Accounting for Aggregation Bias in Almost Ideal Demand Systems

Ron C. Mittelhammer, Hongqi Shi, and Thomas I. Wahl

This study revisits the consistent aggregation (over households) property of almost ideal demand system (AIDS) models and presents a method to explicitly account for expenditure aggregation bias when estimating the aggregate AIDS model with time-series data. Ignoring aggregation bias can lead to biased and inconsistent parameter estimates and can cause aggregate demand functions to be inconsistent with the demand functions at the individual household level. Recognizing the generally limited information contained in aggregate time-series data for explicitly modeling aggregation bias, we present a new method of constructing an aggregation bias term that is derived from the proportions of households in different income groups. This information is generally available in developed economies. We use this framework to estimate aggregate meat demand within a complete demand system based on U.S. annual expenditure data.

Key words: aggregation bias, AIDS model, bias correction

Introduction

Empirical demand analyses often are based on aggregate time-series data. An assumption often made in these studies is that the market demand functions are consistent with the demand functions of individual households so that the neoclassical restrictions apply to market demands. However, this assumption is only valid under restrictive conditions presented by previous researchers such as Gorman (1959) and Muellbauer (1975, 1976). According to Gorman, market demands expressed as a function of aggregate income are consistent with household demand functions when all households have identical marginal propensities to consume. Muellbauer (1975) proposed a class of preferences called price independent generalized linearity (PIGL) and demonstrated how individual demand functions based on this type of preference structure lead to exact nonlinear aggregation. Under PIGL preferences, demand functions need not be linear in total (or per capita) income in order to obtain consistent aggregation from household demands to market demand. However, market demand generally will depend on the distribution of income across households. A special case of PIGL preferences is its logarithmic form, called PIGLOG preferences, from which the almost ideal demand system (AIDS) is derived.

One of the most important reasons for the popularity of the AIDS model among empirical demand analysts is its property of consistent aggregation across households. However, for AIDS demand functions to exhibit this property, aggregate income must be equally distributed among households and the income distribution across households

must be stable over time if, as is typical, aggregate demand is specified as a function of aggregate (or per capita) income. This is a very restrictive and unrealistic assumption. Furthermore, ignoring the income distributional effect in the aggregated demand model generally results in biased parameter estimates and the aggregate demand model does not properly represent the underlying houshold demand functions (Muellbauer 1975, 1976; Stoker; Blundell, Pashardes, and Weber). Nevertheless, most empirical demand studies have taken the aggregation property of the AIDS model for granted or simply ignored the aggregation problem entirely.

In the exceptional case where extensive pooled cross-section/time-series data and sufficient research resources are available, as in the study by Blundell, Pashardes, and Weber, the aggregation problem can be circumvented, at least in principle, by explicitly aggregating household demand functions. Alternatively, household data can be used to calculate variables (called "aggregation factors" by Blundell, Pashardes, and Weber) that are subsequently added to the aggregate demand model specification to represent the effect of aggregation across households. In this article, we first identify an explicit aggregation bias term whose omission is typical in empirical work and leads to biased and inconsistent estimates of model parameters. We then introduce a new procedure for estimating the expenditure aggregation bias effect that is based on accessible and easily processed information about the income distribution of households. The income distribution itself is estimated using a procedure recently introduced by Majumder and Chakravarty. The goal is to obtain parameter estimates of the aggregate market demand functions that are unbiased, consistent, and represent valid aggregations of household demand functions. The procedure is applied to an AIDS model of the aggregate U.S. domestic demand for meats and conclusions are drawn.

Aggregation Theory Background

The problem of aggregation in demand analysis has received considerable attention in the economics literature. The early works in this area are concerned primarily with consistent linear aggregation, including Samuelson, Theil (1954), Gorman (1953, 1959), and Green (1964). The necessary and sufficient condition for household demand (consumption) functions to consistently and linearly aggregate to a market demand (consumption) function is that the marginal propensity to consume (MPC) is identical across households, which leads to market demand functions that are independent of the income distribution. Muellbauer (1975, 1976) established more general conditions based on PIGL preferences for market demand functions to be consistent with household demand functions. A major advantage of PIGL preferences is that they allow nonlinear forms of demand (and Engel) functions and yet still allow for aggregate demand at the market level to be consistent with household demands.

Following Deaton and Muellbauer, a household-specific AIDS model derived from the logarithmic form of PIGL preferences can be expressed in share form as

$$(1) \qquad w_{ih} = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \log\left(\frac{x_h}{k_h P}\right), \qquad \forall i, h$$

$$\log(P) = \alpha_0 + \sum_i \alpha_i \log(p_i) + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \log(p_i) \log(p_j),$$

where

$$\sum_i \alpha_i = 1, \qquad \sum_i \gamma_{ij} = 0, \qquad \sum_j \gamma_{ij} = 0, \qquad \sum_i \beta_i = 0 \qquad \gamma_{ij} = \gamma_{ji} \quad \forall i \neq j,$$

and $k_h > 0$, $\forall h$, are taste difference parameters allowing for different preference relations across households. The share of aggregate expenditure allocated to good i can be defined as

$$(2) \qquad \bar{w}_i = \frac{\sum_h p_i q_{ih}}{\sum_h x_h} = \frac{\sum_h x_h w_{ih}}{\sum_h x_h},$$

where q_{ih} is the quantity of commodity i consumed by household h, p_i is the price of commodity i, x_h is the total expenditure of household h, and w_{ih} is the share of total expenditure allocated to commodity i by household h.

