U.S. Income Distribution and Gorman Engel Curves for Food

J. T. LaFrance and T. K. M. Beatty, University of California - Berkeley
R. D. Pope, Brigham Young University
G. K. Agnew, University of Arizona

Abstract

A method for nesting, estimating and testing for the rank and functional form of the income terms in an incomplete system of aggregable and integrable demand equations is derived. Information theory is applied to the problem of inferring the U.S. income distribution using annual time series data on quintile and top five percentile income ranges and intra-quintile and top five percentile mean incomes. Estimates for the year-to-year income distribution are combined with annual time series data on the U.S. consumption of and retail prices for twenty-one food items to estimate the rank and functional form of the income terms in U.S. food demand over the period 1919-95, excluding 1942-46 to allow for the structural impacts of World War II.

1. Introduction

Following Muellbauer's (1975) extension of the Gorman polar form to a nonlinear function of income to obtain the price independent generalized linear (PIGL) and price independent generalized logarithmic (PIGLOG) functional forms, much progress has been made in the past 25 years on aggregation theory in consumption. The Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980) implements Muellbauer's results to produce demands with budget shares expressed as functions of linear and quadratic terms in the logarithm of prices and a linear term in the logarithm of income. The AIDS and its linear approximation (LA-AIDS) have been linchpins in applied demand analysis since their introduction. Most applications of the AIDS and LA-AIDS either assume separability and estimate a complete system of demands for a disaggregate group of commodities as functions of prices for the goods in the group and total expenditure on the group, or estimate a complete system of demands with highly aggregated commodities as functions of aggregate price indices and total consumption expenditures (hereafter, income, which we denote by *m*).

Shortly after the article by Deaton and Muellbauer, in a remarkable and elegant contribution to the festschrift to Sir Richard Stone, Gorman (1981) derived the set of functional forms for demand models that can be written in terms of any additive set of functions of income. Any complete system of demand equations in the class of "Gorman Engel curves" must satisfy two properties in addition to homogeneity, adding up and symmetry. First, if the number of independent functions of income is at least three, then the functions all must be either (a) polynomials in income, (b) polynomials in some non-integer power of income, (c) polynomials in the natural logarithm of income, or (d) a series of sine and cosine functions of the natural logarithm of income. Second, the number of "linearly independent" functions of income in this class of demand systems at most equals three, where linear independence refers to the rank of the matrix of price functions that premultiply the income functions. One important implication is that theoretically consistent demand aggregation in models that have full column rank for this matrix requires three summary statistics from the distribution of income to estimate the demand parameters with aggregate data.

Gorman (1981) also conjectured that second-order polynomials are the most general non-degenerate cases of demand systems that have full rank three. Pursuing this conjecture by exploiting the methods of van Daal and Merkies (1989), Lewbell (1990) was able to show that all full rank three generalizations of Muellbauer's PIGL and PIGLOG demand models are quadratic forms analogous to the quadratic expenditure system (QES) developed by Howe, Pollak and Wales (1979) and perfected by van Daal and Merkies (1989). Lewbell (1990) also derived a full rank three trigonometric model.

All of the above results on the rank of the coefficient matrix and the functional form of the income terms in the class of Gorman Engel curve demand models require the adding up property of a complete demand system. However, often we are interested in the demands for a subset of goods that make up only part of the consumption budget. In such a case, separability is a strong assumption, and it is undesirable to impose strong restrictions without good reason or prior evidence. Without separability, there is little reason to impose the same functional form on the demand equations for the goods of interest and all of the other goods for which we have little or no price or quantity information. This implies that the above results cannot be applied directly to incomplete demand systems.

In an ambitious paper, Gorman (1965; 1995) considered the structure of the demands for groups of goods in which each group's total expenditure is a function of income and a set of aggregate price indices for each group, and derived the restrictions on the individual demand equations and the properties of the indirect utility function under this set of restrictions. Independently and more recently, but along a similar line of thought, Epstein (1982), LaFrance (1985) and LaFrance and Hanemann (1989) developed a theory for the *weak integrability* of the demands for a single proper subset of all goods that does not exhaust the consumer's budget, regardless of the number of prices that enter the demand equations. The conditions for weak integrability of an incomplete demand system are that the demands are positive valued, 0° homogeneous in all prices and income, the budget restriction takes the form of a strict inequality (not all of income is exhausted by the subset of goods under study), and the submatrix of Slutsky substitution terms associated with this subset of demands is symmetric and negative semidefinite. These conditions exhaust the properties implied by consumer theory for any proper subset of all goods and are necessary and sufficient for the recovery of the conditional preference functions (both direct and indirect) for those goods, with prices of all other goods acting as conditioning variables (LaFrance (1985); LaFrance and Hanemann (1989)). *Inter alia*, the set of incomplete demand models that satisfy weak integrability is much richer than the corresponding set of integrable complete demand systems.

