

COMPLETENESS, DISTRIBUTION RESTRICTIONS AND
THE FORM OF AGGREGATE FUNCTIONS

by

Thomas M. Stoker

August 1982

(revised: August 1983)

WP #1345-82

*Thomas M. Stoker is Assistant Professor of Economics and Management at the Sloan School of Management, Massachusetts Institute of Technology. The author wishes to thank Julio Rotemberg and Dale Jorgenson for continuing discussion of this topic, as well as Takeshi Amemiya, William Barnett, Angus Deaton, Terence Gorman, Zvi Griliches, Jerry Hausman, James Heckman, Paul Krugman, Ed Kuh, Ed Leamer, Dan McFadden, Peter Phillips, Robert Pindyck, Jim Powell and Richard Schmalensee for useful comments on many aspects of this research.

ABSTRACT

This paper presents the problem of aggregation over individual agents as a basic identification problem inherent to interpreting relationships between averaged economic variables. The concept of a complete aggregation structure is introduced, which embodies the correct condition for identification. Several examples of complete aggregation structures are provided by previous work on the aggregation problem in economics. Examples are also provided by work on complete distribution families in statistics, which in turn provide the correct conceptual framework for developing a theory of parameter estimation and tests of specific aggregation assumptions. The potential lack of correspondence between microeconomic behavior and estimated relations between averaged data is illustrated by distribution families obeying linear probability movement, which induce an extreme failure of the completeness property. Certain topics regarding empirical applications of aggregation results are reviewed.

1. Introduction

The problem of aggregation over individual agents concerns the connection between micro behavior and macro behavior. Micro behavior refers to the relation between dependent and predictor variables for individual agents, while macro behavior refers to the relation between aggregate (averaged or totaled) dependent and predictor variables. Specific solutions to the aggregation problem require a specification of micro behavior together with a specification of the distribution of predictor variables across agents from which the aggregate function, the relation between aggregate variables, is derived. The purpose of this paper is to analyze the problem of aggregation over individual agents, with special emphasis on nonlinear micro behavior and the role of the distribution of predictor variables across agents.¹

The analysis is concerned with the empirical practice of treating parameters estimated with aggregate data as though they were behavioral parameters of individuals. We provide the required theoretical foundation for this practice, termed complete aggregation structure, which implies a unique correspondence between micro functions (representing individual behavior) and aggregate functions. The foremost examples of previous approaches providing this property are those of linear aggregation (c.f. Gorman (1953), Muellbauer (1975, 1976), and Lau (1977, 1982) as well as Theil (1954, 1971, 1975) and Barnett (1979, 1981)). These approaches are useful because structure can be applied to aggregate function coefficients, which is provided by the theory of individual behavior.

The first major advantage of the concept of complete aggregation structure is that it provides a general framework for incorporating nonlinear microeconomic behavior and specific restrictions on the distribution of predictor variables. The "aggregation problem" is presented in a familiar guise, namely as a problem of identification. An aggregation structure is a structural model and an aggregate function is analogous to a reduced form equation, with completeness the precise condition under which micro behavior can be identified from an aggregate

function. Moreover, the advantages of linear aggregation approaches can be stated precisely, as providing the major "distribution free" complete aggregation structure.

The second major advantage of the concept of complete aggregation structure is that it contains the concept of complete distribution families central to the theory of hypothesis testing and unbiased estimation of statistics. In addition to providing examples of complete aggregation structures with general nonlinear microeconomic behavior, this correspondence provides the correct conceptual framework for developing a unified theory of parameter estimation and tests of specific aggregation assumptions, such as those of linear aggregation. Although the general estimation theory will be presented as part of future research, several existing results are reviewed and discussed here.

The exposition begins with the definitions of basic concepts in Section 2.1, with interpretive remarks and examples in Section 2.2. Section 3 presents some results characterizing completeness, and then considers several specific aggregation assumptions for illustration -- linear aggregation, distributions of the exponential family form, "fixed" distributions and homogeneous micro functions, and discrete and segmented population distributions. Section 4 begins with a numerical example to illustrate the completeness property, discusses the failure of completeness induced by a linear probability movement structure, and reviews some results on empirical techniques. Section 5 contains some concluding remarks.²

2. Notation and Definitions of Basic Concepts

2.1 Aggregation Structures and Completeness

Each individual agent is characterized by an M vector of predictor variables X , which determine a dependent variable Y via a micro function

$$Y = F_{\gamma}(X) \quad (2.1)$$

where γ is a Q vector of structural parameters $\gamma \in \Gamma \subset \mathbb{R}^Q$, with Γ the set of all

parameter values. The class of possible micro functions is denoted as $C = \{F_Y(X)\}_{Y \in \Gamma}$.

The population of agents is assumed to be a random sample from a distribution with density $p(X|\theta)$, where θ is an L vector of parameters, $\theta \in \Theta \subset \mathbb{R}^L$, with Θ the set of all parameter values. The domain of definition of $p(X|\theta)$ is Ω , a full dimensional³ subset of \mathbb{R}^M . The class of all possible densities is denoted $P = \{p(X|\theta)\}_{\theta \in \Theta}$. We require

ASSUMPTION 1: The expectation $E(Y) = E(F_Y(X))$ exists and is finite for all $Y \in \Gamma$ and $\theta \in \Theta$.

We now define the concepts required for considering aggregation over individuals.⁴

DEFINITION 1: A is an aggregation structure if $A = \{C, P\}$, where $C = \{F_Y(X)\}_{Y \in \Gamma}$ is a class of micro functions and $P = \{p(X|\theta)\}_{\theta \in \Theta}$ is a class of densities of X .

DEFINITION 2: Let $A = \{C, P\}$ be an aggregation structure. For each $F_Y(X) \in C$, the aggregate function derived from $F_Y(X)$ is defined as:

$$E_\theta(y) = \int F_Y(X) p(X|\theta) dX \equiv \phi_Y(\theta) \quad (2.2)$$

DEFINITION 3: Let $A = \{C, P\}$ be an aggregation structure with $C = \{F_Y(X)\}_{Y \in \Gamma}$ and $P = \{p(X|\theta)\}_{\theta \in \Theta}$. P is complete for C if and only if for all $F_Y(X), F_{Y'}(X) \in C$ such that

$$\phi_Y(\theta) = \int_\Omega F_Y(X) p(X|\theta) dX = \int_\Omega F_{Y'}(X) p(X|\theta) dX = \phi_{Y'}(\theta) \quad (2.3)$$

for all $\theta \in \Theta$, then

$$F_Y(X) = F_{Y'}(X) \text{ almost surely (p):} \quad (2.4)$$

We say that $A = \{C, P\}$ is complete if P is complete for C .⁵

Completeness says that a particular aggregate function $\phi_Y(\theta)$ is associated with a unique form of micro function $F_Y(X)$. Completeness fails if several different micro functions imply the same aggregate function. In other words, if the aggregation structure $A = \{C, P\}$ is regarded as a structural model, an aggregate

function is analogous to a reduced form equation, Completeness is the precise condition under which $E(Y) = \phi_Y(\theta)$ identifies $F_Y(X)$, the micro function which implied it.⁶

2.2 Interpretative Remarks and Examples

2.2a Micro Functions

The micro function $Y = F_Y(X)$ is defined as common to all agents, which is not restrictive because X captures all agent differences relevant to Y . For interpretation, however, it is useful to consider situations where X represents the observed predictor variables in an econometric model.