The aggregate expenditure share equation in the AIDS model can be obtained by substituting equation (1) into (2), obtaining

$$(3) \qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \left(\frac{\sum_h x_h \log\left(\frac{x_h}{k_h P}\right)}{\sum_h x_h} \right).$$

Letting $r_h = x_h / \sum_h x_h$ represent the hth household's share of aggregate expenditure, equation (3) can be reexpressed as

$$(4) \qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \left[\sum_h r_h \log\left(\frac{x_h}{k_h P}\right) \right],$$

where $r_h \in [0, 1]$ and $\sum_h r_h = 1$. Letting x^* and k^* denote the respective weighted (by r_hs) geometric means of expenditures and taste difference parameters, equation (4) can be rewritten as

$$(5) \qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \log\left(\frac{x^*}{k^* P}\right),$$

where

$$x^* = \prod_h x_h^{r_h} \qquad \text{and} \qquad k^* = \prod_h k_h^{r_h}.$$

Defining N to be the number of households and $\bar{x}$ to be the simple arithmetic mean of household expenditures, it follows that $x_h = r_h(\sum_h x_h) = r_h(N\bar{x})$, and the weighted geometric mean of expenditure can be rewritten as

$$(6) \qquad x^* = \prod_h (x_h)^{r_h} = \left(\prod_h r_h^{r_h} \right)(N\bar{x}) = \left(\frac{N}{Z} \right)\bar{x}, \qquad \text{where} \qquad Z = \left(\prod_h r_h^{r_h} \right)^{-1}.$$

Note that $\log Z = -\sum_h r_h \log(r_h)$ is the entropy measure of the distribution (dispersion) of household's expenditure shares. It can be shown that Z achieves its maximum value of N when the households' expenditure shares are identical, namely, $r_h = 1/N \ \forall \ h$ (Theil 1971). In the general case where the households' aggregate expenditure shares are not identical, $N / Z > 1$, which implies that x^* is larger than $\bar{x}$. This indicates that under

PIGLOG preferences, the simple mean of households' expenditures always underestimates the true value of the aggregate representative expenditure, x^*. Therefore, the parameter estimates (β_is) associated with real expenditure are generally biased and inconsistent if the simple mean is used in place of the geometric mean of households' expenditures.

Regarding the geometric mean of taste change parameters, k^*, in the aggregate AIDS model (5), first note that if all households have the same tastes, so that $k_h = 1 \; \forall h$, then $k^* = 1$ and the taste variable vanishes. Alternatively, if tastes differ across households, but tastes and the distribution of income shares across households remain stable over time, then k^* is a constant that can be subsumed into the intercept term of (5), as $\alpha_i^* = \alpha_i - \beta_i \log(k^*)$. Finally, if tastes and/or income shares change over time, k^* can change over time, leading to the intercept α_i^* changing over time as well. Thus, even if the tastes of individual households do not change over time, the fact that tastes are different across households can induce a taste effect on aggregate demand through a changing income distribution. In modeling aggregate demand, it will then be necessary to account for a taste effect even if individual household preference relations do not change.

Modeling the Aggregation Bias in AIDS

From the preceding discussion, the true geometric mean of household expenditure in the AIDS model can be expressed as $x^* = (N/Z)\bar{x}$. Substituting this expression into (5) and subsuming any taste effect into the intercept term yields the following aggregate AIDS model:

$$(7) \qquad \bar{w}_i = \alpha_i^* + \sum_j \gamma_{ij} \log(p_j) + \beta_i \log\left(\frac{\bar{x}}{P}\right) + \beta_i[\log(N) - \log(Z)], \qquad \forall i.$$

We refer to the entire term in brackets as the expenditure aggregation bias term, which represents an omitted variable when using simple mean expenditure in place of the weighted geometric mean of households' expenditures. From (7) it is evident that, to calculate the expenditure aggregation bias term, time-series information on the number of households and on individual households' shares of aggregate expenditure are needed.

Information on the shares of aggregate expenditure across households is generally unavailable or inaccessible. However, time-series information on the number of households in different income categories is readily available for most developed economies and can provide valuable information for closely approximating the income distribution and aggregation bias term in the aggregate AIDS model. We now discuss an approximation to the expenditure aggregation bias term that can be used to correct or reduce aggregation bias.

Assume initially that all households have the same PIGLOG preference structure. Let $\phi(x)$ denote the density function of household income, so that the expected value of income across all households can be expressed as

$$(8) \qquad \bar{x} = E(x) = \int_{x_L}^{x_U} x\phi(x) \, dx,$$

where x_L and x_U are, respectively, the lower and upper bounds of the income distribution,

and $E(\cdot)$ denotes an expectation taken with respect to the density $\phi(x)$. Express the AIDS model of a household in quantity-dependent form as follows:

$$(9) \qquad q_i = \frac{x}{p_i}\left[\alpha_i + \sum_j \gamma_{ij}\log(p_j) + \beta_i\log\left(\frac{x}{P}\right)\right],$$

where q_i is quantity demanded for commodity i. The expected value of quantity demanded of commodity i across all households (i.e., market demand on a per household basis) can be obtained by taking the expectation of (9) with respect to the income distribution as

$$(10) \qquad \bar{q}_i = \int_{x_L}^{x_U} q_i\phi(x)\, dx = \int_{x_L}^{x_U}\left\{\frac{x}{p_i}\left[\alpha_i + \sum_j \gamma_{ij}\log(p_j) + \beta_i\log\left(\frac{x}{P}\right)\right]\right\}\phi(x)\, dx$$

$$= \frac{1}{p_i}\left[\alpha_i + \sum_j \gamma_{ij}\log(p_j)\right]\int_{x_L}^{x_U} x\,\phi(x)\, dx + \frac{\beta_i}{p_i}\int_{x_L}^{x_U} x\log\left(\frac{x}{P}\right)\phi(x)\, dx.$$

Based on (10), the average budget share of good i can be expressed as

$$(11) \qquad \bar{w}_i = \frac{p_i\bar{q}_i}{\bar{x}} = \frac{1}{\bar{x}}\left[\alpha_i + \sum_j \gamma_{ij}\log(p_j)\right]\int_{x_L}^{x_U} x\phi(x)\, dx + \frac{\beta_i}{\bar{x}}\int_{x_L}^{x_U} x\log\left(\frac{x}{P}\right)\phi(x)\, dx$$

$$= \alpha_i + \sum_j \gamma_{ij}\log(p_j) + \beta_i\log\left(\frac{\bar{x}}{P}\right) + \beta_i\left[\frac{E(x\log(x))}{\bar{x}} - \log(\bar{x})\right],$$

where $E(x\log(x))/\bar{x}$ is the analog to the logarithm of the geometric mean of income, and the bracketed term in (11) is the analog to the expenditure aggregation bias term of (7).[1] Consistent with our previous observation, note that the aggregation bias term in (11) vanishes when households' aggregate expenditure shares are identical, which in the current context is represented by a degenerate income distribution defined as $\phi(x) = 1$ when $x = \bar{x}$, $\phi(x) = 0$ otherwise. The expenditure aggregation bias term of the AIDS model in (11) is directly interpretable as the standard measure of household income inequality (Theil 1971, p. 653).