This paper exploits the richness of the set of weakly integrable demand models to extend aggregation in nonlinear functions of income to incomplete demand systems for the PIGL and PIGLOG members of Gorman Engel curves. These extensions permit us to develop a method to nest weakly integrable LA-AIDS, AIDS, quadratic AIDS (QAIDS), quadratic PIGL (QPIGL), and extended QES[1] models to simultaneously test for and estimate both the rank and functional form of the income terms in aggregable incomplete demand systems.

As noted above, a full rank three Gorman Engel curve demand model requires three summary statistics from the income distribution, e.g., for a QPIGL model in expenditure form we need the cross-sectional means of $m_h^{1-\kappa}$, m_h, and $m_h^{1+\kappa}$, where m_h is the income level of family *h*, *h* = 1, ..., *H*, say, and κ is the PIGL coefficient on income, while for a QAIDS model we need the means of m_h, $m_h \ln(m_h)$, and $m_h[\ln(m_h)]^2$. To calculate these means, we need information on the distribution of income. The U.S. Bureau of the Census annually publishes the quintile ranges, intra-quintile means, top five-percentile lower bound for income, and the mean income within the top five-percentile range for all U.S. families. We use Bayesian methods to obtain annual information theoretic density functions that satisfy each of these percentile and conditional mean conditions for the period 1910-1999. These *maximum entropy* densities and the resulting food demand estimates are compared with those obtained from a truncated three-parameter lognormal distribution and a piecewise uniform distribution for each year.

The income distribution estimates are combined with aggregate annual time series data on per family U.S. food expenditures for 21 individual food items over the period 1919-1995, excluding 1942-1946 to account for the structural impacts of World War II.[2] In addition to annual measures of food expenditures, prices, and the income distribution, we incorporate measures for the distribution of the U.S. population by age and the ethnicity of the U.S. population in the incomplete demand model's specification. The results of the empirical application strongly suggest that a full rank three model is essential, and that the QAIDS is strongly rejected in favor of an extended QES.

The rest of the paper is organized as follows. The next section extends the aggregation results of Gorman and others to incomplete demand systems that can be written in a PIGL/PIGLOG form. The third section describes the estimates of the U.S. income distribution. Section 4 presents a summary and discussion of a subset of the empirical results, focusing primarily on the rank of the demand model and the functional form of the income terms. The final section summarizes the findings of the paper and discusses possible limitations of the analysis and possible directions for further research. Additional detailed derivations, discussions, and proofs of our main results are contained in an expanded paper that is available from the authors upon request.

2. Nesting LA-AIDS, AIDS and QAIDS within a QPIGL-IDS

In the two decades since its introduction by Deaton and Muellbauer, the AIDS has been widely used in demand analysis. The vast majority of empirical applications follows Deaton and Muellbauer's suggestion and replaces the translog price index that deflates income with Stone's index, which generates the LA-AIDS. Although Deaton and Muellbauer (1980: 317-320) cautioned against and avoided the practice, most empirical applications of the LA-AIDS include tests for and the imposition of an approximate version of Slutsky symmetry by restricting the matrix of logarithmic price coefficients to be symmetric. Important examples include Anderson and Blundell (1983), Buse (1998), Moschini (1995), Moschini and Meilke (1989), and Pashardes (1993).[3] In this section, we derive a simple method for nesting the weakly integrable LA-AIDS model within a general class of QPIGL demand models.

