Journal of Business & Economic Statistics, July 1990, Vol. 8, No. 3

Demand Systems Estimation With Microdata: A Censored Regression Approach

Dale Heien
Department of Agricultural Economics, University of California, Davis, CA 95616

Cathy Roheim Wessells
Department of Resource Economics, University of Rhode Island, Kingston, RI 02881

Demand systems estimation increasingly makes use of household-level microdata, mainly to measure the effects of demographic variables. Data based on these household-expenditure surveys present a major estimation problem. For any given household, many of the goods have zero consumption, implying a censored dependent variable. Techniques which do not take this censored dependent variable into account will yield biased results. We utilize a censored regression approach that is computationally simple, consistent, and asymptotically efficient. The results are then presented and compared with those obtained using an uncensored technique.

KEY WORDS: Demand for dairy products; Survey data; Zero consumption.

1. INTRODUCTION

Although early demand systems estimation was undertaken utilizing aggregate time series data, recent interest in demographic effects has led researchers to use cross-section data. Examples of recent demand systems research using household-level microdata include Barnes and Gillingham (1984) and Pitt (1983). The use of household-level data is preferable, since it avoids the problem of aggregation over consumers and often provides a large and statistically rich sample. The use of household data for detailed commodities, however, presents a major estimation problem. This problem stems from the fact that, for many items in the budget, households are observed to consume zero amounts of the various commodities under consideration. In this article, we present a technique for dealing with this problem. A data base which contains numerous zero observations for some items is purposely utilized and the estimator employed is a computationally simple two-step procedure. We compare the results of a demand system estimated as a censored regression with a system estimated by conventional methods. The estimated demand system encompasses 11 food items with emphasis on dairy products.

2. SYSTEMS ESTIMATION AS A CENSORED REGRESSION PROBLEM

The Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980) was selected as the specification for the demand system. The technique presented here can, in principle, be applied to any demand system. The AIDS demand system was chosen because of its flexibility and linearity and because it is a complete system; that is, it can be restricted to satisfy the conditions of adding up, homogeneity, and symmetry. The AIDS demand relations, in budget-share form, are given by

$$w_i = \alpha_i + \sum_{i=1}^{n} \gamma_{ij} \ln p_j + \beta_i \ln(m/P), \qquad i = 1, \ldots, n, \quad (1)$$

where m is total expenditure, p_j is the price of the jth good, and P is a price index given by

$$\ln P = \alpha_o + \sum_{i=1}^{n} \alpha_i \ln p_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln p_i \ln p_j. \quad (2)$$

It has long been recognized that food demand, especially for dairy products, is influenced by the age structure of the population and various other demographic factors. [See Boehm and Babb (1975), Heien and Wessells (1988), and Kinnucan (1986) and the references cited therein.] To incorporate demographic variables, the AIDS model was modified by specifying

$$\alpha_i = \rho_{io} + \sum_{k=1}^{s} \rho_{ik} d_k, \qquad i = 1, \ldots, n, \quad (3)$$

where ρ_{io} and the ρ_{ik}'s are parameters to be estimated and the d_k are the demographic variables, of which there are s. This method for incorporating demographic variables is known as demographic translation. The other widely used technique is demographic scaling. Translation preserves the linearity of the system, whereas scaling is a highly nonlinear specification. For a discussion of the two methods, see Pollak and Wales (1981).

The data base used in this study is the U.S. Department of Agriculture's (USDA's) 1977–1978 household food consumption survey (HFCS). This survey contains data on food consumption and the money value of food used at home during one week for 14,930 households of one or more members. Socioeconomic characteristics of the household are also contained in the data set and include urbanization, region, tenancy of residents, number and age of people living in the household, race of the respondent, recipients of food stamps, frequency of major food shopping occasions, and classification of household by sex of the head.

The survey was conducted by the Consumer Nutrition Center of the Human Nutrition Information Service. The basic sample was a multistage, stratified probability sample of all private households in the 48 conterminous states, stratified by region, urbanization, and geographic or demographic similarities. Separate surveys were conducted for Alaska, Hawaii, and Puerto Rico. Hence they are not represented in this survey. Separate surveys were also run for the elderly and those on food stamps. The survey used in this study does contain individuals in these categories, but they are not specifically targeted as they are in the special supplemental surveys. Within this sample, four independent samples were drawn and implemented in four separate quarters of data collection. Data collection for the samples took place April 1, 1977–March 31, 1978, with approximately 3,750 sample households surveyed during each three-month period. To choose the households, the 48 states were divided into 114 strata on the basis of three levels of stratification—geographic division, urbanization or zone, and demographic or other geographic similarities. Each of the nine geographic divisions that comprise the four census regions was divided into three census zones—central city, suburban, and nonmetropolitan. The procedure resulted in 114 homogeneous strata of approximately 600,000 households each. These strata were then divided into one or more primary sampling units formed from cities, counties, or portions of cities and counties containing at least 10,000 housing units. Each primary sampling unit was divided into area segments containing 100 or more housing units. From all of the primary sampling units, 2,550 segments were drawn. Estimation of occupied housing units permitted calculation of sampling ratios for the area segments that would yield 3,750 households per quarter. For each quarter, a sample was systematically selected from each segment, without replacement, after a random start.