In order to estimate the aggregate budget share equations as defined in (11), information on households' income distribution for calculating $E(x\log(x))$ is required in addition to information on $\bar{x} = E(x)$, prices, and aggregate budget shares. A method of estimating households' income distribution using time-series information on the number of households in different income categories is presented below.

Estimation of Households' Income Distribution

A number of methods have been presented in the literature for estimating the income distribution of households based on census data relating to the proportions of households

---

[1] To maintain consistency with our empirical procedure, we tacitly assume here that the number of households is large enough to assume the income distribution is continuous, and thus Riemann integrals are used. In this context, the bracketed term in (11) is the continuous analog to the bracketed term in (7). The entire derivation in this section could be repeated using a discrete income distribution $\Phi(x)$, say, and using Stieltjes integrals, in which case the bracketed terms in (7) and (11) would be identical. For example, using Steiltjes integrals it would then follow that (8) could be rewritten as

$$\bar{x} = N^{-1}\sum_{h=1}^{N} x_h \equiv \int_{x_L}^{x_U} x\,d\Phi(x),$$

and so on.

in various income categories (e.g., Champernowne; Salem and Mount; Singh and Maddala; Dagum; McDonald; Esteban). Recently, Majumder and Chakravarty proposed a method for estimating a four-parameter income distribution based on Esteban's "income share elasticity" approach. The new method subsumes the three- and four-parameter distributions of Singh and Maddala, Dagum, and McDonald as special cases.[2] Based on empirical evidence, as well as theoretical considerations relating to the satisfaction of the Weak Pareto Law, Majumder and Chakravarty document their approach as being significantly better than its predecessors for estimating households' income distribution. Their method is adapted to obtain the household income distribution information necessary for calculating the AIDS aggregation bias term. The density function used to represent the households' income distribution is given by

$$(12) \qquad f(x; a, b, c, d) = \frac{b d^{a/b} c^{\,(b/d)-a} x^{(b/d)-a-1}}{B\left(\left(\frac{1}{d}\right) - \left(\frac{a}{b}\right), \frac{a}{b}\right)} ((cx)^b + d)^{-1/d},$$

where $x \geq 0$, $b > ad$, and $B(m,n)$ is the beta function.

Given k income classes defined by the income partition $0, x_1, x_2, \ldots, x_k$, and sample observations on household membership in the various income classes, the distribution is fit by calculating the values of the parameters a, b, c, d according to the minimum χ^2 method as

$$(13) \qquad \hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \sum_{i=1}^{k} \frac{[n_i - \lambda_i(\Theta)]^2}{[n\lambda_i(\Theta)]},$$

where

$$\Theta = (a, b, c, d),$$

$$n_i = \text{number of households in income class } i,$$

$$n = \sum_{i=1}^{k} n_i \text{ is the sample size,}$$

and

$$(14) \qquad \lambda_i(\Theta) = \int_{x_{i-1}}^{x_i} f(x; \Theta) \, dx$$

is the proportion of households assigned to income class i when the value of the parameter vector equals Θ.[3] The estimated density can then be used to estimate the expectation terms needed to specify the aggregate budget share equation (11), as

---

[2] The Majumder and Chakravarty approach does not subsume the two-parameter Pareto, log normal, and gamma distributions, but the empirical performance of these distributions has been shown to be poor when the entire income range of households is considered.

[3] Note that $f(x;a,b,c,d)$ is a continuous density having the nonnegative real line for its support, and so the upper bound of the highest income category is set to $x_k = \infty$ while the lower bound of the lowest income category is set to $x_o = 0$. A discussion of the advantages of using the minimum chi-squared method relative to other estimation methods when using grouped data can be found in McDonald and Ransom.

$$(15) \qquad\qquad \hat{\bar{x}} = \hat{E}(x) = \int_0^\infty xf(x; \hat{\theta}) \, dx,$$

and

$$(16) \qquad\qquad \hat{E}(x \log(x)) = \int_0^\infty (x \log(x)) f(x; \hat{\theta}) \, dx,$$

where $\wedge$ denotes an estimate based on the estimated income distribution. A time series of expectation terms can be generated by applying the preceding estimation procedure repeatedly to annual (or other periodic) observations on the proportions of households contained in various income classes.

Annual household income data for the United States as reported in *Current Population Reports: Consumer Income* by the U.S. Bureau of Census publications for the years 1963 through 1989 were used to estimate yearly income distributions based on the aforementioned procedure of Majumder and Chakravarty. Calculations were performed using the OPTMUM and INTQUAD procedures in the GAUSS programming language. The estimated parameter values for the income distributions, as well as goodness-of-fit measures, are presented in table 1. Graphs of the income distributions generated by the estimation procedure are provided in figures 1 and 2 for the years 1989 and 1963, respectively. (The uppermost income categories are not graphed for *actual* observations since the upper bound of the category is unknown.) Consistent with the findings of Majumder and Chakravarty, the estimated income distributions fit the household income distribution data very well.[4] Based on the estimated parameters of income distributions for the years 1963–89, the values of $\hat{E}(x\log(x))/\hat{\bar{x}}$, $\log(\hat{\bar{x}})$, and the expenditure aggregation bias term (the difference between the two) for the years 1963–89 were calculated and used in an analysis of aggregate U.S. meat demand within a complete demand system, as described in the next section.

To this point we have suppressed the taste parameter k_h [recall equation (1)]. Reintroducing k_h values in the derivation of (11) can be viewed as altering the intercept of the share equations from α_i to $\alpha_i^* = \alpha_i - \beta_i\log(k^*)$, where k^* is the aggregate income share-weighted geometric mean of the k_hs. In the event that aggregate demand is dependent on a taste effect due to either changing households' tastes or a changing income distribution, or both, the intercept would need to be modeled as a function of time, indicator variables, and/or sociodemographic characteristics.

Aggregate U.S. Domestic Demand for Meats

The above theoretical framework for modeling the expenditure aggregation bias term is used to estimate a complete AIDS demand system for beef, pork, poultry, nonmeat foods, and nonfood commodity groups. In the model, total expenditure is equal to aggregate income. The data consist of annual per capita consumption and price indices from 1963 to 1989, as defined by Eales and Unnevehr (1993). Nonmeat food quantity is food quantity [the ratio

---

[4] The estimated values of some of the parameters, especially a and c, were not stable for certain years and the apparent problem was not mitigated by examining alternative starting values for the parameters. Nonetheless, the graphs and probabilities assigned to income intervals, as well as the expenditure aggregation bias term, changed only gradually over time.