Suppose for each agent that there is an observed dependent variable y which is given by a behavioral model as

$$y = f_Y(X, u) \quad (2.5)$$

where u is an unobserved stochastic disturbance distributed conditional on X with respect to the density $g(u|X)$, where we take $E(u|X) = 0$ without loss of generality.⁷ The important feature of $g(u|X)$ is that it is invariant to changes in θ , the X density parameters. Our concept of micro function is defined as:

$$Y = E(f_Y(X, u) | X) = \int f_Y(X, u) g(u|X) du | X \equiv F_Y(X) \quad (2.6)$$

Under this interpretation, $F_Y(X)$ represents a subaggregated version of the behavioral model (2.5). The aggregate functions implied by (2.5) and (2.6) are the same; $E(y) = E(Y) = \phi_Y(\theta)$, which under completeness identifies $F_Y(X)$. Finally, if $g(u|X)$ depends upon some unknown parameters, they can easily be included into the definition of γ .

We now present some examples.

EXAMPLE 2.1: Linear Micro Functions. The class of linear micro functions is defined as $C_L = \{F_Y(X)\}_{\gamma \in \Gamma}$ where $\gamma = (\gamma_0, \gamma_1')$, $\Gamma = R^{M+1}$ and $F_Y(X) = \gamma_0 + X'\gamma_1$. This class is implied by a standard linear behavioral model with additive error⁸

$$y = f_Y(X, u) = (\gamma_0 + u) + X'\gamma_1 \quad (2.7a)$$

or by a linear random coefficients model.⁹

$$y = f_{\gamma}(X,u) = (\gamma_0 + u_0) + X'(\gamma_1 + u_1) \quad (2.7b)$$

EXAMPLE 2.2: Quadratic Micro Functions. The simplest examples of nonlinear micro functions are polynomial forms. For instance, we can define $C_Q = \{F_{\gamma}(X)\}_{\gamma \in \Gamma}$ as the class of quadratic micro functions via

$$F_{\gamma}(X) = \gamma_0 + \sum_i \gamma_{1i} X_i + \sum_{ij} \gamma_{2ij} X_i X_j \quad (2.8)$$

where $\gamma = (\gamma_0, (\gamma_{1i}), (\gamma_{2ij})) \in \Gamma = R^{1+M+(M+1)M/2}$.

EXAMPLE 2.3: Probit Micro Functions. The class $C_{PB} = \{F_{\gamma}(X)\}_{\gamma \in \Gamma}$ of (homogeneous) probit micro functions is defined via $\gamma = (\gamma_0, \gamma_1') \in \Gamma = R^{M+1}$ and

$$F_{\gamma}(X) = \Phi(\gamma_0 + X'\gamma_1) \quad (2.9)$$

where Φ is the standard normal distribution function. This class is implied by a discrete choice behavioral model (c.f. Amemiya (1981), Manski and McFadden (1981)) of the form

$$y = f_{\gamma}(X,u) = \begin{cases} 1 & \text{if } u \leq \gamma_0 + X'\gamma_1 \\ 0 & \text{if } u > \gamma_0 + X'\gamma_1 \end{cases} \quad (2.10)$$

where $g(u|X)$ is normal with mean zero and variance 1. Other discrete choice classes (logit, etc.) can be defined by adopting different assumptions on $g(u|X)$.

Other examples are given in Sections 3.3 and 3.4.

In general, the specification of the class C captures all theoretical restrictions and/or assumptions imposed on micro functions. When no restrictions are appropriate, we specify C as C_U , the class of all functions measurable with respect to $p(X|\theta)$ obeying Assumption 1.

2.2b Density Classes and Aggregate Functions

The density form $p(X|\theta)$ is usefully interpreted as representing how the X density changes through time. For each time period $t \in \tau$ (either discrete or continuous), θ has value $\theta(t)$ and $p(X|\theta(t))$ is the density of the true X distribution. The aggregate function $E_t(Y) = \phi_Y(\theta(t))$ gives the time path of the dependent variable mean, when $F_Y(X)$ is the true micro function. To consider data, suppose that for each t , \bar{Y}_t and \bar{X}_t are consistent estimators of $E_t(Y)$ and $\theta(t)$ (and ϕ_Y is continuous), then $\bar{Y}_t \approx \phi_Y(\bar{X}_t)$ is the true macro model, with equality holding up to sampling variation. The completeness property says that knowledge of the true ϕ_Y suffices to identify the true micro function F_Y .

The specification of the class P captures all restrictions placed on the X densities. At times we will define density classes using specific functional forms, as in

EXAMPLE 2.4: Normal Density Classes. Let $\theta = (\mu_1, \dots, \mu_M, \sigma_{11}, \sigma_{12}, \sigma_{22}, \dots, \sigma_{MM})$ be the $L = M + (M+1)M/2$ vector of means and covariances of X, by defining the normal class $P_N = \{p(X|\theta)\}_{\theta \in \Omega}$ as

$$p(X|\theta) = \frac{1}{(2\pi)^{M/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (X-\mu)' \Sigma^{-1} (X-\mu) \right) \quad (2.11)$$

where $\mu = (\mu_1, \dots, \mu_M)'$, $\Sigma = [(\sigma_{ij})]$, $\theta = \{\theta | \Sigma \text{ is positive definite}\}$ and $\Omega = \mathbb{R}^M$.

The normal translation class is given as $P_{NT} = \{p \in P_N | \Sigma = \Sigma_0\}$, the class of normal distributions with fixed covariance matrix Σ_0 . Similarly $P_{NV} = \{p \in P_N | \mu = \mu_0\}$ is the class of normal distributions with fixed mean vector.