The data were collected by interview with the household member identified as the person most responsible for food planning and preparation. Households were contacted at least seven days prior to the interview and asked to keep grocery receipts or other aids to help them recall the food purchased in the seven-day period. Interviewers then recorded the kind, form, quantity, and cost of each food or beverage purchased by the household during those seven days. The survey defines food consumption in terms of the products as they enter the kitchen, not after they are transformed into meals. A further discussion of this survey can be found in USDA (1983).

The sample size used in the estimation in this paper was 10,746. The entire sample of 14,930 could not be used because some households either did not report their income or had nonresponses in other categories. Some of the data chosen as the most relevant explanatory variables for food consumption for this study were the region of the country in which the household resided, the season or quarter of the year in which the response was given, the race of the respondent, the number of household members of each gender in each age group, the household size in 21-meal-at-home equivalents (defined by the USDA as the number of 21-meal-at-home equivalent persons in the household), the housing tenure of the respondent (i.e., owner or renter), and the employment status of the household. A complete list of household characteristics used as explanatory variables can be found in Table 1, along with the parameter estimates and their associated t ratios.

The data were aggregated into the following 11 categories: (a) milk (95.0%), (b) cheese (77.4%), (c) cottage cheese (27.9%), (d) butter (26.6%) and margarine (78.6%), (e) ice cream (49.2%), (f) coffee and tea (86.8%), (g) sodas and fruit ades (74.9%), (h) vegetable and citrus juices (74.6%), (i) meat (99.3%), and (j) all other food (98.8%). The percentages in parentheses give the proportion of households in the survey that consume the item in question. For example, 27.9% of the households in the survey consumed cottage cheese during the survey period.

This specification implies that the various food items are separable from the other (nonfood) items in the consumer's budget. The commodity selection and aggregation used here were done in conjunction with a previous study of demand focusing specifically on dairy products (Heien and Wessells 1988). That study did not, however, employ a censored regression technique. We use this configuration of aggregates for this study because it contains some goods with relatively low percentages of consuming households (e.g., cottage cheese and butter) and others with relatively high percentages. This heterogeneity aids in comparing the censored regression results with the conventional techniques.

As indicated previously, not every household consumed something in every category. The figures in parentheses indicate the proportion of the total sample for which observations on that item are nonzero. From a theoretical point of view, demand is constrained to be nonnegative. Houthakker (1954) initially recognized the nonnegativity constraint and treated it as a special case of rationing. Recently Wales and Woodland

Table 1. Estimates of AIDS Food-Demand System Using the Censored-Regression Method