Table 1. Income Distribution Parameter Estimates and Goodness of Fit

Year	a	b	c	d	r	MSE ($\times 10^5$)	MAPE (%)
1989	9.3001112	1.3426029	0.0015382	0.1245056	0.99	1.462	6.33
1988	5.8542338	1.7102324	0.0054715	0.2372958	0.99	1.422	7.00
1987	147.74599	1.0983423	5.4126951	0.0073651	0.92	1.847	7.33
1986	19.965	1.2253179	0.0003772	0.0574771	0.92	1.736	6.90
1985	147.74599	1.0505792	4.064289	0.0070414	0.95	1.446	6.23
1984	111.00374	1.0287858	6.1556595	0.0091457	0.97	1.333	5.80
1983	11.432500	1.3580654	0.0016607	0.1061946	0.97	1.617	6.80
1982	8.6642657	1.4189186	0.0030473	0.1411644	0.97	2.075	7.45
1981	15.803459	1.2682110	0.0008908	0.0735502	0.98	2.306	8.27
1980	17.380112	1.1835337	0.0006270	0.0625766	0.97	3.666	9.94
1979	6.1034639	1.7187399	0.0096094	0.2301832	0.98	4.093	10.33
1978	91.738635	1.0411857	1.6074304	0.0111572	0.99	2.612	8.48
1977	51.784068	1.0264895	4.8119187	0.0192196	0.98	3.177	8.86
1976	9.1243425	1.4898642	0.0053865	0.1411470	0.97	3.698	9.92
1975	7.0741371	1.7081771	0.0112586	0.2018450	0.95	3.645	10.61
1974	4.9038478	2.3399253	0.0282748	0.3804337	0.99	3.440	12.97
1973	4.0196473	2.8846116	0.0429040	0.5537529	0.99	2.954	11.20
1972	3.7238222	3.3285241	0.0529536	0.6850654	0.99	2.174	9.87
1971	3.5658722	3.9907260	0.0644445	0.8538593	0.99	2.362	9.59
1970	3.4529287	4.5935116	0.0714653	1.0118616	0.99	2.230	8.38
1969	3.4334308	5.2638547	0.0779547	1.1654743	0.99	2.051	8.03
1968	3.4440264	5.4420382	0.0859030	1.1969583	0.99	2.098	8.88
1967	3.1268873	6.2167642	0.0983388	1.4837389	0.98	2.155	7.61
1966	3.3538768	6.0064850	0.1012142	1.3587836	0.97	2.979	7.44
1965	3.1954798	6.2979538	0.1115857	1.4854700	0.98	2.032	5.72
1964	3.4092788	5.7618499	0.1106460	1.2962403	0.98	1.833	5.39
1963	3.3385654	6.3419247	0.1182020	1.4533337	0.98	2.375	6.47

Notes: a, b, c, and d are the estimates for the parameters of the Majumder and Chakravarty income distribution given in equation (13). r, MSE, and MAPE are, respectively, the correlation, mean square error (multiplied by 10^5), and mean absolute percent error of the relationship between predicted ($\hat{p}$) and actual (p) proportions of households in the income categories reported by the U.S. Bureau of the Census in a given year. Income categories used (measured in thousands of dollars) are defined by the following break points:
1988–89: 5, 10, 15, . . . , 100
1979–87: 2.5, 5.0, 7.5, . . . , 40, 45, 50, 60, 75
1975–78: 2, 3, 4, . . . , 18, 20, 25, 50
1967–74: 1.0, 1.5, 2.0, . . . , 4, 5, 6, . . . , 10, 12, 15, 25, 50
1963–66: 1.0, 1,5, 2.0, . . . , 4, 5, 6, . . . , 10, 12, 15, 25

of food expenditures to the food consumer price index (CPI)] minus the sum of beef, pork, and poultry quantities. Nonmeat food price is the ratio of nonmeat food expenditures to nonmeat food quantity. The nonfood CPI is used as the price of nonfood, and nonfood quantity is defined to be nonfood expenditures divided by nonfood CPI.

Models are estimated with and without the expenditure aggregation bias term. In order to analyze potential taste effects on aggregate demand, three intercept-shifting terms were added to these two models which were motivated by a host of previous studies suggesting that taste/structural change had occurred around the mid-1970s (Braschler; Chavas; Dahlgran; Eales and Unnevehr; Moschini and Mielke; and Thurman). The first term is a time-
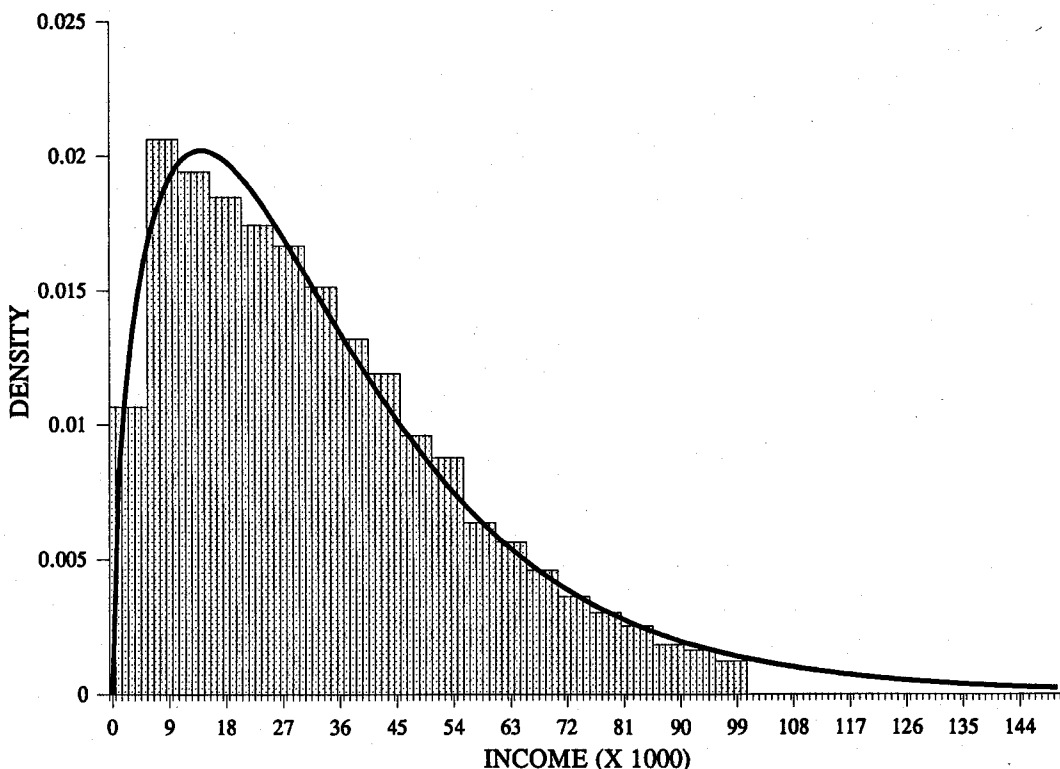
Figure 1. Income distribution 1989: actual vs. predicted