We will also consider density classes which just obey general properties (such as "mean full dimensionality" of Section 3.2). Our framework does assume that a class P is defined using a common functional form $p(X|\theta)$ -- this is not too restrictive since θ can be a vector of any finite dimension L.

In practice, the specification of $p(X|\theta)$ should represent all X density changes relevant to the particular application. If a specified functional form for $p(X|\theta)$ is appropriate, θ is naturally taken as the parameters of the assumed form. If a particular aggregate function is of interest, say $\bar{Y}_t = \phi_Y(\bar{X}_t)$, then θ should include the probability limit of \bar{X}_t (e.g. if \bar{X}_t is the sample average of X , θ should include $E(X)$).¹⁰

3. Completeness and Aggregation

3.1 Discussion and Basic Properties

As indicated above, completeness is the basic identification condition of aggregation structures. We should note, however, that our definition isolates an aggregation and not the identification of the structural parameters γ of the function $Y = F_Y(X)$. If the defining condition (2.4) were replaced by

$$\gamma = \gamma' \tag{3.1}$$

then completeness would require that $\phi_Y(\theta)$ identifies $F_Y(X)$ and that $F_Y(X)$ identifies γ . Condition (3.1) is clearly sufficient for (2.4), and will often be used to establish completeness in the examples which follow.

Our definition of completeness contains as a special case the definition of completeness familiar to the statistical theory of unbiased estimation and hypothesis testing. In this context a class of densities P is complete if P is complete for the class C_U of all measurable functions. P is boundedly complete if P is complete for the class of all bounded measurable functions. The usual definition of completeness in statistics replaces our defining conditions (2.3) and (2.4) with -- "If $\int F_Y(X)p(X|\theta)dX = 0$ for all $\theta \in \theta$, then $F_Y(X) = 0$ almost surely (p)." This condition is obviously equivalent to (2.3) and (2.4) if the class of functions C is closed under addition and subtraction.

In the theory of unbiased estimation and hypothesis testing, the class $P = \{p(X|\theta)\}_{\theta \in \theta}$ is regarded as the set of possible densities for a problem of statistical inference, with θ a parameter to be estimated. X is regarded as

a statistic, with completeness of P implying that if $E_{\theta}(h(X)) = \phi(\theta)$, for any ϕ and for all $\theta \in \Theta$, then the function h is unique. If X is a sufficient statistic for θ , and P is boundedly complete, then X is minimal sufficient for θ .

Completeness also provides many other useful properties concerning the existence and construction of uniformly most powerful tests on the value of θ using the statistic X (c.f. Lehmann (1959) and Ferguson (1967), among others).

The following properties follow immediately from the definition of completeness.

THEOREM 1:

- (a) Let C and C' denote classes of micro functions with $C \subset C'$, and P and P' denote classes of densities with $P \subset P'$ where sets of measure zero under P have measure zero under P' . If P is complete for C' , then P is complete for C and P' is complete for C and C' . If P' is not complete for C , then P' is not complete for C' , and P is not complete for C or C' .
- (b) Let $P = \{p(X|\theta)\}_{\theta \in \Theta}$ and $r: \theta \rightarrow \theta^*$, $r(\theta) = \theta^*$, be a (reparameterizing) function which is one-to-one and onto. The completeness of a class P for C is invariant to whether P is parameterized by θ or θ^* , i.e. $P = \{p(X|\theta)\}_{\theta \in \Theta} = \{p(X|r^{-1}(\theta^*))\}_{\theta^* \in \Theta^*}$.

Property (a) says heuristically that completeness will fail in cases where the admissible function class C is "too large," or in cases where the admissible density class P is "too small." Property (b) says that completeness is a non-parameteric property given the functional form of $p(X|\theta)$ for some definition of θ . Thus while the precise form of the aggregate functions do depend on the parameterizations chosen ($\phi_Y(\theta) = \phi_Y(r^{-1}(\theta^*)) = \phi_Y^*(\theta^*)$), the completeness property does not. In the development and examples to follow, we will choose the parameterization of θ which is most convenient for the question at hand.

3.2 Linear Aggregation

Here we consider aggregation when the class of micro functions is $C_L = \{F_\gamma(X)\}_{\gamma \in \Gamma}$ where $F_\gamma(X) = \gamma_0 + X'\gamma_1$. Given a class of densities $P = \{p(X|\theta)\}_{\theta \in \Theta}$ where the dimension L of θ is greater than M , a useful reparameterization of P is constructed as follows. Denote the means of X as functions of θ by $E_\theta(X) = \mu(\theta)$, where the range of μ is denoted $\mu(\Theta) \subset R^M$. Suppose we consider reparameterizing $p(X|\theta)$ by $\theta^* = (\mu, \theta_1)$, with θ_1 an $L-M$ vector, via a mapping r defined as:

$$\theta^* = r(\theta) = (\mu(\theta), \theta_1(\theta)) \quad (3.2)$$

We provide for the existence of such a mapping r via the following definition.

DEFINITION 4: A class of distributions $P = \{p(X|\theta)\}_{\theta \in \Theta}$ is mean full dimensional if there exists a (reparameterizing) function r of the form (3.2) which is one-to-one and onto its range, where $\mu(\Theta)$ is a full dimensional subset of R^M .

Given a micro function $Y = F_\gamma(X) = \gamma_0 + X'\gamma_1$, the aggregate function is $E_\theta(Y) = \gamma_0 + [\mu(\theta)]'\gamma_1$. The advantages of micro linearity in aggregation can now be stated precisely as Theorem 2 (which is proved in the Appendix).

THEOREM 2:

- (a) Let P denote any mean full dimensional class of densities. Then P is complete for C_L , the class of linear micro functions.
- (b) Let C denote a class of functions which strictly contains C_L -- i.e. $C_L \subset C$ but there exists $G(X) \in C$, $G(X) \notin C_L$. Then there exists a mean full dimensional class of densities P such that P is not complete for C .

In words, Theorem 2 says that completeness is assured for a general (mean full dimensional) class of densities if micro functions are linear, and if the class C of interest contains both linear and nonlinear micro functions, then the density class P must be restricted to insure completeness.¹¹ In this sense,

linear micro functions provide the major "distribution free" complete aggregation structures.¹²

In many modeling situations, linearity of microeconomic behavior is quite unrealistic, such as in typical discrete choice situations (see Example 2.3). Consequently, it is of interest to consider situations for which completeness is assured for a class C containing nonlinear micro functions. We now turn to several such examples.