Let $\boldsymbol{p}$ be the *n*-vector of market prices for goods, let *u* be the utility index, let $e(\boldsymbol{p},u)$ be the consumer's expenditure function, and let $\boldsymbol{w}$ be the *n*-vector of budget shares. In this study, we nest the LA-AIDS, AIDS, LES, and PIGL demand models within a general rank three quadratic PIGL incomplete demand system (QPIGL-IDS). The *quasi-indirect utility function* (Hausman (1981); LaFrance (1985); LaFrance and Hanemann (1989)) for this model can be written in a form that is consistent with the QES originally developed in Howe, Pollak and Wales (1979),

$$(1) \qquad \varphi(\boldsymbol{p},m) =$$

$$-\left\{\frac{1}{m(\kappa)-\alpha_0-\boldsymbol{\alpha}'\boldsymbol{p}(\lambda)-\frac{1}{2}\,\boldsymbol{p}(\lambda)'\boldsymbol{B}\boldsymbol{p}(\lambda)}+\boldsymbol{\delta}'\boldsymbol{p}(\lambda)\right\}e^{\boldsymbol{\gamma}'\boldsymbol{p}(\lambda)}.$$

Applying Roy's identity to (1) generates a QPIGL-IDS in budget share form as

(2) $$w = m^{-\kappa} P^{\lambda} \Big\{ \alpha + B p(\lambda) + \gamma \big[m(\kappa) - \alpha_0 - \alpha' p(\lambda) - \tfrac{1}{2} p(\lambda)' B p(\lambda) \big] + \big[I + \gamma' p(\lambda) \big] \delta \big[m(\kappa) - \alpha_0 - \alpha' p(\lambda) - \tfrac{1}{2} p(\lambda)' B p(\lambda) \big]^2 \Big\}.$$

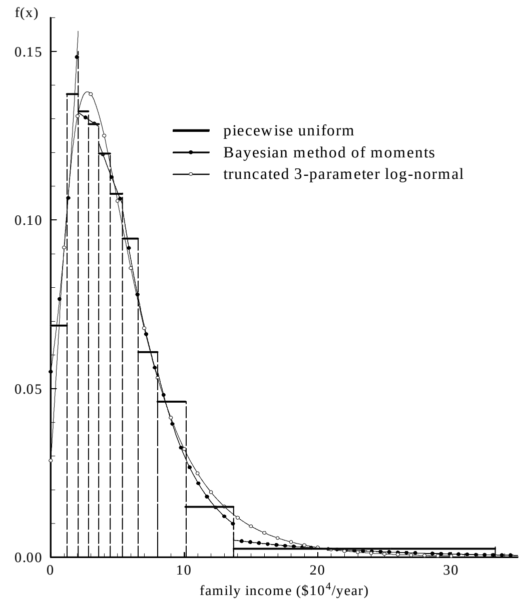
Assuming that α and $\mathbf{B}$ do not completely vanish simultaneously, it follows that: (a) $\gamma \neq \mathbf{0}, \delta \neq \mathbf{0}$ is necessary and sufficient for a full rank three QPIGL-IDS; (b) $\gamma \neq \mathbf{0}, \delta = \mathbf{0}$ is necessary and sufficient for a full rank two, non-homothetic PIGL-IDS; (c) $\gamma = \mathbf{0}, \delta \neq \mathbf{0}$ is necessary and sufficient for a full rank two QPIGL-IDS that excludes the linear term; and (d) $\gamma = \delta = \mathbf{0}$ is necessary and sufficient for a homothetic PIGL-IDS. Thus, we obtain a rich class of models that permits nesting, testing and estimating the rank and functional form of the income aggregation terms in incomplete demand systems.

3. Estimating the U.S. Income Distribution

When a demand model is nonlinear in income, the demand equations do not aggregate directly across individual decision units to average (per capita or per family) income at the market level. The advantage of the Gorman class of Engel curves is that, when information on the income distribution across economic units is available, only a small number of summary statistics from this distribution are required to obtain a theoretically consistent, aggregable demand model. Indeed, all full rank three Gorman Engel curve demand models require three summary statistics from the income distribution, e.g., a QPIGL requires the cross-sectional means of $m_h^{1-\kappa}$, m_h, and $m_h^{1+\kappa}$. To calculate these means, however, we need information on the distribution of income.

The U.S. Bureau of the Census publishes annually quintile ranges, intra-quintile means, the top five-percentile lower bound for income, and the mean income within the top five-percentile range for all U.S. families. These data are currently available for 1947-1998 on the U.S. Bureau of the Census World Wide Web site, and for the years 1929, 1935/36, 1941, 1944 and 1946 from the Census Bureau's historical statistics (U.S. Department of Commerce, 1972). Several issues arise regarding the use of these data to estimate the U.S. income distribution. First and perhaps foremost is an appropriate methodology for obtaining a reasonable density function given the probability ranges and intra-range means. In this paper, we consider three possibilities, depicted in figure 1 for 1997, which are developed and explained in this section.