trend variable ($T = 1,2,3, \ldots$) which is included to proxy a secular taste change effect. The second term is an indicator variable ($I_t = 1$ if $T \le t$ and $= 0$ otherwise) for capturing a structural break in households' preferences occurring after time period t. The third term is a trend shifter (T multiplied by I_t) for modeling a possible change in taste-induced consumption trends caused by a structural break in preferences. The two aggregate AIDS models of U.S. demand are given as follows:

$$\text{Model I} \qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \log\left(\frac{\hat{\bar{x}}}{P}\right) + d_{i1} T + d_{i2} I_t + d_{i3} T I_t,$$

and

$$\text{Model II} \qquad \bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log(p_j) + \beta_i \log\left(\frac{\hat{\bar{x}}}{P}\right) + \beta_i \left[\frac{\hat{E}(x \log(x))}{\hat{\bar{x}}} - \log(\hat{\bar{x}})\right]$$

$$+ d_{i1} T + d_{i2} I_t + d_{i3} T I_t.$$

Iterated three-stage least squares (IT3SLS) is used to estimate the two AIDS models. In order to identify a starting point for the structural break terms, the value of t in the definition of the indicator variable I_t in Models I and II was treated as an unknown integer-valued parameter in the range $\{1968, 1969, \ldots, 1985\}$. The instruments used in estimation include the aforementioned indicator and time-trend variables and their inter-action, as well as the natural logarithms, and square of the natural logarithms, of the
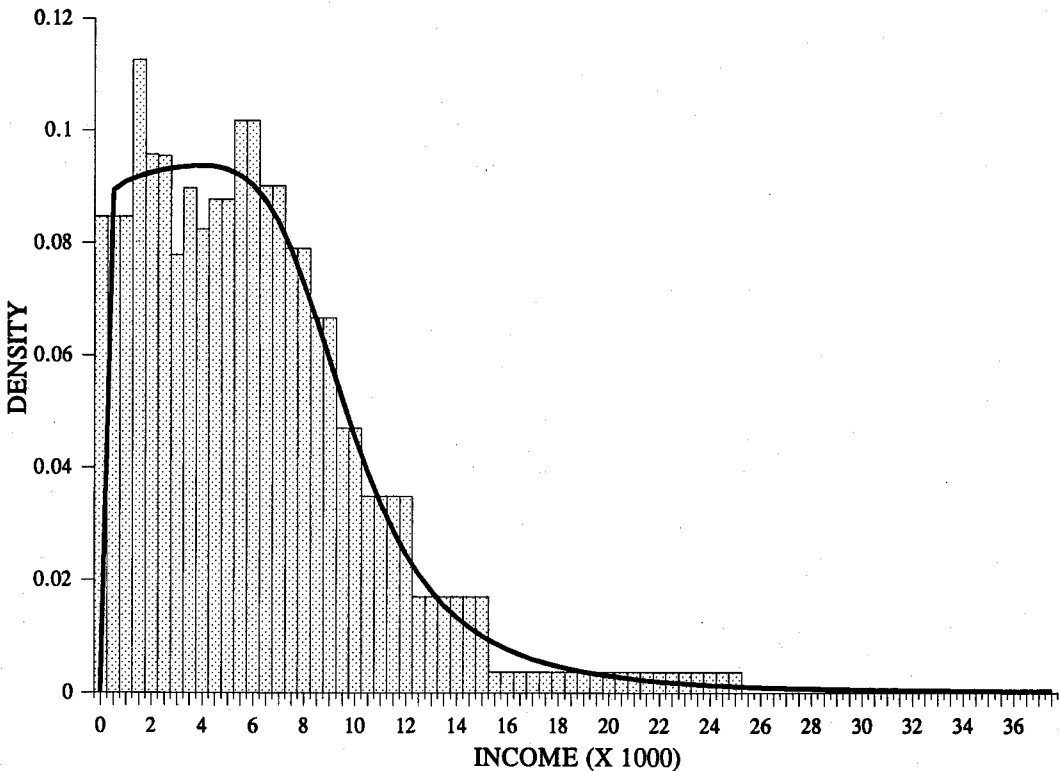
Figure 2.   Income distribution 1963: actual vs. predicted

U.S. population, the treasury bill rate (Economic Report of the President, 1990, pp. 329, 362, and 376), the CPI for fuel and energy, and the wage rate for meat-packing plant workers (U.S. Department of Commerce). The models were estimated in linearized form using the modified Stone's price index suggested by Moschini. Moschini showed that the Stone price index typically used in estimating the linear AIDS model is not invariant to changes in units of measurement. One solution proposed by Moschini is to use scaled prices in Stone's price index, such as scaling prices by their means, so that the Stone price index is invariant to the measurement unit. In this study, all price variables are scaled by their means. The particular form of Stone's price index used is given by $\log(P) = \Sigma_i W_{it-1} \log(P_{it})$. Budget shares are lagged one period to circumvent the problem of endogenous budget shares in the definition of Stone's price index. For both Models I and II, $t = 1974$ was estimated to be the last year preceding the structural break, indicating that Models I and II with I_t set to I_{1975} were the best estimates of the structural equations (in the sense of minimizing the iterated weighted sum of squared residuals inherent in the definition of the IT3SLS estimator). Parameter estimates for the two models based on a structural break occurring in 1975 are presented in table 2.