Variable	Mean	Milk	Cheese	Cottage cheese	Butter	Margarine	Ice cream	Coffee and tea	Sodas and fruit ades	Vegetable and citrus juices	Meat	All other food at home
Intercept		.14000	.03100	.01230	.01028	.01470	.01976	.09972	.05254	.07379	.09126	.45465
		(28.56)	(11.89)	(13.97)	(13.77)	(19.52)	(13.45)	(23.20)	(15.13)	(27.78)	(9.52)	(45.24)
Percent of meals at home	89.300	.00007	.00003	.00002	.000004	.00002	.00004	−.00001	−.00007	−.000009	.00029	−.00039
		(3.16)	(2.22)	(5.43)	(1.34)	(6.75)	(5.59)	(.43)	(4.16)	(.78)	(6.35)	(.60)
Location												
Metro	.356	.002	.00160	.00006	.00012	.00035	.00139	.00086	.00116	−.00292	−.00573	.00111
		(1.36)	(2.50)	(.26)	(.65)	(1.89)	(3.72)	(.76)	(1.20)	(4.18)	(2.13)	(.45)
Rural	.338	.00600	−.00350	.00045	−.00127	.00133	.00109	.00254	−.00015	−.00297	−.00998	.00646
		(4.06)	(5.10)	(1.95)	(6.67)	(6.90)	(2.79)	(2.15)	(.15)	(4.11)	(3.58)	(2.36)
Season												
Spring	.243	−.00300	−.00260	.00026	−.00025	−.00120	.00069	−.00528	.00682	−.00262	.009596	−.00104
		(1.91)	(3.61)	(1.07)	(1.24)	(5.8)	(1.71)	(4.27)	(6.54)	(3.48)	(3.31)	(.92)
Summer	.231	−.00400	−.00170	.00015	−.00019	−.00110	.00112	−.00305	.00913	−.00356	−.02007	.02377
		(2.36)	(2.33)	(.59)	(.95)	(5.40)	(2.72)	(2.85)	(8.65)	(4.68)	(6.85)	(7.48)
Fall	.266	.00100	−.00016	−.00024	.00023	−.00017	.00045	−.00125	.00124	.000597	.00167	−.00337
		(.91)	(.23)	(1.04)	(1.19)	(.88)	(1.15)	(1.04)	(1.22)	(.82)	(.59)	(1.22)
Region												
North central	.242	−.00200	−.00190	.00127	−.00312	.00209	−.00053	.001459	.00660	−.00761	.00253	.00121
		(1.01)	(2.52)	(4.95)	(14.76)	(9.79)	(1.23)	(1.13)	(6.06)	(9.60)	(.83)	(.21)
South	.339	.00090	−.00460	−.00288	−.00477	.00130	.00078	.00612	.00615	−.00679	.01501	−.01122
		(.66)	(6.57)	(11.96)	(24.04)	(6.51)	(1.93)	(5.03)	(5.96)	(9.08)	(5.25)	(3.73)
West	.173	.00400	.00200	.00223	−.00370	.001297	−.00037	−.005489	.00276	−.00149	−.01515	.00224
		(2.31)	(2.52)	(7.73)	(15.76)	(5.55)	(.79)	(3.95)	(2.34)	(1.75)	(4.65)	(4.14)
Tenancy: Owner	.660	.00100	−.00180	.00093	−.00009	−.00064	.00468	.00168	−.00510	.00069	−.00452	.00189
		(.85)	(2.97)	(4.38)	(.51)	(3.63)	(13.01)	(1.54)	(5.55)	(1.05)	(1.77)	(1.63)
Male occupation												
Professional	.127	.00300	.00120	−.00006	−.00023	.00032	.00355	−.00231	.00198	.00138	−.00603	−.00280
		(1.05)	(.96)	(.14)	(.67)	(.91)	(5.07)	(1.08)	(1.10)	(1.06)	(1.20)	(.44)
Manager	.101	−.29700	.00490	−.00081	.00031	−.00139	−.00005	−.00574	−.000168	.00117	−.00408	.30285
		(1.02)	(3.32)	(1.61)	(.74)	(3.33)	(.62)	(2.24)	(.08)	(.75)	(.68)	(1.47)
Farmer	.016	−.00400	.00270	−.00104	−.00012	−.000935	−.00006	−.00409	.00426	.00050	.00310	−.00032
		(1.44)	(.176)	(2.00)	(.29)	(2.16)	(.07)	(1.54)	(1.89)	(.31)	(.50)	(.01)
Clerical	.050	−.00030	−.00180	−.00114	.00401	−.00180	−.00041	−.00420	−.00388	.00134	.02500	−.01732
		(.66)	(.83)	(1.52)	(6.51)	(2.88)	(.33)	(1.09)	(1.19)	(.57)	(2.82)	(1.82)
Craftsman	.011	−.00200	.00040	−.00189	.00016	−.00056	.00016	.00110	.00344	−.00211	.002496	−.00120