Figure 1. U.S. Income Distribution, 1997.



A simple, naive and uninformative approach is to construct a sequence of piecewise uniform densities on each of the first four quintile ranges, the 85-95 percentile range, and the top five percentile range.[4] However, these piecewise uniform densities generally do not satisfy the intra-quintile and top five percentile mean conditions. A more informative solution is to construct a pair of uniform densities on each range, separated at the intra-range mean, and with total probabilities that sum to .20, .15, or .05, as appropriate. Letting $[\ell_{i-1}, \ell_i)$ denote the i^{th} income range, μ_i the i^{th} intra-range mean, and p_i the proportion of the total number of U.S. families whose incomes that fall within this range, we calculate a piecewise uniform density on $[\ell_{i-1}, \ell_i)$ that satisfies

(3) $$f(x) = \left(\frac{p_i}{\ell_i - \ell_{i-1}} \right) \begin{cases} \left(\dfrac{\ell_i - \mu_i}{\mu_i - \ell_{i-1}} \right), & \forall\ x \in [\ell_{i-1}, \mu_i) \\ \left(\dfrac{\mu_i - \ell_{i-1}}{\ell_i - \mu_i} \right), & \forall\ x \in [\mu_i, \ell_i) \end{cases}.$$

This density is illustrated in figure 1 for 1997 by the series of horizontal line segments.

The piecewise uniform density is *ad hoc* and discontinuous at eleven points[5]. We have a fixed (and small) number of observations in each year on quintile limits, intra-quintile means, and the top five percentile lower limit and mean, so we cannot appeal to properties like consistency. Therefore, alternative estimators warrant consideration. Two approaches are considered here. One is based on the principle of maximum entropy and information theory. This density is well-known to possess several desirable properties (Zellner 1988). This approach generates a piecewise exponential density that is smooth and monotone within each income range and satisfies the probability and intra-range mean conditions exactly, but is discontinuous at the boundary between each pair of contiguous income ranges,

$$(4) \qquad f(x) = \begin{cases} \dfrac{-p_i \lambda_i e^{-\lambda_i x}}{e^{-\lambda_i \ell_i} - e^{-\lambda_i \ell_{i-1}}} \ \forall \ x \in [\ell_{i-1}, \ell_i),\ i \le 5 \\[2ex] p_6 \lambda_6 e^{-\lambda_6 (x-\ell_5)} \ \forall \ x \in [\ell_5, \infty) \end{cases}$$

with p_i = 0.20, i = 1…4, p_5 = 0.15, and p_6 = 0.05, and the Lagrange multipliers for the mean constraints satisfy

$$(5) \qquad e^{\lambda_i(\ell_i - \ell_{i-1})} - \left[\frac{1+\lambda_i(\ell_i - \mu_i)}{1-\lambda_i(\mu_i - \ell_{i-1})}\right] = 0,\ i \le 5,$$

$$\lambda_6 = 1/(\mu_6 - \ell_5).$$

For 1997, this density is depicted in figure 1 by the series of piecewise exponential curves marked with solid black circles.

The second density is a parametric, truncated three-parameter lognormal density. This density is smooth everywhere and has a general shape that is similar to the piecewise uniform and maximum entropy densities, but does not satisfy either the probability or mean conditions exactly in any range of income. Suppose that $z = [\ln(x-\theta) - \mu]/\sigma$ has a standard normal distribution, with α, σ, and θ parameters and $x > \theta$. Define the standardized zero income limit by $z_0 = (\ln(-\theta) - \mu)/\sigma$ and denote the standard normal *cdf* at z_0 by $\Phi(z_0) = \int_{-\infty}^{z_0} \varphi(z)dz$, where $\varphi(z) = (1/\sqrt{2\pi})e^{-z^2/2}$ is the standard normal *pdf*. Then the truncated three-parameter log-normal density for $x \ge 0$ is defined by

$$(6) \qquad f(x \mid x \ge 0; \mu, \sigma, \theta) = \frac{1}{\sqrt{2\pi}\sigma(x-\theta)(1-\Phi(z_0))}$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}[\ln(x-\theta) - \mu]^2\right\}.$$

For 1997, this density is depicted in figure 1 by the smooth curve with empty circles.