An important feature of Model II is the attempt to segregate the effects of taste/structural change from expenditure aggregation bias. In empirical demand studies based on time-series data, indicator variables, and/or functions of time are often included to represent changes in consumer preferences or to model structural breaks. A potential complication in analyses that do not incorporate the expenditure aggregation bias cor-

rection term is the degree to which the term is correlated with taste/structural change variables. Estimates of the parameters associated with taste/structural change variables and statistical tests for taste/structural change can be misleading if the expenditure aggregation bias term is entangled with, or proxied by, the taste/structural change variables. In the case at hand, the correlation is 0.68 between actual values and linear least squares predictions of the expenditure aggregation bias term using the taste/structural change variables with a structural break in 1975 as explanatory variables. The correlations are 0.62, 0.95, and 0.80 for the periods 1963–74, 1975–87, and 1975–89, respectively. These interrelationships have a notable effect on the interpretation of the taste/structural change parameters and the significance of the expenditure aggregation bias terms.

Differences in corresponding price parameter values between the models with and without bias correction range from as little as 0.47% to a high of 17.88% with a mean absolute percent difference of 4.77%.[5] The effects of aggregation bias on the expenditure term parameters are more pronounced, ranging in magnitude between 8.25% and 15.97% with a mean absolute percentage difference of 11.99%. The intercepts of the share equations were also affected by the exclusion of the expenditure aggregation bias term.

In both models, joint χ^2-tests (table 2) suggest significant taste/structural change.[6] Furthermore, reestimating the models without the taste/structural change variables yielded several implausible (wrong sign, suspect magnitude) elasticities and substantively reduced explanations of the historical budget shares. However, qualitatively the individual taste/structural change directions are identical between models. In particular, the downward trends in expenditure shares for all food types accelerated for beef, decelerated for pork, and reversed for poultry and nonmeat foods after 1974 (see table 3). The magnitudes of differences in rates of change and intercept shifts range from 1.98% to 317.24% with a mean absolute difference of 45.45%.

Price and expenditure elasticities calculated from the two models are presented in table 4. Since lagged budget shares were used in Stone's price index, Chalfant's formula is used to calculate the price elasticities. The respective signs of all elasticities are identical between the two models. Both models imply inelastic price responses for all commodity categories, with inelastic expenditure elasticities for food items and a slightly elastic expenditure elasticity for nonfood items. Differences in the magnitudes of direct price elasticities range from 0.15% to 15.14% with a mean absolute percentage difference of 5.94%. The differences in expenditure elasticities are larger on average, ranging from 0.98% to 19.19% with a mean absolute percentage difference of 8.23%.

Overall, the two models lead to the same qualitative conclusions regarding demand response. Differences in magnitude between parameters, elasticities, and taste/structural change effects are generally 10% or less with almost all differences being less than 20%. On average, the differences between the two models tended to be larger for expenditure and taste/structural change effects than for price effects. Given these results and the

---

[5] The percentage differences reported here are calculated as [(bias corrected value − uncorrected value)/uncorrected value] × 100.

[6] Note that while the χ^2-statistics are impressively large, the χ^2-tests should be interpreted conservatively here. The maintained hypothesis is that a structural break did occur within a specified range of years, as suggested by past research, and the best estimate of the break point is chosen. In the event that there is no structural change, the estimation procedure will nonetheless choose the best breakpoint from the feasible set, and the approach will have a tendency to overstate the statistical significance of the breakpoint. A statistical test with asymptotically correct size for the type of structural break analyzed in this study has been recently introduced by Andrews. However, because of the limited sample size relative to the number of parameters in the model, Andrews's approach was not pursued.

Table 2. Parameter Estimates for U.S. AIDS Models with and without Agression Bias

Variable	Model I AIDS w/o Bias Correction		Model II AIDS with Bias Correction		|Difference| in Parame- ter Values|
	Parameter (×100)	|t-Value|	Parameter (×100)	|t-Value|	|(%)|
Beef equation:					
Intercept	15.3773	5.263	16.6800	4.602	8.47
$\log(P_{beef})$	1.0905	5.263	1.0854	5.197	0.47
$\log(P_{pork})$	0.2724	1.913	0.2520	1.753	7.49
$\log(P_{poultry})$	0.1939	4.368	0.1980	4.369	2.11
$\log(P_{nonmeat})$	0.3109	0.604	0.3665	0.696	17.88
$\log(P_{nonfood})$	−1.8677	4.145	1.9019	4.046	1.83
$\log(\hat{x}/P)$ or $[\hat{E}(x \log x)/\hat{x}] - \log(P)$	−1.2152	3.953	−1.3154	3.546	8.25
I_{1975}	0.4209	2.786	0.3558	2.263	15.47
Trend	−0.0604	7.531	−0.0579	6.674	4.14
I^*_{1975} trend	0.0313	2.510	0.0261	2.082	16.61
Pork equation:					
Intercept	8.7734	5.235	10.0000	4.868	13.98
$\log(P_{pork})$	0.2030	1.226	0.1851	1.105	8.82
$\log(P_{poultry})$	0.0667	1.366	0.0706	1.414	5.85
$\log(P_{nonmeat})$	0.8227	1.973	0.8924	2.125	8.47
$\log(P_{nonfood})$	−1.3648	5.207	−1.4001	5.184	2.59
$\log(\hat{x}/P)$ or $[\hat{E}(x \log x)/\hat{x}] - \log(P)$	−0.7628	4.268	−0.8668	4.083	13.63
I_{1975}	0.2567	2.636	0.2938	2.879	14.45
Trend	−0.0119	1.882	−0.0101	1.528	15.13
I^*_{1975} trend	−0.0134	1.619	−0.0157	1.851	17.16
Poultry equation:					
Intercept	1.8402	2.967	2.1320	2.834	15.86
$\log(P_{poultry})$	0.3367	5.509	0.3458	5.531	2.70
$\log(P_{nonmeat})$	−0.2262	1.639	−0.2458	1.784	8.66
$\log(P_{nonfood})$	−0.3710	3.642	−0.3659	3.528	1.37
$\log(\hat{x}/P)$ or $[\hat{E}(x \log x)/\hat{x}] - \log(P)$	−0.1641	2.496	−0.1903	2.457	15.97
I_{1975}	0.1943	6.068	0.2070	6.301	6.54
Trend	0.0128	5.838	0.0137	5.963	7.03
I^*_{1975} trend	−0.0174	6.623	−0.0185	7.011	6.32
Nonmeat food:					
Intercept	75.8328	5.166	83.6334	4.539	10.29
$\log(P_{nonmeat})$	7.1979	3.267	7.2499	3.163	0.72
$\log(P_{nonfood})$	−8.1053	3.502	−8.2604	3.383	1.91
$\log(\hat{x}/P)$ or $[\hat{E}(x \log x)/\hat{x}] - \log(P)$	−6.2706	4.070	−6.9050	3.667	10.12
I_{1975}	3.5124	5.055	3.9204	5.354	11.62
Trend	0.0058	0.174	0.0242	0.645	317.24
I^*_{1975} trend	−0.3157	5.926	−0.3478	6.488	10.17

