkies, A. H. G. M. (1989) A note on the quadratic expenditure model, *Econometrica*, **57**, 1939–43.

Appendix: Proof of Proposition 1

Inheritance of symmetry

Assume the indirect utility function and Marshallian demand for the kth good of each consumer i are given by $U_t(\mathbf{p}_t, M_{it})$ and $f^k(\mathbf{p}_t, M_{it})$, respectively, and satisfy

$$U_t(\mathbf{p}_t, M_{it}) = U_t(\mathbf{p}_t, \underline{M}_t) + Q_{it} \tag{A1}$$

where $U_t(\mathbf{p}_t, M_t)$ is defined as a zero-order approximation of $U_t(\mathbf{p}_t, M_{it})$ around mean income and Q_{it} represents higher-order terms of the Taylor expansion. From Roy's identity,

$$\frac{\partial U_t(\mathbf{p}_t, M_{it})}{\partial p_{kt}} = -\lambda(\mathbf{p}_t, M_{it}) f^k(\mathbf{p}_t, M_{it}) \tag{A2}$$

where

$$\lambda(\mathbf{p}_t, M_{it}) = \frac{\partial U_t(\mathbf{p}_t, M_{it})}{\partial M_{it}}$$

Evaluating Equation A2 at $\underline{M}_t$ and substituting into Equation A1 after differentiation with respect to p_{kt} obtains

$$\frac{\partial U_t(\mathbf{p}_t, \underline{M}_t)}{\partial p_{kt}} + \frac{\partial Q_{it}}{\partial p_{kt}} = -\lambda(\mathbf{p}_t, \underline{M}_t) f^k(\mathbf{p}_t, \underline{M}_t) + \frac{\partial Q_{it}}{\partial p_{kt}} \tag{A3}$$

and

$$\lambda(\mathbf{p}_t, M_{it}) f^k(\mathbf{p}_t, M_{it}) = \lambda(\mathbf{p}_t, \underline{M}_t) f^k(\mathbf{p}_t, \underline{M}_t) - \frac{\partial Q_{it}}{\partial p_{kt}}$$

Aggregating over consumers yields

$$\frac{\sum_i \lambda(\mathbf{p}_t, M_{it}) f^k(\mathbf{p}_t, M_{it})}{n} = \lambda(\mathbf{p}_t, \underline{M}_t) f^k(\mathbf{p}_t, \underline{M}_t) - \left(\sum_i \frac{\partial Q_{it}}{\partial p_{kt}}\right) \Big/ n \tag{A4}$$

Now, defining $\overline{U}_t(\mathbf{p}_t, \underline{M}_t)$ as the level curve of $U_t(\mathbf{p}_t, M_t)$, it follows that

$$\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} = \frac{\partial f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} + \left[\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial U_t}\right] \times \left[\frac{\partial U_t(\mathbf{p}_t, \underline{M}_t)}{\partial p_j}\right]$$

can be represented from Equations A3 and A4 as

$$\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} = \frac{\partial f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} + \left[f^j(\mathbf{p}_t, \partial U_t(\mathbf{p}_t, \underline{M}_t))\right] \times \left[\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial \underline{M}_t}\right]$$

because, from Roy's identity,

$$f^j(\mathbf{p}_t, \partial U_t(\mathbf{p}_t, \underline{M}_t)) = \frac{\partial U_t(\mathbf{p}_t, \underline{M}_t)}{\partial \underline{M}_t} = -\frac{\partial U_t(\mathbf{p}_t, \underline{M}_t)}{\partial p_j}$$

and, from the chain rule,

$$\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial \underline{M}_t} = \left[\frac{\partial f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))}{\partial U_t}\right]\left[\frac{U_t(\mathbf{p}_t, \underline{M}_t)}{\partial \underline{M}_t}\right]$$

To conclude the proof requires showing that

$$\frac{\partial f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} = \frac{\partial f^j(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_k}$$

To focus on the representative consumer model, consider a Taylor expansion of the expenditure function of consumer i, $e(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}))$, around mean income, M_t,

$$e(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it})) = e(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t)) + \Gamma(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t), M_{it} - \underline{M}_t) \tag{A5}$$

where $e(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))$ is the zero-order term of the Taylor expansion of $e(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}))$ around mean income, and the residual of the Taylor expansion is $\Gamma(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t), M_{it} - \underline{M}_t)$.