Data for U.S. food consumption and retail prices, as well as additional variables that are described in the next section, have been obtained from LaFrance (1999a) for the years 1918–1995. However, observations on the Census Bureau's summary data for the income distribution are available for 1929, 1935/36, 1941 and 1946–98. One issue that arises in using this data in an aggregate U.S. food demand model, then, centers on predicting or extrapolating this income data for the years 1918–1928, 1930–40, 1942–43, and 1945. We forecast these missing observations utilizing data on per capita disposable personal income and the unemployment rate as predictors and following a recursive forecasting procedure. The natural logarithms of the first quintile upper limit and conditional mean income are predicted with a constant term, the log of per capita disposable income, the squared log of per capita disposable income, and the unemployment rate. Each successive income limit and condition mean are then recursively predicted with ordinary least squares using a constant term and first- second- and third-order powers of the log of the closest smaller limit or conditional mean, as appropriate, as regressors.

4. Estimating a Nested QPIGL-IDS for Food Demand

The system of empirical nested QPIGL-IDS demand equations that we estimate for U.S. food consumption for the years 1918–1995, excluding 1942–1946, can be written in deflated expenditure form as

$$(7) \qquad \boldsymbol{e}_t = m_t^{1-\kappa} \boldsymbol{P}_t^{\lambda} \Big\{ \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{p}_t(\lambda)$$

$$+ \boldsymbol{\gamma}\Big[m_t(\kappa) - \boldsymbol{p}(\lambda)'\boldsymbol{A}\boldsymbol{s}_t - \tfrac{1}{2}\boldsymbol{p}_t(\lambda)'\boldsymbol{B}\boldsymbol{p}_t(\lambda)\Big]$$

$$+ [\boldsymbol{I} + \boldsymbol{\gamma}'\boldsymbol{p}_t(\lambda)]\boldsymbol{\delta}\Big[m_t(\kappa) - \boldsymbol{p}_t(\lambda)'\boldsymbol{A}\boldsymbol{s}_t - \tfrac{1}{2}\boldsymbol{p}_t(\lambda)'\boldsymbol{B}\boldsymbol{p}_t(\lambda)\Big]^2 \Big\}$$

$$+ \boldsymbol{\varepsilon}_t,\ t = 1, \ldots, T,$$

where $\boldsymbol{e}_t = [p_{1t}q_{1t} \ \ldots \ p_{nt}q_{nt}]'$ is the n-vector of deflated per family annual expenditures on individual food items, $\boldsymbol{s}_t$ is a vector that includes a constant, the mean, variance and skewness of the U.S. population's age distribution, the proportion of the U.S. population that is Black and the proportion of the population that is neither Black nor White, and $\boldsymbol{\varepsilon}_t$ is an n-vector of mean zero, identically distributed error terms. We specify the empirical model in expenditure form to keep all income terms on the right-hand side so that the mean values of all of the appropriate transformations of income are properly calculated across all U.S. families during the econometric estimation of the demand parameters.

Estimation of the model's parameters requires, for given κ ∈ (0, 1], numerical integration to evaluate the expected values of the three powers of income at each year in the sample period, where the expectation is taken over that year's estimated income distribution. To accomplish this, we transform the positive half line into the unit interval [0, 1) through a change of variables to $y = 10^{-4}x/(1+10^{-4}|x|)$ and use Simpson's rule on a grid over the unit interval.

We used two-step nonlinear seemingly unrelated regressions equations (NLSURE) estimation methods, combined with a one dimensional search over the income term's Box-Cox parameter κ. Only one iteration on the residual covariance matrix was calculated to avoid numerically over fitting one or more equations, which can occur with iterative NLSURE in large, highly parameterized demand models such as this.[6] A search over κ was used to incorporate the numerical integrations required to generate the aggregate income variables, which in turn depend upon the parameter κ. Symmetry of the coefficient matrix ***B*** is maintained throughout the estimation process in order to reduce the dimension of the parameter space from 527 to 317 estimated parameters. The optimal first round value for κ was found to be 1.00 for the truncated three-parameter lognormal (T3PLN) density, 1.03 for the piecewise exponential, and 0.97 for the piecewise uniform income distribution. Conversely, the optimal values for κ obtained in the second iteration of the NLSURE procedure are 1.03, 1.00, and 0.98 for the T3PLN, piecewise exponential, and piecewise uniform income distributions, respectively.