		(.45)	(.20)	(3.29)	(.33)	(1.17)	(.17)	(.37)	(1.38)	(1.17)	(.36)	(.22)
Operative	.075	−.00300	.00110	−.00191	−.00022	−.00049	−.00055	.00032	.00445	−.00476	.006195	−.00114
		(1.21)	(.80)	(4.05)	(.56)	(1.25)	(.69)	(.13)	(2.18)	(3.24)	(1.09)	(.14)
Service	.061	−.00200	−.00076	−.00211	−.00051	−.00026	.00065	.00329	.00447	−.00351	.00603	.00529
		(.72)	(.49)	(4.02)	(1.18)	(.59)	(.73)	(1.23)	(1.97)	(2.14)	(.95)	(.76)
Food shopper												
Female	.708	−.00800	−.00005	−.00229	−.000299	.000084	.00065	−.00125	.00051	−.00486	.00905	.00646
		(2.38)	(.33)	(4.26)	(.68)	(.19)	(.72)	(.45)	(.22)	(2.90)	(1.39)	(.87)
Male	.105	−.00020	.00300	−.00015	.00033	−.00029	.00074	−.00210	.00392	−.00145	.00352	−.00732
		(.13)	(5.33)	(.76)	(2.07)	(1.77)	(2.26)	(2.11)	(4.66)	(2.40)	(1.51)	(3.01)
Female and male	.133	.00060	−.00300	−.00123	−.00158	−.00246	−.00063	.00164	−.00073	.00187	.02850	−.02298
		(.23)	(2.41)	(2.92)	(4.54)	(7.00)	(.89)	(.76)	(.41)	(1.42)	(5.63)	(4.30)
Ethnic												
Black	.117	−.00490	.00140	−.00045	−.000024	.000593	.00064	−.00241	.00248	.00212	.01400	−.01345
		(2.21)	(1.24)	(1.19)	(.08)	(1.89)	(1.00)	(1.25)	(1.52)	(1.80)	(3.08)	(2.81)
Spanish	.048	−.00240	.00220	−.00154	−.000496	.000319	.00235	−.00702	.00488	.00365	.00280	−.00474
		(.85)	(1.52)	(3.15)	(1.2)	(.78)	(2.85)	(2.79)	(2.29)	(2.39)	(.48)	(.75)
Pregnancy	.021	−.00480	.00120	−.00083	−.000008	.000887	.00126	−.00394	.00245	.00101	.00926	−.00952
		(1.89)	(.92)	(1.89)	(.02)	(2.44)	(1.72)	(1.76)	(1.30)	(.74)	(1.76)	(1.17)
Nursing	.011	−.02500	−.01070	−.00451	−.00012	−.00187	−.00062	−.01872	−.00155	.00065	.10740	−.04496
Employment												
Male employed	.572	.00750	−.00198	−.000497	−.00039	.000605	−.00156	−.00730	−.00122	.00277	−.00994	.01201
		(2.14)	(1.12)	(.83)	(.81)	(1.21)	(1.55)	(2.37)	(.47)	(1.48)	(1.37)	(11.94)
Female employed	.383	−.00700	.00350	−.00212	.000611	−.00077	−.00313	−.00330	−.00048	.00603	.00012	.00654
		(1.41)	(1.42)	(2.50)	(.87)	(1.08)	(2.19)	(.75)	(.13)	(2.28)	(.02)	(1.57)
Number of male household members between ages of												
0–10	.304	.01540	−.00160	.00014	.000084	.00053	−.000009	−.00077	.00104	.00099	−.00655	−.00718
		(17.15)	(3.54)	(.92)	(.66)	(4.17)	(.03)	(.99)	(.157)	(2.09)	(3.55)	(4.77)
11–20	.289	.01420	.00029	.00031	.00026	.00069	.00136	−.00188	.00232	.00151	.00299	−.01607
		(15.365)	(.62)	(1.95)	(2.02)	(5.22)	(5.14)	(2.32)	(3.39)	(3.069)	(1.57)	(8.03)
21–35	.316	.00930	.00100	−.00033	.00046	.00062	.00059	.−00054	.00275	.00103	.00886	−.02374
		(6.15)	(1.33)	(1.29)	(2.15)	(2.89)	(1.35)	(.41)	(2.45)	(1.27)	(2.85)	(7.24)
36–50	.226	.00430	−.00088	−.00045	.00056	.00068	.00062	.00919	.000395	.00034	.01112	−.02588
		(2.25)	(.93)	(1.38)	(2.10)	(2.54)	(1.15)	(5.54)	(.28)	(.34)	(2.85)	(6.29)
Over 51	.293	.00840	−.00130	−.00071	.00016	.000597	−.00018	.00992	−.00137	−.00104	.01389	−.02837
		(5.26)	(1.59)	(2.59)	(.72)	(2.63)	(.39)	(7.11)	(1.16)	(1.22)	(4.24)	(8.21)