R^2:					
Beef		0.990		0.989	
Pork		0.985		0.984	
Poultry		0.947		0.945	
Nonmeat food		0.953		0.948	

Table 2. Continued

Variable	Model I AIDS w/o Bias Correction	Model II AIDS with Bias Correction
Structural change χ^2-tests:		
I_{1975} (4df)	156.080	159.524
Trend (4df)	130.327	123.151
I^*_{1975} trend (4df)	130.956	141.779
Homogeneity and symmetry χ^2-test:		
Appropriate linear restrictions (10df)	8.105	7.859
Nonnested P-tests (χ^2-test):		
H_0: AIDS w/o bias correction (4df)	2.255	
H_0: AIDS with bias correction (4df)		3.964

Note: Missing parameter values can be obtained by the use of symmetry and adding up conditions. The percentage difference is calculated as [(bias corrected value − uncorrected value)/uncorrected value] × 100.

Table 3. Taste/Structural Change Effects for the U.S. AIDS Models with and without Aggregation Bias Term

Commodity	Difference (%)	$\dot{W}_i$–No Bias Correction (×100)	$\dot{W}_i$–with Bias Correction (×100)
1963–75:			
Beef	4.28	−0.0291	−0.0318
Pork	1.98	−0.0253	−0.0258
Poultry	4.35	−0.0046	−0.0048
Nonmeat food	4.42	−0.3099	−0.3236
1976–89:			
Beef	−4.14	−0.0604	−0.0579
Pork	−15.13	−0.0119	−0.0101
Poultry	7.03	0.0128	0.0137
Nonmeat food	317.24	0.0058	0.0242

Commodity	Difference (%)	ΔIntercept– No Bias Correction (×100)	ΔIntercept– with Bias Correction (×100)
Beef	17.86	−0.0140	−0.0165
Pork	8.73	0.0825	0.0897
Poultry	7.11	0.149	0.1596
Nonmeat food	1.57	−0.5917	−0.6010

Note: $\dot{W}_i$ is the ceteris paribus rate of change in W_i with respect to time; ΔIntercept is calculated as the difference in the level of W_i in 1975 ($T = 13$) between when $I_{1975} = 0$ and $I_{1975} = 1$. The percent difference is calculated as [(bias corrected value − uncorrected value)/uncorrected value] × 100.

Table 4. Marshallian Elasticities and Standard Errors for the U.S. AIDS Model with and without Accounting for Aggregation Bias

	Beef	Pork	Poultry	Nonmeat	Nonfood	Expenditure
Without aggregation bias correction:						
Beef	−0.607	0.086	0.082	0.263	−0.341	0.518
	(0.076)	(0.057)	(0.015)	(0.170)	(0.201)	(0.101)
Pork	0.178	−0.828	0.063	0.599	−0.530	0.517
	(0.119)	(0.151)	(0.037)	(0.342)	(0.228)	(0.123)
Poultry	0.444	0.166	−0.206	−0.549	−0.413	0.557
	(0.083)	(0.098)	(0.129)	(0.296)	(0.267)	(0.130)
Nonmeat	0.045	0.049	−0.017	−0.374	−0.242	0.540
	(0.030)	(0.029)	(0.009)	(0.146)	(0.186)	(0.098)
Nonfood	−0.028	−0.017	−0.005	−0.139	−0.929	1.118
	(0.006)	(0.004)	(0.001)	(0.032)	(0.044)	(0.022)
With aggregation bias correction:						
Beef	−0.610	0.078	0.083	0.294	−0.320	0.475
	(0.077)	(0.058)	(0.016)	(0.174)	(0.214)	(0.119)
Pork	0.162	−0.841	0.066	0.650	−0.493	0.456
	(0.121)	(0.155)	(0.039)	(0.347)	(0.282)	(0.143)
Poultry	0.454	0.172	−0.179	−0.593	−0.355	0.501
	(0.086)	(0.103)	(0.136)	(0.303)	(0.280)	(0.151)
Nonmeat	0.050	0.053	−0.019	−0.361	−0.220	0.497
	(0.031)	(0.030)	(0.010)	(0.151)	(0.199)	(0.116)
Nonfood	−0.029	−0.017	−0.005	−0.143	−0.935	1.129
	(0.006)	(0.004)	(0.001)	(0.033)	(0.047)	(0.026)

Note: Elasticities are measured at sample means. Figures in parentheses are bootstrapped standard errors.

variability in the parameter estimates, one wonders to what extent the models can be differentiated on a statistical basis?

To assess the models' consistency with neoclassical theory, a Wald χ^2-test of the homogeneity and symmetry restrictions was conducted. The neoclassical restrictions could not be rejected at any reasonable level of type I error for either model. Thus, the models could not be differentiated on the basis of adherence to the neoclassical restrictions.

Using a multivariate version of MacKinnon and Davidson's nonnested P-test neither the uncorrected or corrected AIDS models could be rejected at any reasonable level of type I error (see table 2). Pursuing this result further, Model (II) was reestimated without requiring that the $\log[(\hat{x})/P]$ and $[\hat{E}(x \log(x))/\hat{x} - \log(\hat{x})]$ terms share the same β_i parameters and it was found by a Wald-test that the vector of parameters on the expenditure aggregation bias terms was not significantly different from the zero vector at a level of type I error ≤ 0.10. Finally, the preceding reestimation of the model was repeated except that the three taste/structural change variables were eliminated. In this case the zero vector hypothesis for the parameters of the aggregation bias terms was rejected through a Wald test (probability value = 0.03). These observations suggest that the time-trends and indicator variables were proxies for the expenditure aggregation bias correction to some degree as anticipated in the previous discussion. As a result, the effect of the expenditure aggregation bias terms on estimates of demand response is moderate but notable.

Conclusions

We have presented a new method of estimating the income distribution characteristics necessary for an appropriate empirical specification of an aggregate AIDS demand model. The method is straightforward to implement and does not require extensive cross-sectional information on households. The procedure is applicable whenever survey or census information is available on the proportions of households belonging to a discrete number of income categories, which is readily available in most developed economies.