QAIDS-IDS is strongly rejected in favor of an extended QES-IDS for this data set, for both income distribution estimates, and at both stages of the NLSURE estimation process. The resulting estimates for the first- and second-order income coefficients, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, respectively, as well as the optimal values for the Box-Cox parameters, κ and λ, are statistically similar across specifications of the income distribution.

Table 1 presents the individual equation summary statistics for the T3PLN income distribution. Results for the other income distribution functional forms were similar, and are not reported here.

Table 1. Equation Summary Statistics, T3PLN.

Equation	$\mathbf{R^2}$	**Durbin-Watson**
Milk & Cream	.9975	1.935
Butter	.9971	1.559
Cheese	.9977	1.353
Frozen Dairy	.9661	1.333
Canned & Powder Milk	.9648	1.287
Beef & Veal	.9741	1.280
Pork	.9266	1.315
Other Red Meat	.9569	1.455
Fish	.9899	1.665
Poultry	.9628	1.098
Fresh Citrus Fruit	.8474	2.084
Fresh Noncitrus Fruit	.9668	2.628
Fresh Vegetables	.9834	1.790
Potatoes	.9671	1.869
Processed Fruit	.9869	1.829
Processed Vegetables	.9785	1.554
Eggs	.9747	1.771
Fats and Oils	.9983	1.551
Cereals and Bakery	.9925	1.279
Sugar	.9828	2.145
Coffee, Tea & Cocoa	.9691	1.930

Table 2 presents the Box-Cox price coefficient and the first- and second-order income coefficients for the T3PLN income distributions. The standard errors reported in this table are conditional on the estimate of κ due to the generated income variables nature of the demand model's parameter estimates. This implies that these standard errors should be interpreted with caution.

Table 2. Income Coefficients, T3PLN: $\hat{\kappa}$ = 1.03

Parameter	**Estimate**	**Conditional Standard Error**
λ	.853915	.033923
γ_1	-.024642	.013093
γ_2	$-.258821\cdot10^{-2}$	$.206153\cdot10^{-2}$
γ_3	$-.911698\cdot10^{-3}$	$.221715\cdot10^{-2}$
γ_4	.017188	$.500290\cdot10^{-2}$
γ_5	$.301518\cdot10^{-2}$	$.592252\cdot10^{-2}$
γ_6	.022654	.010523
γ_7	.010451	$.995331\cdot10^{-2}$
γ_8	$-.313399\cdot10^{-2}$	$.371416\cdot10^{-2}$
γ_9	$.339135\cdot10^{-2}$	$.209848\cdot10^{-2}$
γ_{10}	$-.93449810^{-3}$	$.425920\cdot10^{-2}$
γ_{11}	$-.968060\cdot10^{-3}$	$.997693\cdot10^{-2}$
γ_{12}	.048544	.015135
γ_{13}	$.768783\cdot10^{-2}$	$.772512\cdot10^{-2}$
γ_{14}	-.020795	.020258
γ_{15}	$-.550843\cdot10^{-2}$	$.685728\cdot10^{-2}$
γ_{16}	.026156	$.743900\cdot10^{-2}$
γ_{17}	.011163	$.439605\cdot10^{-2}$
γ_{18}	$.615151\cdot10^{-2}$	$.353595\cdot10^{-2}$
γ_{19}	.021258	.013952
γ_{20}	.019811	$.939334\cdot10^{-2}$
γ_{21}	$.254267\cdot10^{-2}$	$.252764\cdot10^{-2}$
δ_1	$.112092\cdot10^{-5}$	$.486111\cdot10^{-6}$
δ_2	$-.187811\cdot10^{-7}$	$.862305\cdot10^{-7}$
δ_3	$.866507\cdot10^{-7}$	$.802000\cdot10^{-7}$
δ_4	$-.463795\cdot10^{-6}$	$.203328\cdot10^{-6}$
δ_5	$-.726773\cdot10^{-9}$	$.226065\cdot10^{-6}$
δ_6	$-.518846\cdot10^{-6}$	$.405252\cdot10^{-6}$
δ_7	$-.316928\cdot10^{-6}$	$.401569\cdot10^{-6}$
δ_8	$.137460\cdot10^{-6}$	$.136176\cdot10^{-6}$
δ_9	$.720693\cdot10^{-8}$	$.714625\cdot10^{-7}$
δ_{10}	$.250661\cdot10^{-6}$	$.199447\cdot10^{-6}$
δ_{11}	$.650096\cdot10^{-7}$	$.391085\cdot10^{-6}$
δ_{12}	$-.193640\cdot10^{-5}$	$.670224\cdot10^{-6}$
δ_{13}	$.496807\cdot10^{-8}$	$.294807\cdot10^{-6}$
δ_{14}	$.959843\cdot10^{-6}$	$.731403\cdot10^{-6}$
δ_{15}	$.510022\cdot10^{-6}$	$.285776\cdot10^{-6}$
δ_{16}	$-.445306\cdot10^{-6}$	$.328907\cdot10^{-6}$
δ_{17}	$-.366017\cdot10^{-6}$	$.187846\cdot10^{-6}$
δ_{18}	$-.159808\cdot10^{-6}$	$.174505\cdot10^{-6}$
δ_{19}	$-.525896\cdot10^{-6}$	$.569711\cdot10^{-6}$
δ_{20}	$-.288873\cdot10^{-6}$	$.386350\cdot10^{-6}$
δ_{21}	$-.138823\cdot10^{-7}$	$.982542\cdot10^{-7}$