Table 1. (continued)

Variable	Mean	Milk	Cheese	Cottage cheese	Butter	Margarine	Ice cream	Coffee and tea	Sodas and fruit ades	Vegetable and citrus juices	Meat	All other food at home
Number of female household members between ages of												
0–10	.299	.01350	−.00140	.000092	−.00006	.00043	.00043	−.00041	−.00005	.00016	−.00715	−.00583
		(15.08)	(3.20)	(.60)	(.50)	(3.35)	(1.68)	(.53)	(.71)	(.33)	(3.88)	(2.85)
11–20	.295	.00940	.000198	.000056	.00016	.00055	.00066	−.00129	.00249	.00133	−.00273	−.01183
		(10.21)	(.43)	(.36)	(1.25)	(4.24)	(2.49)	(1.60)	(3.69)	(2.73)	(1.45)	(5.44)
21–35	.364	.00170	.00240	−.00014	.00017	.00058	−.00032	.00031	.00333	.00139	.00674	−.01616
		(1.04)	(2.87)	(.63)	(.75)	(2.46)	(.67)	(.21)	(2.73)	(1.58)	(1.98)	(4.50)
36–50	.247	.00380	.00011	−.00043	.00001	.00075	−.00053	.00722	.00210	.00077	.00890	−.02270
		(1.99)	(.11)	(1.29)	(.04)	(2.73)	(.96)	(4.28)	(1.48)	(.75)	(2.25)	(5.45)
51+	.364	.00160	−.00270	.00061	.00005	.00151	−.00060	.00844	−.00381	.00098	.00631	−.01239
		(.97)	(3.29)	(2.13)	(.21)	(6.39)	(1.26)	(5.81)	(3.11)	(1.11)	(1.85)	(3.43)
Food stamp participant	.071	.01980	−.00110	−.00311	−.00168	−.00005	.00063	−.00363	.00096	−.00411	.01137	−.01908
		(5.28)	(.56)	(4.86)	(3.19)	(.85)	(.58)	(1.11)	(.35)	(2.06)	(1.47)	(2.35)
Value of food stamps	8.510	−.00009	.00001	.00001	.000002	.000005	.0000001	.000033	.000028	.00001	−.00005	.00004
		(3.45)	(.79)	(2.33)	(.65)	(1.40)	(.01)	(1.42)	(1.44)	(.79)	(.95)	(.74)
Price												
Milk	.253	.01380	−.00400	−.00006	−.00113	−.00060	−.00099	.00355	.00009	.00116	−.01606	.00064
		(11.80)	(1.70)	(.29)	(6.45)	(3.48)	(3.02)	(5.23)	(1.77)	(2.18)	(11.38)	(.45)
Cheese	1.700	.00000	.01710	−.00009	−.00043	−.00092	−.00069	−.000069	−.00084	−.00067	−.01053	.00245
		(.0002)	(20.18)	(.31)	(1.70)	(3.83)	(1.93)	(.16)	(2.51)	(1.52)	(12.23)	(3.29)
Cottage cheese	.845	−.00006	−.00009	.00608	−.00178	.00052	−.00014	.000215	−.00025	−.00023	−.00136	−.00291
		(.29)	(.31)	(14.39)	(5.89)	(2.34)	(.78)	(1.35)	(2.08)	(1.33)	(4.30)	(9.15)
Butter	1.229	−.00110	−.00043	−.00178	.00449	.00036	.00043	−.000647	.00013	.00005	−.00105	−.00019
		(6.45)	(1.70)	(5.89)	(10.58)	(1.72)	(2.86)	(4.89)	(1.27)	(.32)	(4.01)	(.63)
Margarine	.637	−.00060	−.00092	.00053	.00036	.00848	−.00045	.000214	.000475	.00049	−.00409	−.00449
		(3.48)	(3.83)	(2.34)	(1.72)	(35.38)	(3.08)	(1.63)	(4.77)	(3.40)	(15.68)	(17.16)
Ice cream	.891	−.00099	−.00069	−.00014	−.00043	−.00045	.01189	.000022	.00016	.00030	.00000	−.00961
		(3.02)	(1.93)	(.78)	(2.86)	(3.08)	(27.11)	(.09)	(.81)	(1.15)	(.000)	(17.09)
Coffee and tea	5.390	.00350	−.00007	.00022	−.00065	.00021	.000022	−.00131	.00139	.00123	−.00441	−.00013
		(5.23)	(.16)	(1.35)	(4.89)	(1.63)	(.08)	(1.66)	(3.14)	(2.90)	(3.80)	(.16)
Sodas and fruit ades	.491	.00095	−.00084	−.00025	−.00013	.00048	−.00016	.00139	−.00399	−.00015	−.00137	.00407
		(1.77)	(2.51)	(2.07)	(1.27)	(4.77)	(.82)	(3.14)	(7.55)	(.44)	(1.46)	(4.20)
Vegetable and citrus juices	.453	.00120	−.00067	−.00024	.00005	.00050	−.00030	.00123	−.00015	.00275	−.00564	.00127
		(2.17)	(1.51)	(1.33)	(.32)	(3.40)	(1.15)	(2.89)	(.44)	(5.05)	(6.71)	(1.53)
Meat	1.151	−.01606	−.01053	−.00136	−.00105	−.00410	−.00490	−.00441	−.00137	−.00564	.15941	−.10999
		(11.38)	(12.23)	(4.30)	(4.01)	(15.68)	(9.67)	(3.80)	(1.46)	(6.71)	(46.60)	(17.10)
All other food at home	.494	−.000645	−.00287	−.00291	−.00017	−.00449	−.00471	−.00019	.00404	.00131	−.11490	.12554
		(.45)	(3.29)	(9.15)	(.63)	(17.16)	(9.21)	(.16)	(4.20)	(1.53)	(39.03)	(34.67)
Expenditure	45.023	−.01800	−.00320	−.00191	−.00108	−.00234	−.00505	−.01148	−.00693	−.00958	.01896	.04061
		(15.17)	(5.29)	(9.24)	(6.29)	(13.69)	(14.58)	(11.08)	(7.92)	(15.07)	(7.78)	(15.81)
Mills ratio	.000	.03620	.02270	.01347	.01075	.00716	.02309	.02512	.02601	.02309	.09625	−.28384
		(24.41)	(61.66)	(112.63)	(109.03)	(68.7)	(85.39)	(33.827)	(46.91)	(61.51)	(13.94)	(39.84)
Mean dependent variable		.07570	.02090	.00600	.00470	.00950	.01330	.04400	.03520	.02670	.34530	.41070
R^2		.159	.315	.568	.542	.393	.488	.192	.255	.301	.297	—
Standard error		.05180	.02590	.00008	.00005	.00730	.01500	.04520	.00145	.02747	.10640	—

NOTE: The numbers in parentheses below the coefficients are the *t* ratios.

(1983), Lee and Pitt (1986), and Ransom (1987) explored the econometric implications of this constraint, basing their approach on the Kuhn–Tucker conditions for nonnegativity. Ransom (1987) proposed using the Amemiya (1974) estimator when the error terms are not heteroscedastic. Other models for dealing with the zero observations problem have been proposed by Deaton and Irish (1984), Kay, Keen, and Morris (1984), Keen (1986), and Blundell and Meghir (1987). These models are based on the discrepancy between observed expenditure and actual consumption. Our approach is to follow Ransom's suggestion and use the Amemiya estimator. A related problem is that the estimated budget shares may not lie between 0 and 1. Woodland (1979) investigated this problem using a Dirichlet distribution to insure the restriction. We make no attempt to implement this restriction here.

The fact that the observed budget shares cannot take on negative values means that the dependent variable is censored. The problem of censored dependent variables was first recognized by Tobin (1958), who showed that the use of ordinary least squares (OLS) for such models results in biased and inconsistent estimates. The technique presented here was first utilized by Gronau (1974) and Lewis (1974) in a single-equation context analyzing labor market behavior. For single-equation models, work by Chung and Goldberger (1984), Greene (1981), and Olsen (1980) indicated that the bias is proportional to the probability of a limit observation; however, the problem here is more complex, involving a set of demand relations interrelated both through the error structure and the cross-equation restrictions. It is possible to estimate models of this type by maximum likelihood, but such procedures generally are not computationally feasible or reliable.