To apply Shephard's lemma, let the level curves of $U_t(\mathbf{p}_t, M_{it})$ and $U_t(\mathbf{p}_t, \underline{M}_t)$ be defined as $\overline{U}_t(\mathbf{p}_t, M_{it})$ and $\overline{U}_t(\mathbf{p}_t, \underline{M}_t)$, respectively. From Equation A5 and Shephard's Lemma,

$$\frac{\partial e_i(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}))}{\partial p_{kt}} = f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, M_{it})) + f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t)) + R_k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t), M_{it} - M_t)$$

where

$$f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t)) + \frac{\partial e(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_{kt}}$$

and

$$R_k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t), M_{it} - \underline{M}_t) = \frac{\partial \Gamma(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t), M_{it} - \underline{M}_t)}{\partial p_{kt}}$$

In terms of the aggregate mean demand,

$$\frac{\sum_i \left[\partial e_i(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}))/\partial p_{jt}\right]}{n} = \frac{\sum_i \left[f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, M_{it}))\right]}{n} = f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_{it})) + \frac{\sum_i \left[R_j(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t), M_{it} - \underline{M}_t)\right]}{n} \tag{A6}$$

To show that symmetry is inherited by the representative consumer model, note that

$$\frac{\partial^2 e(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_{kt} \partial p_{jt}} = \frac{\partial^2 e(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_{jt} \partial p_{kt}}$$

and, thus, in Equation A6,

$$\frac{\partial f^k(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_j} = \frac{\partial f^j(\mathbf{p}_t, \bar{U}_t(\mathbf{p}_t, \underline{M}_t))}{\partial p_k}$$

Therefore, symmetry holds in the representative consumer model.

Inheritance of summability

By definition, the expenditure function is the sum of the value of each commodity consumed by individual i,

$$e\big(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it})\big) = \sum\nolimits_j p_j f^j(\mathbf{p}_t, M_{it})$$

A zero-order Taylor approximation of $e(\mathbf{p}_t, U_t \times (\mathbf{p}_t, M_{it}))$ around mean income $\underline{M}_t$ is

$$e\big(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_{it})\big) = \sum\nolimits_j p_j f^j(\mathbf{p}_t, \underline{M}_t)$$

which verifies summability in the representative consumer model.

Inheritance of homogeneity

The zero-order Taylor approximation of f^k $(\kappa\mathbf{p}_t, U_t \times (\kappa\mathbf{p}_t, \kappa M_{it}))$ and f^k $(\mathbf{p}_t, U_t$ $(\mathbf{p}_t, M_{it}))$ around mean income are f^k $(\kappa\mathbf{p}_t, U_t(\kappa\mathbf{p}_t, \kappa\underline{M}_t))$ and $f^k(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t))$, respectively. Because Marshallian demands are homogeneous of degree zero,

$$f^k\big(\kappa\mathbf{p}_t, U_t(\kappa\mathbf{p}_t, \kappa M_{it})\big) = f^k\big(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it})\big)$$

implies

$$f^k\big(\kappa\mathbf{p}_t, U_t(\kappa\mathbf{p}_t, \kappa\underline{M}_t)\big) = f^k\big(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t)\big)$$

Thus, the representative consumer in Equation 3 inherits homogeneity from Equation 2.

The case of heterogeneous tastes

For the case of heterogeneous preferences, an analogous proof can be used to show that the representative consumer model inherits integrability properties. The outline of the proof requires modelling the expenditure function of consumer i as $e(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}, \alpha_{it}), \alpha_{it})$ where preferences for individual i are indexed by α_{it}. A Taylor approximation around mean income, $\underline{M}_t$, and mean preferences, $\underline{\alpha}_t$, obtains

$$\begin{aligned} e\big(\mathbf{p}_t, U_t(\mathbf{p}_t, M_{it}, \alpha_{it}), \alpha_{it}\big) &= e\big(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t, \underline{\alpha}_t), \underline{\alpha}_t\big) \\ &\quad + \Gamma\big(\mathbf{p}_t, M_{it}, \underline{M}_t, \alpha_{it}, \underline{\alpha}_t\big) \end{aligned}$$

Continuing the analysis as above shows that the representative consumer model derived from $e(\mathbf{p}_t, U_t(\mathbf{p}_t, \underline{M}_t, \underline{\alpha}_t), \underline{\alpha}_t)$ inherits integrability.