The empirical example suggests the effects of expenditure aggregation bias can confound the interpretation of the taste/structural change variables to the extent that the latter proxies the former. One might argue that this is not a serious complication and might even consider this proxy characteristic to be a virtue of time-trends and indicator variables if one's main interest centers not on an analysis of the effects of taste/structural change but rather on the effects of changing prices and income. However, one cannot rely on the fortunate happenstance of taste/structural change variables fully respresenting expenditure aggregation bias effects. Furthermore, designing time-trends and indicator variables for the explicit purpose of modeling expenditure aggregation bias would appear to be a circular and counterproductive exercise since the effectiveness of such a modeling effort will rely on knowledge of the expenditure aggregation bias terms themselves.

Both the conceptual and empirical models underscore an important dichotomy in the types of aggregation bias effects that must be considered in the specification of aggregate demand: (*a*) expenditure aggregation effects, and (*b*) aggregate taste effects induced by either a changing income distribution or changing household preferences, or both. The degree to which the expenditure aggregation effect is fully accounted for is solely dependent on the accuracy with which households' income distribution can be estimated. However, the aggregate taste effect also depends on the distribution of taste differences across households as well as on changes in this distribution over time. Lack of data relegates the modeling of aggregate taste effects to proxy variables, including functions of time, indicator variables, and/or functions of sociodemographic variables, as in Blundell, Pashardes, and Weber (1993). An important topic for future research is the degree to which the functional representation of aggregate taste effects can be refined in the absence of detailed information on the distribution of household tastes.

We hope that the accessibility of our method will create opportunities for widespread use of expenditure aggregation bias correction in aggregate demand analyses. Such aggregation-bias-corrected models will not only be theoretically consistent but should lead to more precise estimates of demand model parameters and provide the researcher with a way to more accurately segregate income distributional effects from taste/structural change effects.

[Received October 1994; final version received May 1996.]

References

Andrews, D. W. K. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61(1993):821–56.

Blundell, R., P. Pashardes, and G. Weber. "What Do We Learn about Consumer Demand Patterns from Micro Data?" *Amer. Econ. Rev.* 83(1993):570–97.

Braschler, C. "The Changing Demand Structure for Pork and Beef in the 1970s: Implications for the 1980s."
    S. J. Agr. Econ. 15(1983):105–10.

Chalfant, J. "A Globally Flexible, Almost Ideal Demand for Applied Demand System." *J. Bus. and Econ.
    Statis.* 5(1987):233–42.

Champernowne, D. G. "A Model for Income Distribution." *Econ. J.* 53(1953):318–51.

Chavas, J.-P. "Structural Change in the Demand for Meat." *Amer. J. Agr. Econ.* 65(1983):148–53.

Dagum, C. "A New Model of Personal Income Distribution: Specification and Estimation." *Economie Ap-
    pliquée* 30(1977):413–37.

Dahlgran, R. "Complete Flexibility Systems and the Stationarity of U.S. Meat Demands." *West. J. Agr. Econ.*
    12(1987):152–63.

Deaton, A., and J. Muellbauer. "An Almost Ideal Demand System." *Amer. Econ. Rev.* 70(1980)312–26.

Eales, J., and L. Unnevehr. "Demand for Beef and Chicken Products: Separability and Structural Change."
    Amer. J. Agr. Econ. 70(1988):521–32.

———. "Simultaneity and Structural Change in U.S. Meat Demand." *Amer. J. Agr. Econ.* 75(1993):259–
    68.

Esteban, J. "Income Share Elasticity and the Size Distribution of Income." *Internat. Econ. Rev.* 27(1986)439–
    44.

Gorman, W. M. "Community Preference Fields." *Econometrica* 21(1953):63–80.

———. "Separable Utility and Aggregation" *Econometrica* 27(1959):469–81.

Green, H. A. J. *Aggregation in Economic Analysis.* Princeton NJ: Princeton University Press, 1964.

Majumder, A., and S. R. Chakravarty. "Distribution of Personal Income: Development of a New Model and
    Its Application to U.S. Income Data." *J. Appl. Econometrics* 5(1990):189–96.

McDonald, J. B. "Some Generalized Functions for the Size Distribution of Income" *Econometrica* 52(1984):
    647–63.

McDonald, J. B., and M. R. Ransom. "Alternative Parameter Estimators Based upon Grouped Data." *Com-
    munications in Statistics-Theory and Methods* A8(1979):899–917.

Moschini, G. "Units of Measurement and the Stone Index in Demand System Estimation." *Amer. J. Agr.
    Econ.* 77(1995):63–68.

Moschini, G., and K. D. Mielke. "The U.S. Demand for Beef—Has There Been a Structural Change?" *West.
    J. Agr. Econ.* 9(1984):271–82.

Muellbauer, J. "Aggregation, Income Distribution and Consumer Demand." *Rev. Econ. Stud.* 42(1975)525–
    543.

———. "Community Preferences and the Representative Consumer." *Econometrica* 44(1976)979–99.

Salem, A. B. Z., and T. D. Mount. "A Convenient Descriptive Model of Income Distribution." *Econometrica*
    42(1974)1115–127.

Samuelson, P. A. *Foundations of Economic Analysis.* Cambridge MA: Harvard University Press, 1947.

Singh, S. K., and G. S. Maddala. "A Function for Size Distribution of Incomes." *Econometrica* 44(1976):
    963–70.

Stoker, T. M. "Simple Tests of Distribution Effects on Macroeconomic Equation." *J. Polit. Econ.* 94(1986):
    763–95.

Theil, H. *Linear Aggregation of Economic Relations.* Amsterdam: North Holland Pub. Co., 1954.

———. *Principles of Econometrics.* New York: John Wiley and Sons, 1971.

Thurman W. "The Poultry Market: Demand Stability and Industry Structure." *Amer. J. Agr. Econ.* 69(1987):
    30–37.

U.S. Department of Commerce, Bureau of the Census. "Current Population Reports: Consumer Income, Series
    P-60, Washington DC: Government Printing Office. Various issues, 1963–89.

———. Bureau of Labor Statistics. "Employment, Hours, and Earning." Washington DC: Government Print-
    ing Office. Various issues, 1963–89.

U.S., President. "Economic Report of the President." Washington DC: Government Printing Office, 1990.