Lee (1978) generalized the two-step Amemiya (1974) estimator to a simultaneous-equation model, which consists of observable endogenous variables, unobservable latent endogenous variables with dichotomous indicators, and limited and censored dependent variables, as well as continuous endogenous variables. The number of equations is arbitrary, and tobit, probit, and censored models are special cases of this general model. Lee proved that the two-stage estimators resulting from this procedure are asymptotically more efficient than other two-stage estimators—namely, those by Nelson and Olsen (1978) and Heckman (1978).

Our particular application involves the special case of the censored simultaneous-equation model in which the dependent variables are censored by a subset of unobservable latent variables. The dependent variables, which are the budget shares for the 11 categories specified, are either 0 or some positive amount for each household. Those shares that are 0 are censored by an unobservable latent variable that induces the decision not to purchase that particular item in the weeklong survey period. The decision to buy or not to buy can be indicated by a binary indicator variable, which is a function of the latent variables and is estimated as a probit model (see Lee 1978, p. 354). The assumptions underlying this model (and its proofs) are that the error terms from the model are approximately normal with zero means and a finite variance–covariance matrix that is constant over all observations—that is, iid.

The estimation procedure then involves two steps. First, a probit regression is computed that determines the probability that a given household will consume the good in question. This regression is then used to compute the inverse Mills ratio for each household. The inverse Mills ratio is used as an instrument that incorporates the censoring latent variables in the second-stage estimation of the demand relations.

In the first stage, the decision to consume is modeled as a dichotomous choice problem,

$$Y_{ih} = f(p_{1h}, \ldots, p_{nh}, m_h, d_{1h}, \ldots, d_{sh}), \quad (4)$$

where Y_{ih} is 1 if the hth household consumes the ith food item, (i.e., if $w_{ih} > 0$) and 0 if the household does not consume the item in question. The other variables are as defined previously. Little, if any, theoretical work has been done regarding the specification of (4); however, prices and demographic effects should play roles similar to those expected in traditional demand analysis. In addition, food expenditure is included in the specification, since Jackson (1984) showed that variety is an increasing function of income, here proxied by expenditure. It can be argued that if the interview period were longer, more items would be observed entering the consumer's market basket. This is especially true for those food categories that include storable items.

The model given by (4) was estimated using the probit technique for each of the 11 items in the food budget. Goodness-of-fit measures such as the pseudo R^2 are found in Table 2. Highly significant variables in all of the probit regressions were deflated expenditures, the proportion of meals eaten at home, and the dummy variables for season, region, and race. In many cases, the own price, use of food stamps, and presence of a pregnant household member were also significant. The estimates of the probit regressions are not given here but are available on request from the authors.

Next, for the ith food item for the hth household, which consumes the item, the inverse Mills ratio

$$R_{ih} = \phi(\mathbf{p}_h, \mathbf{d}_h, m_h)/\Phi(\mathbf{p}_h, \mathbf{d}_h, m_h) \quad (5)$$

was computed, where $\mathbf{p}_h$ is a vector of prices for the

Table 2. Comparison Statistics for Seemingly Unrelated Regression Estimates of AIDS Food Demand System

Category	Budget share	Proportion nonzero observations	Probit psuedo R^{2*}	R^2		Own-price elasticity		Expenditure elasticity	
				Uncensored	Censored	Uncensored	Censored	Uncensored	Censored
Milk	.076	.95	.24	.113	.159	−.81	−.77	.78	.77
Cheese	.021	.77	.13	.061	.315	−.57	−.37	1.02	.89
Cottage cheese	.006	.28	.08	.058	.568	−1.01	−.03	1.05	.70
Butter	.005	.27	.09	.041	.542	−.91	−.00	1.10	.76
Margarine	.010	.79	.09	.129	.393	−.25	−.08	.79	.75
Ice cream	.013	.49	.03	.141	.488	−2.19	−.05	.98	.61
Coffee and tea	.044	.87	.13	.099	.192	−1.08	−1.01	.78	.73
Sodas and fruit ades	.035	.75	.12	.092	.255	−1.38	−1.10	.82	.79
Vegetable and citrus juice	.027	.75	.07	.049	.301	−1.06	−.87	.79	.62
Meat	.345	.99	.54	.280	.297	−.56	−.42	1.06	1.06
All other	.411	.99	—	—	—	−.77	−.89	1.04	1.10

NOTE: $\hat{l}(\Omega)$ is the value of likelihood function evaluated at maximum likelihood estimates. $\hat{l}(w)$ is the value of likelihood function under hypothesis $\beta_1 = \cdots = \beta_k = 0$.
* Pseudo $R^2 = 1 \ln \hat{l}(\Omega)/\ln \hat{l}(w)$.