3.3 Distribution Classes of the Exponential Family

The connection between completeness as defined here and completeness as defined in statistics is quite useful for considering aggregation over nonlinear micro functions, since an aggregation structure $A = \{C, P\}$ is guaranteed to be complete if P is any complete class of densities known to statistics. The foremost examples of complete density classes are those of the exponential family form (c.f. Ferguson (1967), Efron (1975, 1978) and Barndorff-Neilsen (1978)).

DEFINITION 5: $P = \{p(X|\theta)\}_{\theta \in \Theta}$ is a class in exponential family form if $p(X|\theta)$ can be written for all $\theta \in \Theta$ as

$$p(X|\theta) = c(\theta)h(X) \exp(\pi(\theta)'D(X)) \quad (3.3)$$

where $\pi: \Theta \rightarrow R^L$ is an L vector function of θ , $D: \Omega \rightarrow R^L$ is a one-to-one L vector function of X and

$$c(\theta) = \left(\int_{\Omega} h(X) \exp(\pi(\theta)'D(X)) dX \right)^{-1}$$

We restrict attention to cases where the range of D is a full dimensional subset of R^L and $c(\theta)$ exists. The following result is well known.

THEOREM 3 (Lehmann and Scheffe (1950-1955), Lehmann (1959)):

If $P = \{p(X|\theta)\}_{\theta \in \Theta}$ is a class of the exponential family form, and the range of π contains an open set in R^L , then P is complete (for the class of all measurable micro functions).

From this point forward, we assume in each case that the function π is one-to-one and onto its range, which contains an open subset of R^L . For later reference we will consider two additional parameterizations of the form (3.3), the "natural parameterization" found by reparameterizing θ by π via the function $\pi(\theta)$, and the "mean parameterization" found by reparameterizing π by $\mu_D = E_{\pi}(D(X))$. The validity of the latter parameterization follows from the development of Efron (1978) or Stoker (1982), as well as the fact that the range of the $\mu_D = r(\pi) = E_{\pi}(D(X))$ function contains an open subset of R^L .

The exponential family form contains many common distribution forms as special cases -- the normal class P_N , gamma and beta, among others (c.f. Ferguson (1967)). In general the form (3.3) restricts only the distribution changes with respect to its parameters, and not its shape for any particular parameter value -- let $\tilde{p}(X)$ represent any desired distribution at $\theta = \theta_0$, then for given $D(X)$ and $\pi(\theta)$, by defining $h(X)$ in (3.3) as $h(X) = \tilde{p}(X)/\exp(\pi(\theta_0)'D(X))$, we construct an exponential family form $p(X|\theta)$ with $p(X|\theta_0) = \tilde{p}(X)$. For example, to obtain P_N , we take $\tilde{p}_0(X)$ as the normal density with fixed mean and covariance matrix, and set $D(X) = (X_1, X_2, \dots, X_M, X_1^2, X_1 X_2, \dots, X_M^2)$. P_{NT} is derived by setting $D(X) = X$.

We illustrate completeness for the exponential family by considering aggregation over P_N . All aggregation structures (C, P_N) are complete by Theorem 3. For example, consider (C_Q, P_N) . The quadratic micro function $F_{\gamma}(X)$ of (2.8) implies the aggregate function

$$\phi_{\gamma}(\theta) = E(F_{\gamma}(X)) = \gamma_0 + \sum_i \gamma_{1i} \mu_i + \sum_{ij} \gamma_{2ij} (\mu_i \mu_j + \sigma_{ij})$$

Now suppose that we have $\gamma, \gamma' \in \Gamma$ such that

$$\phi_{\gamma}(\theta) = \phi_{\gamma'}(\theta) \tag{3.4}$$

for all $\theta \in \Theta$. Then by taking first and second partial derivatives of (3.4) with respect to μ_i and μ_j , $i, j=1, \dots, M$, we easily establish that $\gamma = \gamma'$, so that (C_Q, P_N) is complete.

A second, more interesting example is (C_{PB}, P_N) . The probit micro function $F_Y(X)$ of (2.7) implies the aggregate function (c.f. McFadden and Reid (1975))

$$\phi_Y(\theta) = E(F_Y(X)) = \Phi \left[\frac{\gamma_0 + \mu' \gamma_1}{\sqrt{1 + \gamma_1' \Sigma \gamma_1}} \right] \quad (3.5)$$

where $\phi_Y(\theta)$ refers to the true proportion of agents with $y = 1$. Completeness can easily be verified as above. Notice that $\phi_Y(\theta)$ coincides with the micro function $F_Y(X)$ evaluated at $X = \mu$ only if $\Sigma = 0$, which is disallowed by assumption. Nevertheless, the completeness property implies that an analyst with many observations on (consistent estimates of) $E(Y)$, μ and Σ , could estimate the micro function parameters γ_0, γ_1 .

In showing the completeness of (C_Q, P_N) , we only needed to take partial derivatives of (3.4) with respect to the components of μ . This reflects the fact that P_{NT} , the normal class with fixed covariance matrix, is also complete by Theorem 3. In contrast, P_{NV} , the normal class with fixed mean, does not insure completeness, as (C_Q, P_{NV}) is not complete (see Kendall and Stuart (1973, p.200) for other examples). While P_{NV} is in exponential family form, it is easily verified that natural parameter range $\pi(\theta)$ does not contain an open set, and so Theorem 3 does not apply.

3.4 "Fixed Distributions" and Homogeneous Functions

Often aggregation questions are posed when the microvariables X have a fixed distribution relative to varying means of X (see Chipman (1974) among others). This does not imply a constant density in our framework, but rather special forms of densities. Here we give examples of complete aggregation structures based on homogeneous micro functions and "fixed" distribution classes.

We consider only non-negative micro variables, with $X \in \Omega = R_+^M = \{X \in R^M \mid X \geq 0\}$. Micro behavior is specified as $y = f_Y(X, u)$, where f_Y is homogeneous of degree $\lambda_Y \geq 0$, and we assume u is distributed independently of X . The micro function corresponding to f_Y , $Y = F_Y(X)$, is clearly homogeneous of degree λ_Y in X . Let

C_H denote the class of all such functions. We will also consider the subclasses $C_H^\lambda \subset C_H$ defined by $C_H^\lambda = \{F_Y(X) \in C_H \mid \lambda_Y = \lambda\}$, the functions homogeneous of degree λ .

Another subclass, denoted C_C , is defined as $C_C = \{F_Y^C(X)\} \subset C_H$ where

$$F_Y^C = \gamma_0 X_1^{\gamma_1} \dots X_M^{\gamma_M}$$

and $\gamma = (\gamma_0, \dots, \gamma_M)$, the set of (Cobb-Douglas) geometric average functions, where $\lambda_Y = \sum_{i=1}^M \gamma_i$ and $\gamma_j \geq 0, j=0, \dots, M$.