However, it is possible to calculate consistent test statistics for the rank of the demand model using a Wald test. For the T3PLN version, we obtain the following:

H_0: $\boldsymbol{\gamma} = \mathbf{0}$
H_1: $\boldsymbol{\gamma} \neq \mathbf{0}$ $\qquad \chi^2(21) = 114.89$

H_0: $\boldsymbol{\delta} = \mathbf{0}$
H_1: $\boldsymbol{\delta} \neq \mathbf{0}$ $\chi^2(21) = 59.99$
H_0: $\boldsymbol{\gamma} = \boldsymbol{\delta} = \mathbf{0}$
H_1: $\boldsymbol{\gamma} \neq \mathbf{0}$ or $\boldsymbol{\delta} \neq 0$ $\chi^2(42) = 349.18$

Similar results were obtained for the other two distributions, and in all cases we are lead to reject all three versions of the null hypothesis at any standard level of significance, and therefore conclude that the full rank three QES-IDS model is a significant improvement over all of the more restrictive versions. We also conclude that any version of integrable AIDS model is significantly inferior to the corresponding alternative with the Box-Cox income parameter statistically very close to unity.

5. Conclusions

This paper presents a method to nest, test and estimate both the rank and functional form of the income terms in an incomplete system of aggregable and integrable demand equations is derived. Bayesian methods are applied to the problem of inferring the U.S. income distribution using annual time series data on quintile and top five percentile income ranges and intra-quintile and top five percentile mean incomes. The results obtained with different functional forms for the income distribution are compared and contrasted. The estimates for the year-to-year income distribution are combined with annual time series data on the U.S. consumption of and retail prices for twenty-one food items over the period 1919–95, excluding 1942–46 to account for the structural impacts of World War II.

The empirical results suggest that all integrable versions of the AIDS model are strongly rejected by this data set, in favor of a full rank three extended QES-IDS. This has potentially significant implications for future demand analysis, particularly with respect to food consumption using aggregate market-level data sets. For example, in his model of the demand for dairy products, Agnew (1998) finds the nonhomothetic, integrable rank two AIDS model to be substantially responsible for rejections of the implications of consumer choice theory – both symmetry and curvature – as well as a similar result as is reported here regarding the inferiority in all statistical respects relative to an extended LES model specification. The extreme level of confidence with which we reject the AIDS forms here suggests that a similar finding is likely. This, of course, must be left for future research.

The empirical results presented in this paper regarding the demand for U.S. food consumption are somewhat limited in their scope and interpretation. The primary reason for this is the fact that all other parameter estimates are conditional on the estimated Box-Cox parameter for the income coefficient. On the other hand, however, if we were to assume *a priori* that a QES model is the best specification – which of course at this stage of the game is unfair play – then we could interpret the remaining parameter estimates in the usual manner. It is interesting to note that, given the QES specification, the moments required from the income distribution for exact aggregation are precisely the mean and the variance. This is an interesting implication of the present study in its own right. No attempt is made in the present empirical work to test or impose the appropriate curvature restrictions necessary for the demand model to be logically consistent with *weak integrability*, and therefore the maximization hypothesis. The empirical results reported here, as a result, should not be use for welfare analysis.