hth household, $\mathbf{d}_h$ is a vector of the demographic variables for the hth household, and ϕ and Φ are the density and cumulative-probability functions, respectively. For those households who do not consume the item in question,

$$R_{ih} = \phi(\mathbf{p}_h, \mathbf{d}_h, m_h)/(1 - \Phi(\mathbf{p}_h, \mathbf{d}_h, m_h)). \quad (6)$$

The inverse Mills ratio for each item is then used as an instrument in the second-stage regression

$$w_{ih} = \rho_{io} + \sum_{k=1}^{s} \rho_{ik} d_{kh} + \sum_{j=1}^{n} \gamma_{ij} \ln p_{jh} + \beta_i \ln(m_h/Z_h) + \delta_i R_{ih}, \quad (7)$$

where, following Deaton and Muellbauer (1980),

$$Z_h = \sum_{i=1}^{n} w_i \ln p_{ih} \quad (8)$$

is used as an approximation to (2) so that the estimation remains linear. Equation (7) is the specification used to estimate the demand relations. Note that the specification given by (7) pertains only to the first $n - 1$ demand relations. It is well known that the variance–covariance matrix of error terms for the complete n equation demand system will be singular due to the adding-up property. The normal procedure is to delete one of the equations, since the parameters for that relation can be computed residually from the others. As Pollak and Wales (1969) have shown, the estimates are invariant to which good is dropped, but if all n relations are specified according to (7), the system will not add up. If all n equations were specified according to (7), adding up would require that $\Sigma_{i=1}^{n} d_i R_{ih} = 0$. Since R_{ih} can take on any value, such a restriction is not possible in general. To preserve the adding-up property, we specify the nth (deleted) relation as

$$w_{ih} = \rho_{io} + \sum_{k=1}^{s} \rho_{ik} d_{kh} + \sum_{j=1}^{n} \gamma_{ij} \ln p_{jh} + \beta_i \ln(m_h/Z_h) - \sum_{j=1}^{n-1} \delta_j R_{jh}. \quad (9)$$

The following restrictions of economic theory are readily imposed on this system: adding up—

$$\sum_{i=1}^{n} \alpha_i = 0; \qquad \sum_{i=1}^{n} \gamma_{ij} = 0, \qquad j = 1, \ldots, n; \qquad \sum_{j=1}^{n} \beta_i = 0; \quad (10)$$

homogeneity—

$$\sum_{j=1}^{n} \gamma_{ij} = 0, \qquad i = 1, \ldots, n; \quad (11)$$

and symmetry—

$$\gamma_{ij} = \gamma_{ji} \quad \text{for all } i, j (i = j). \quad (12)$$

Another problem arising due to zero consumption is that of missing prices. In order to estimate a complete system, prices must be available for all items for all households; however, for households not consuming a particular item, there will be no data on the price of that item. The procedure employed was to estimate the missing prices. This was done by performing a regression with the data on the price of the item from those households who did consume it. These regressions specified the price as a function of regional dummies, seasonal dummies, and income. These regressions were then used to estimate the missing prices for those households which did not consume that particular item. The properties of estimates using data obtained in this manner were discussed by Dagenais (1973) and Gourieroux and Monfort (1981). It should be pointed out that these properties hold only for noncensored variables.

3. ECONOMETRIC RESULTS AND CONCLUSIONS

The complete demand system given by (7) was estimated by seemingly unrelated regression (SUR) with the restrictions of economic theory, (9) and (10)–(12), imposed. The estimates of this system are given in Table 1. The demand system was also estimated by SUR under restrictions (9) and (10)–(12) but not including the inverse Mills ratio—that is, (7) with the R_{ih}'s deleted. Due to space limitations, this set of estimates is not given, but several statistics that enable comparisons between the two sets of estimates are given in Table 2. One of the most striking features of the comparison between the two estimation techniques is the improvement in the goodness of fit of the censored model. The average proportionate increase in goodness of fit was over fivefold as measured by the R^2.

Equally interesting was the change in the own-price elasticities, which are also given in Table 2. For items of which the number of zero observations was quite small (e.g., meat and milk), the own-price elasticity changed very little, but for items such as cottage cheese and butter which have a large proportion of zero responses, the elasticities changed dramatically bearing out the results noted previously that the bias is proportional to the probability of a limit observation. It is worth noting that demand became more inelastic under the censored regression estimation for all items except all other foods. The budget shares for cottage cheese, butter, margarine, and ice cream are all relatively small, indicating a priori that demand for these products is likely to be inelastic. Demand for these four products is highly inelastic under the censored estimation but was elastic in the uncensored regime for ice cream and cottage cheese and nearly so for butter. Expenditure elasticities for the censored model are lower for all items except meat and all other foods, which have virtually no zero observations.

The differences in the cross-price elasticities between the censored and uncensored regressions were generally minimal with two exceptions. The uncensored demand system tends to overstate the substitutibility and understate the complementarity between food items specified in (a) the demand equations for those items with the smallest budget shares—that is, butter, cottage cheese, margarine, and ice cream—and (b) all demand equations with respect to the price of meat and the price of all other goods. Note also that there were minor changes in the demographic effects between the two estimation techniques and that the demographic effects were in general agreement with expectations in so far as a priori expectations can be formed.