We consider three versions of fixed density classes. Each is derived by specifying a base variable x with density $\tilde{p}(x)$, for which it is assumed that all (uncentered) moments are finite, and a mapping $X = \delta(x, \theta)$, from which densities of X are induced:

$P_1 = \{p(X|\theta)\}_{\theta \in \Theta}$, where x is a univariate random variable with density $\tilde{p}(x)$, $E(x) = 1$, and $X = \theta \cdot x$, where $\theta \in \Theta = R_+^M$.

$P_2 = \{p(X|\theta)\}_{\theta \in \Theta}$, where x is an M component random variable with density $\tilde{p}(x)$, $E(x) = i$ (the vector of ones), and $X = \theta \cdot x$, where θ is a scalar parameter, $\theta \in \Theta = R_+$.

$P_3 = \{p(X|\theta)\}_{\theta \in \Theta}$, where x is an M component random variable with density $\tilde{p}(x)$, $E(x) = i$ and $X = (\theta_1 x_1, \dots, \theta_M x_M)$ where $\theta = (\theta_1, \dots, \theta_M) \in \Theta = R_+^M$.

P_1 denotes densities where the micro variables are perfectly correlated, with $E_\theta(X) = \theta$. P_2 denotes densities with proportional expansion (rescaling) of the base distribution, with $E_\theta(X) = \theta i$. P_3 denotes densities where each micro variable expands (i.e. is rescaled) by different amounts, with $E_\theta(X) = \theta$.

These families are identical when $M = 1$, which is the case studied by Chipman (1974).

It is easy to see that P_1 is complete for C_H , and consequently for C_H^λ , all $\lambda > 0$ and C_C . If $F_Y(X) \in C_H$ is homogeneous of degree λ_Y , then

$$E_\theta(y) = \phi_Y(\theta) = \int_{\Omega} F_Y(\theta x) \tilde{p}(x) dx = F_Y(\theta) \int_{\Omega} x^{\lambda_Y} \tilde{p}(x) dx = m_{\lambda_Y} \cdot F_Y(\theta)$$

where m_{λ_Y} is the λ_Y moment of p defined above. Obviously if $\phi_Y(\theta) = \phi_{Y'}(\theta)$ for all $\theta \in \Theta$, then $\gamma = \gamma'$. Note also that for C_H^1 , an aggregate function is of exactly the same form as the corresponding micro function.

It is also easy to see that if $M \geq 2$, then P_2 is not complete for C_H , any C_H^λ or C_C . Take $M = 2$, and $F_Y(X) \in C_C$ be given by $F_Y(X) = \gamma_0 X_1^{\gamma_1} X_2^{\gamma_2}$. Then

$$\begin{aligned} \phi_Y(\theta) &= \int_{\Omega} \gamma_0 (\theta x_1)^{\gamma_1} (\theta x_2)^{\gamma_2} \tilde{p}(x_1, x_2) dx_1 dx_2 \\ &= \gamma_0 \theta^{\gamma_1 + \gamma_2} \int x_1^{\gamma_1} x_2^{\gamma_2} \tilde{p}(x_1, x_2) dx_1 dx_2 \\ &= \gamma_0 \theta^{\gamma_1 + \gamma_2} m_{\gamma_1, \gamma_2} \end{aligned}$$

where m_{γ_1, γ_2} denotes the γ_1, γ_2 moment of \tilde{p} defined above. The failure of completeness is evident by noting that $\phi_Y(\theta) = \phi_{Y'}(\theta)$ where $F_{Y'}(X) = \gamma_0' X_1^{\gamma_1 + \gamma_2}$, with $\gamma_0' = \gamma_0 m_{\gamma_1, \gamma_2} / m_{\gamma_1 + \gamma_2}$. It is also obvious that every aggregate function derivable from $A = \{C_H, P_2\}$ takes the form of a constant times θ raised to a power.

In addition, we can show that P_3 is complete for C_C , although whether P_3 is complete for C_H is not known. To see that P_3 is complete for C_C , let $F_Y(X) \in C_C$ where $F_Y(X) = \gamma_0 X_1^{\gamma_1} \dots X_M^{\gamma_M}$. Then

$$\begin{aligned} \phi_Y(\theta) &= \gamma_0 \theta_1^{\gamma_1} \dots \theta_M^{\gamma_M} \int_{\Omega} x_1^{\gamma_1} \dots x_M^{\gamma_M} \tilde{p}(x) dx \\ &= \gamma_0 \theta_1^{\gamma_1} \dots \theta_M^{\gamma_M} m_{\gamma_1, \dots, \gamma_M} \end{aligned}$$

If $\phi_Y(\theta) = \phi_{Y'}(\theta)$ for all $\theta \in \Theta$, then clearly $\gamma_j = \gamma_j'$, $j=1, \dots, M$ from which $\gamma_0 = \gamma_0'$ follows.

3.5 Discrete and Segmented Population Distributions

There are situations where the basic micro variables X can assume only a finite number of different values, or where the population is naturally segmented into a finite number of groups. For example, a simple "end use" model

of aggregate gasoline demand may specify individual family gasoline purchases as dependent on the number of cars owned. Alternatively, in building demographic effects into aggregate demand, one may regard the total expenditure distribution as segmented by family size and age-of-head classes (see Jorgenson, Lau and Stoker (1982)). An example where the total expenditure distribution is segmented into expenditure size classes is Shapiro and Braithwaite (1979). In addition, if a continuous distribution (such as the income distribution) were observed by cell proportions (income classes), and the within cell distributions were unchanging, then the proper model is discrete (see the example in Section 4). In this section we first consider completeness for discrete distributions and then generalize to segmented distributions.

Suppose that there are N possible different values of the predictor variables $X \in \Omega = \{X^1, \dots, X^N\}$. A density $p(\theta) = (p_1(\theta), \dots, p_N(\theta))'$ is an N vector such that $i'p(\theta) = 1$, and $p_j(\theta) \geq 0$ denotes the proportion of agents with $X = X^j, j=1, \dots, N$. The most natural (and completely unrestricted) parameterization of a discrete density is $p(\theta) = \theta$, where $\theta \in \Theta = \{\theta \in \mathbb{R}^N \mid i'\theta = 1, \theta \geq 0\}$. However, since we wish to include situations where the distribution of X is structured, for example, if $X = X^j$ itself results from a discrete choice process, we retain the general specification $p(\theta)$. The expected value of X is just $\mu(\theta) = \sum p_j(\theta) X^j$. A micro function $Y = F_Y(X)$ is an N vector $Y_Y = (Y_Y^1, \dots, Y_Y^N)'$ defined by $Y_Y^j = F_Y(X^j)$.