References

Agnew, G. K. *Linquad* Unpublished M.S. Thesis, Department of Agricultural and Resource Economics, University of Arizona, Tucson, 1998.

Anderson, G. and R. Blundell. "Testing Restrictions in a Flexible Dynamic Demand System: An Application to Consumer's expenditure in Canada." *Review of Economic Studies* 50 (1983): 397-410.

Browning, M. and C. Meghir. "The Effects of Male and Female Labor Supply on Commodity Demands." *Econometrica* 59 (1991): 925-951.

Buse, A. "Testing Homogeneity in the Linearized Almost Ideal Demand System." *American Journal of Agricultural Economics* 80 (1998): 208-220.

Deaton, A. and Muellbauer, J. "An Almost Ideal Demand system." *American Economic Review*, 70 (1980): 312-326.

Epstein, L. "Integrability of Incomplete Systems of Demand Functions." *Review of Economic Studies* 49 (1982): 411-425.

Gorman, W. M. "Consumer Budgets and Price Indices." Unpublished typescript, 1965. Published as Chapter 5 in Blackorby, C. and T. Shorrocks, Eds. *Separability and Aggregation: collected Works of W. M. Gorman, Volume I* Oxford: Clarendon Press, 1995: 61-88.

_________. "Some Engel Curves." In A. Deaton, ed. *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone,* Cambridge: Cambridge University Press, 1981. Republished as Chapter 20 in Blackorby, C. and T. Shorrocks, Eds. *Separability and Aggregation: collected Works of W. M. Gorman, Volume I* Oxford: Clarendon Press, 1995: 351-376.

Howe, H., R. A. Pollak, and T. J. Wales. "Theory and Time Series Estimation of the Quadratic Expenditure System." *Econometrica* 47 (1979): 1231-1247.

Zellner, A. "Optimal Information Processing and Bayes Theorem." *American Statistician* (1988): 278-84.

Endnotes

[1] "Extended QES"indicates that *supernumerary income* is income minus a quadratic form in prices and that there is an $n \times n$ matrix of price effects in addition to the intercepts in the QES demands.

[2] See LaFrance (1999a, 1999b) for empirical evidence for the exclusion of World War II and the stability of U.S. food demands over this long sample period. The twenty-one food items included in the data set can be conveniently grouped into four categories: (1) *dairy products*, including fresh milk and cream, butter, cheese, ice cream and frozen yogurt, and canned and dried milk; (2) *meats, fish and poultry*, including beef and veal, pork, other red meat, fish, and poultry; (3) *fruits and vegetables*, including fresh citrus fruit, fresh non-citrus fruit, fresh vegetables, potatoes and sweet potatoes, processed fruit, and processed vegetables; and (4) *miscellaneous foods*, including fats and oils excluding butter, eggs, cereals, sugar and sweeteners, and coffee, tea and cocoa.

[3] However, see Browning and Meghir (1991) for an application of estimating the integrable AIDS, using the LA-AIDS with a symmetric matrix of log-price coefficients to obtain starting values for the nonlinear estimation procedure.

[4] The mean for the 80-95 percentile range is calculated as $\mu_{80-95\%} = (.20\mu_{80-100\%} - .05\mu_{95-100\%})/.15$. The 85-95 percentile range is the interval from the lower limit of the top quintile to the lower limit of the top five percentile range, while the top five percentile mean is assumed to be the midpoint of that range for the piecewise uniform densities discussed in this subsection.

[5] For simplicity, the intra-range mean of the top five percentile group is assumed to be located at the center of that range, making the top percentile uniform density continuous up to the point $x_{.95} + 2\mu_{.95}$, which reduces the number of discontinuities from twelve to eleven.

[6] See LaFrance (1999b), footnote 12 for a discussion of this issue. The crux of the matter is that all of the model parameters, which in the present case total 317, enter each of the demand equations, while there are only 76 time series observations. This creates a numerical possibility for a singular estimated covariance matrix when iterative NLSURE is employed, which generates an unbounded likelihood function.