Hence, on the basis of prior reasoning and empirical results, we conclude that treating microdata demand systems as censored multiple-regression systems provides substantially improved results. These improved results are demonstrable in terms of goodness of fit and the conformity of price elasticities with prior expectations.

ACKNOWLEDGMENTS

The comments from an associate editor and two anonymous referees were useful in revising the article, which originated as Giannini Foundation Research Paper 941.

[*Received October 1988. Revised January 1990.*]

REFERENCES

Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," *Econometrica*, 42, 999–1012.

Barnes, R., and Gillingham, R. (1984), "Demographic Effects in Demand Analysis: Estimation of the Quadratic Expenditure System Using Microdata," *The Review of Economics and Statistics*, 66, 591–601.

Blundell, R., and Meghir, C. (1987), "Bivariate Alternatives to the Tobit Model," *Journal of Econometrics*, 34, 179–200.

Boehm, W., and Babb, E. (1975), "Household Consumption of Beverage Milk Product," Bulletin 75, Purdue University, Agricultural Experiment Station.

Chung, C.-F., and Goldberger, A. (1984), "Proportional Projections in Limited Dependent Variable Models," *Econometrica*, 52, 531–534.

Dagenais, M. G. (1973), "The Use of Incomplete Observations in Multiple Regression Analysis," *Journal of Econometrics*, 1, 317–328.

Deaton, A., and Irish, M. (1984), "Statistical Models for Zero Expenditures in Households Budgets," *Journal of Public Economics*, 23, 59–80.

Deaton, A., and Muellbauer, J. (1980), "An Almost Ideal Demand System," *The American Economic Review*, 70, 312–326.

Gourieroux, C., and Monfort, A. (1981), "On the Problem of Missing Data in Linear Models," *Review of Economic Studies*, 48, 579–586.

Greene, W. H. (1981), "On the Asymptotic Bias of the Ordinary Least-Squares Estimator of the Tobit Model," *Econometrica*, 49, 505–514.

Gronau, R. (1974), "Wage Comparisons: A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1144.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931–959.

Heien, D. M., and Wessells, C. R. (1988), "The Demand for Dairy Products: Structure, Prediction, and Decomposition," *American Journal of Agricultural Economics*, 70, 219–228.

Houthakker, H. S. (1954), "La Forme des Courbes d'Engel," *Cahiers du Seminaire d'Econometrie*, 2, 59–66.

Jackson, L. F. (1984), "Hierarchic Demand and the Engel Curve for Variety," *The Review of Economics and Statistics*, 66, 8–15.

Kay, J. A., Keen, M. J., and Morris, C. N. (1984), "Estimating Consumption From Expenditure Data," *Journal of Public Economics*, 23, 169–182.

Keen, M. (1986), "Zero Expenditures and the Estimation of Engel Curves," *Journal of Applied Econometrics*, 1, 277–286.

Kinnucan, H. W. (1986), "Demographic Versus Media Advertising Effects on Milk Demand: The Case of the New York City Market," *Northeast Journal of Agricultural and Resource Economics*, 15, 66–74.

Lee, L.-F. (1978), "Simultaneous Equations Models With Discrete and Censored Dependent Variables," in *Structural Analysis of Discrete Data With Econometric Applications*, eds. P. Manski and D. McFadden, Cambridge, MA: MIT Press, pp. 346–364.

Lee, L.-F., and Pitt, M. M. (1986), "Microeconometric Demand Systems With Binding Nonnegativity Constraints: The Dual Approach," *Econometricia*, 5, 1237–1242.

Lewis, H. G. (1974), "Comments on Selectivity Biases in Wage Comparisons," *Journal of Political Economy*, 82, 1145–1156.

Nelson, F. D., and Olson, L. (1978), "Specifications and Estimation of a Simultaneous Equation Model With Limited Dependent Variables," *International Economic Review*, 19, 695–710.

Olsen, R. J. (1980), "Approximating a Truncated Normal Regression With the Method of Moments," *Econometrica*, 48, 1099–1105.

Pitt, M. M. (1983), "Food Preferences and Nutrition in Rural Bangladesh," *The Review of Economics and Statistics*, 65, 105–114.

Pollak, R. A., and Wales, T. J. (1969), "Estimation of the Linear Expenditure System," *Econometrica*, 37, 611–628.

——— (1981), "Demographic Variables in Demand Analysis," *Econometrica*, 49, 1533–1558.

Ransom, M. R. (1987), "A Comment on Consumer Demand System With Binding Non-negativity Constraints," *Journal of Econometrics*, 34, 355–359.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36.

U.S. Department of Agriculture (1983), "Food Consumption: Households in the United States, Seasons and Year 1977–78: Nationwide Food Consumption Survey 1977–78," Report H-6, U.S. Department of Agriculture, Human Nutrition Information Service, Consumer Nutrition Division.

Wales, T. J., and Woodland, A. D. (1983), "Estimation of Consumer Demand Systems With Binding Non-negativity Constraints," *Journal of Econometrics*, 21, 263–285.

Woodland, A. D. (1979), "Stochastic Specification and the Estimation of Share Equations," *Journal of Econometrics*, 10, 361–383.