Suppose that $P = \{p(\theta)\}_{\theta \in \Theta}$ represents a class of N cell discrete distributions. For the completeness property, we would like to pick $\theta^1, \dots, \theta^N$, such that the linear equations

$$\begin{aligned} Y_Y' p(\theta^1) &= \phi_Y(\theta^1) \\ &\vdots \\ Y_Y' p(\theta^N) &= \phi_Y(\theta^N) \end{aligned} \tag{3.6}$$

could be solved for Y_Y . But this is possible for all $Y_Y \in \mathbb{R}^N$ if and only if $p(\theta^1), \dots, p(\theta^N)$ are linearly independent. Consequently, we have shown

THEOREM 4:

A class $P = \{p(\theta)\}_{\theta \in \Theta}$ of N cell discrete distributions is complete for the class of all micro functions if and only if there exists $\theta^1, \dots, \theta^N$ such that $p(\theta^1), \dots, p(\theta^N)$ span R^N .

Completeness (for all micro functions) fails if the definition of the class P implies a linear constraint on $p(\theta)$ for all θ other than $\sum p(\theta) = 1$. Completeness here is literally the absence of collinearity between the elements of P .

We model segmented distributions as follows. Let $\Omega^1, \dots, \Omega^N$ denote a mutually exhaustive partition of Ω , and partition the parameter vector θ as $\theta = (\theta_0, \theta_1, \dots, \theta_N) \in \Theta = \theta^0 \times \theta^1 \times \dots \times \theta^N$. Let $p^0(\theta_0) = (p_1^0(\theta_0), \dots, p_N^0(\theta_0))$ denote an N cell discrete density, with $p_j^0(\theta_0)$ denoting the probability that $X \in \Omega^j$. Let $p^j(X|\theta_j)$ denote the distribution of X conditional on $X \in \Omega_j$. Define $P^0 = \{p^0(\theta_0)\}_{\theta_0 \in \Theta^0}$ as the class of "between cell" distributions, and $P^j = \{p^j(X|\theta_j)\}_{\theta_j \in \Theta^j}; j=1, \dots, N$ as the classes of within cell distributions. The class of segmented distributions $P = \{p(X|\theta)\}_{\theta \in \Theta}$ induced by P^0, P^1, \dots, P^N is defined via $p(X|\theta) = \sum_j p_j^0(\theta_0) p^j(X|\theta_j)$ where $\theta = (\theta_0, \theta_1, \dots, \theta_N)$. For any class of micro functions $C = \{F_\gamma(X)\}_{\gamma \in \Gamma}$, define N classes of functions C^1, \dots, C^N as $C^j = \{F_\gamma^j(X)\}_{\gamma \in \Gamma}$, where

$$\begin{aligned} F_\gamma^j(X) &= F_\gamma(X) & \text{if } X \in \Omega^j \\ &= 0 & \text{if } X \notin \Omega^j \end{aligned} \tag{3.7}$$

Given the partitioning above, for any class of functions $C = \{F_\gamma(X)\}_{\gamma \in \Gamma}$, we can define a new class of functions $C_P = \{F_{\gamma^*}(X)\}_{\gamma^* \in \Gamma^*}$ by first constructing $C^j, j=1, \dots, N$ as in (3.7), defining $\Gamma^* = (\gamma_1, \dots, \gamma_N) \in \Gamma^* = \Gamma^N$ and $F_{\gamma^*}(X)$ via $F_{\gamma^*}(X) = F_{\gamma_j}(X)$ if $X \in \Omega_j$. In particular, if the original class is C_L , the class of linear micro functions, we derive the class of piecewise linear functions C_{PL} as above.

It is easy to show (see the Appendix) the following characterization of completeness for segmented distributions.

THEOREM 5:

Let P be induced by P^0, P^1, \dots, P^N where P^0 is complete in the sense of Theorem 4. Then P is complete for C_P if and only if P^j is complete for $C^j, j=1, \dots, N$.

If P^0 is not complete, then the class P can be re-defined with respect to a coarser partition $\bar{\omega}^1, \dots, \bar{\omega}^{N'}$, with $N' < N$, so that the resulting class \bar{P}^0 is complete. Clearly if P is complete for C_P , then it is complete for C .

By applying Theorems 1 and 5, we see that the class of piecewise linear functions C_{PL} is "within distribution free" for segmented distributions.

THEOREM 6:

Let P be induced by P^0, P^1, \dots, P^N and let P^0 be complete in the sense of Theorem 4. If each $P^j, j=1, \dots, N$ is mean full dimensional, then P is complete for C_{PL} . If a class C strictly contains C_{PL} , i.e. there exists $G(X) \notin C_{PL}, G(X) \in C$ where $C_{PL} \subset C$, then there exists mean full dimensional within cell density classes $P_j, j=1, \dots, N$ such that P is not complete for C .

An excellent example of modeling based on Theorem 6 is Shapiro and Braithwaite (1979), who characterize empirically the coarsest partition (smallest N) of the total expenditure distribution, for which micro Engel curves can be adequately taken as elements of C_{PL} .

4. Remarks on the Use of Aggregate Functions

4.1 A Numerical Example

In this section we illustrate completeness using a very simple example. Suppose that we are interested in the demand for a particular commodity Y over a time period of constant prices, and our theory says that Y is some function $F(X)$ of total expenditure X for all individuals. Figure 1 shows four possible Engel curve shapes -- F_1 represents unitary elastic demand, F_2 a luxury good,

with income elasticity greater than one, F_3 a normal good with positive elasticity less than one, and F_4 a luxury good for low total expenditures and a normal good for high total expenditures. Suppose that we will collect data on average demand $\phi = E(Y)$ and average total expenditure $\mu = E(X)$ over six time periods with which to characterize the demand structure.

For simplicity, we characterize the problem as one of discrete aggregation by assuming that the total expenditure distribution is delineated into five classes, with average demand and average total expenditures within each class constant over time. Total expenditure and the four possible Engel curves are now characterized by their within cell averages, given in Table 1, with movements in overall average consumption and average total expenditure determined by distributional shifts between classes.

We now consider two different distributional scenarios over the six time periods. The first, in Table 2, gives rise to different time patterns of average demand for each of the four Engel curves. Here, if the distribution data were observed, it could be used together with the (overall) average demand data to characterize the true Engel curve (i.e. pattern of within cell averages). Formally, equations (3.6) can be solved for $Y_Y = F_Y$, since the distribution vectors $p_t, t=1, \dots, 6$ span a five-dimensional set. The completeness property obtains in this data.

Now consider the distributional scenario in Table 3. Here, each of the four Engel curve shapes gives rise to the same pattern of overall average demand, which is 20% of average total expenditure in each time period. Here, very little is learned about behavior from the average demand data, unless the Engel curve is a priori restricted to be linear, or F_1 here. The distribution vectors, $p_t, t=1, \dots, 6$ span a two-dimensional set (following LPM structure of Section 4.2), and represent an extreme failure of the completeness property.

This example similarly illustrates the classic concerns about aggregation raised by Gorman (1953) and De Wolff (1941). Suppose that an analyst collected

the ϕ_t and μ_t data from Table 3, and to his delight, found that the equation $\phi_t = .2\mu_t$ explained the data perfectly. The problem occurs when he interprets .2 as the marginal propensity to spend (MPS) on Y. The question here is whose MPS? If the answer is the 'representative consumer,' the ambiguity is shifted to the definition of 'representative.' Only when micro behavior is $Y = F_1(X)$ is the answer easy -- 'everyone's MPS.' This is Gorman's point. And if the equation were used to forecast, it would rely on the existing pattern of distribution movements continuing in the forecasted period, unless $Y = F_1(X)$ truly represents behavior. This point was raised by De Wolff.

4.2 The Failure of Completeness and Linear Aggregate Functions

The data of Table 3 above underscore a central concern of this paper that not much can be learned about micro behavior from aggregate data unless a complete aggregation structure is specified. For more general illustration of this important point, consider an aggregate function which is linear in the means of X

$$E_{\theta}(Y) = a + \mu'b \quad (4.1)$$

where $\mu = E_{\theta}(X)$. From Theorem 2a), if the class of micro functions is C_L , then (4.1) identifies $F_Y(X) = a + X'b$. However, from the proof of Theorem 2b), we see that if $P = \{p(X|\theta)\}$ is defined by setting $\theta = \mu$ and

$$p(X|\mu) = p_0(X) + \mu_1 p_1(X) + \dots + \mu_M p_M(X) \quad (4.2)$$

where by construction $\int p_0 dX = 1$, $\int p_j dX = 0, j=1, \dots, M$ and $E(X) = \mu = (\mu_1, \dots, \mu_M)'$, then (4.1) is guaranteed. Namely, if $F(X)$ is an arbitrary (measurable) function, then

$$\begin{aligned} E(F(X)) &= \int F(X) p_0(X) dX + \sum_i \mu_i \left[\int F(X) p_i(X) dX \right] \\ &= a + \sum_i \mu_i b_i = a + \mu'b \end{aligned}$$

The class defined by (4.2) is said to obey linear probability movement (LPM). The function $p_0(X)$ is a base density, with changes in $p(X|\theta)$ occurring linearly with

respect to μ through the M "density shift" functions $p_j(X), j=1, \dots, M$.

The LPM form (4.2) represents a very strong restriction on the form of distribution movement, however its existence points out a troublesome aspect of interpreting the estimated aggregate functions. Suppose that in an empirical application, the actual population densities approximately follow LPM, then it can be shown (c.f. Stoker (1981)) that an estimated linear aggregate function of form (4.1) will fit the observed data configuration quite closely, regardless of the true form of micro behavior. Thus, standard goodness of fit tests can fail to indicate any problems of model specification, but the estimates of a and b of (4.4) reveal little about behavior at the individual level. This danger exists when interpreting any estimated aggregate function unless a complete aggregate structure is specified.¹³

4.3 Remarks on Related Empirical Techniques

4.3a Completeness of Distribution Data

When the proper model for an application is discrete, it is often possible to nonparametrically characterize the true micro function using both average data and distribution data over time (as in Table 2). This method of micro function characterization is detailed in Stoker (1981), including simple tests of micro linearity. A natural question is whether general (non-discrete) distribution data can insure completeness, so that specific assumptions on C are not required to characterize the true micro function.

Formally, suppose that we are in an idealized situation where the true density of X , denoted $p_t(X)$, is observed for $t \in \tau = \{1, \dots, T\}$, and define $P_T = \{p_t(X)\}_{t \in \tau}$ as the observed class. Unfortunately, if X is continuously distributed and T is finite, it is impossible for P_T to be complete. Technically the reason is that the observed data places only T constraints on the true micro functions -- $E_t(Y) = \int F(X)p_t(X)dX$ -- but the space C_U of all measurable functions has an infinite basis, so that an infinite number of functions

satisfy the constraints. Consequently, C must be restricted to imply completeness of (C, P_T) .

4.3b Characterization with Individual Cross Section Data

As a final remark, we indicate a few results which allow the characterization of aggregate functions using cross section data $y_k, X_k, k=1, \dots, K$ observed for a particular time period. Suppose that $\theta = \mu = E(X)$ where $\mu = \mu_0$ is the cross section parameter value, and that $E(y) = \phi(\mu)$ denotes the true aggregate function. It is shown in Stoker (1983) (under some regularity conditions) that the macroeconomic effects $\frac{\partial \phi}{\partial \mu}(\mu_0)$ are consistently estimated by the slope coefficients of the instrumental variables regression of y_k on X_k , where the instruments are components of the locally unbiased and efficient estimator of $\mu = \mu_0$ based on the cross section data. When the density class is in exponential family form with $D(X) = X$, this result says that the OLS slope coefficients of y_k on X_k consistently estimate $\frac{\partial \phi}{\partial \mu}(\mu_0)$. These results provide the basis for tests of aggregate function structure using cross section data, such as linearity of $\phi(\mu)$ in μ (see Stoker (1982, 1983)). Moreover, the results provide general interpretations of cross section regression coefficients in misspecified circumstances, as they are valid when no specific restrictions are imposed on either the true micro function or the cross section distribution of X .

5. Conclusion

In this paper we have introduced the property of completeness of aggregation structures, argued for its importance, and provided several illustrative examples. Brushing all the technical detail aside, a major lesson of this paper is that aggregation problems can be very serious in the context of interpreting aggregate functions as representing behavior. As with all identification problems, difficulties with aggregation are particularly onerous in that they are often undetectable by specification (goodness of fit) tests applied to estimated aggregate functions.

The issues highlighted here are somewhat worrisome for studies of aggregate data using models derived from the theory of individual behavior, such as is commonplace in macroeconomics. This is not to say that any existing behavioral interpretations of estimated aggregate functions are incorrect. However, it is true that any behavioral interpretation of an aggregate function made without mention of a complete aggregation structure is without proper foundation. Consequently, an implication of this paper is that the aggregation assumptions underlying an estimated aggregate function should always be made explicit, so that the quality of particular interpretations can be judged in the light of all the available evidence, including that on micro behavior obtained from data on individuals.

The only valid alternative to using a complete aggregation structure is to adopt a purely reduced form approach to studying aggregate data, as is suggested in the work of Sonnenschein and others on excess demand functions, and advocated by prominent econometricians such as Granger and Simms. But by relinquishing the ability to attach behavioral interpretations to estimated aggregate functions, their use in applications is limited.

APPENDIX: PROOFS OF THEOREMS 2 AND 5

Proof of Theorem (2a): Let $F_Y(X) \in C_L$, $F_Y(X) = \gamma_0 + X'\gamma_1$, $\gamma = (\gamma_0, \gamma_1')'$. Let $P = \{p(X|\mu, \theta_1)\}$ be a mean full dimensional family of densities, reparameterized by the mapping (3.2). The aggregate function ϕ_Y corresponding to $F_Y(X)$ is just $\phi_Y(\mu, \theta_1) = \gamma_0 + \mu'\gamma_1$. Now, suppose that $\gamma = (\gamma_0, \gamma_1')'$, $\bar{\gamma} = (\bar{\gamma}_0, \bar{\gamma}_1')'$ are such that $\phi_Y(\mu, \theta_1) = \phi_{\bar{Y}}(\mu, \theta_1)$ for all μ, θ_1 . We have that

$$(\mu_1 - \mu_2) (\gamma_1 - \bar{\gamma}_1) = 0$$

for all μ_1, μ_2 . Since P is mean full dimensional this implies that $\gamma_1 - \bar{\gamma}_1 = 0$, which implies that $\bar{\gamma}_0 = \gamma_0$, i.e. that $\gamma = \bar{\gamma}$. Q.E.D.

Proof of Theorem (2b): We take $\mu = (\mu_1, \dots, \mu_M) = \theta \in \Theta \subset R^M$. We seek a density class $P = \{p(X|\mu)\}_{\mu \in \Theta}$ defined as

$$p(X|\mu) = p_0(X) + \mu_1 p_1(X) + \dots + \mu_M p_M(X) \quad (A.1)$$

where Θ contains an open subset of R^M , $E_\mu(X) = \mu$ and $\int p_0(X) dX = 1$, $\int p_j(X) dX = 0$ for $j=1, \dots, M$. If there exists such a class, then for $G(X) \in C$, $G(X) \notin C_L$ we have

$$\begin{aligned} E_\mu(G(X)) &= \int G(X) p_0(X) dX + \sum_{i=1}^M \mu_i \int G(X) p_i(X) dX \\ &= a + \sum_{i=1}^M \mu_i b_i = a + \mu'b = E_\mu(F_Y(X)) \end{aligned}$$

where $\gamma = (a, b')'$, $F_Y(X) \in C_L$, and so P is not complete for C .

To show the existence of such a density class, we construct one for $\Omega = [a, b]^M \subset R^M$, i.e. for X bounded (just to insure the existence of various integrals). Let $\tilde{p}_0(X)$ denote any density of X on Ω , such that $\tilde{p}_0(X) > 0$ for $X \in \Omega$, with $\int p_0(X) dX = 1$ and $\int X \tilde{p}_0(X) dX = \mu_0$, a constant value. For example, \tilde{p}_0 could be taken as the uniform distribution; $\tilde{p}_0(X) = 1/(b-a)^M$.

Now, we can apply the Gram Schmidt orthogonalization process (c.f. Hadley (1961)) to the functions $\tilde{p}_0(X)$, $g_1(X) = X_1, \dots, g_M(X) = X_M$ to construct $\tilde{p}_j(X), j=1, \dots, M$ recursively as

$$\tilde{p}_1(x) = x_1 - d_1 \tilde{p}_0(x)$$

$$p_j(x) = x_j - \sum_{i=0}^{j-1} d_{ij} \tilde{p}_i(x)$$

where $d_{i0} = \int x_i \tilde{p}_0(x) dx / s_0$, $d_{ij} = \int x_j \tilde{p}_i(x) dx / s_i$, $s_j = \int [\tilde{p}_j(x)]^2 dx$ (s_j, d_{ij} , $i=1, \dots, N, j=0, 1, \dots, N$ exist because each of the integrands is obviously bounded). $\tilde{p}_0(x), \tilde{p}_1(x), \dots, \tilde{p}_M(x)$ are now nonzero functions such that $\int \tilde{p}_i(x) \tilde{p}_j(x) dx = 0$ if $i \neq j$. Now define $D_1(\mu_1), D_2(\mu_1, \mu_2), \dots, D_M(\mu_1, \dots, \mu_M)$ recursively as

$$D_1(\mu_1) = \frac{1}{s_1} (\mu_1 - d_{10} s_0)$$

$$D_j(\mu_1, \dots, \mu_j) = \frac{1}{s_j} (\mu_j - d_{j0} s_0 - \sum_{i=1}^{j-1} D_i(\mu_1, \dots, \mu_i) d_{ij} s_i) \quad j=2, \dots, M$$

and form

$$p(x|\mu) = \tilde{p}_0(x) + \sum_{i=1}^M D_i(\mu_1, \dots, \mu_i) \tilde{p}_i(x) \quad (A.2)$$

By construction we have $\int x p(x|\mu) dx = \mu$ and $p(x|\mu_0) = \tilde{p}_0(x) > 0$ for $x \in \Omega$. Since each D_i function is linear in its arguments, there exists an open neighborhood θ of μ_0 , $\theta \subset \mathbb{R}^M$ such that if $\mu \in \theta$, then $p(x|\mu) > 0$ for $x \in \Omega$. Also since each $D_i(\mu_1, \dots, \mu_i)$ is linear in its arguments, (A.2) can be rewritten in form (A.1), with the other conditions easily verifiable. Thus, the family $P = \{p(x|\mu)\}_{\mu \in \theta}$ serves the purpose of the theorem.

Q.E.D.

Proof of Theorem 5:

Let $F_{\gamma^*}(x) \in C_p$, with $\gamma^* = (\gamma_1, \dots, \gamma_N) \in \Gamma^N$. Then

$$\phi_{\gamma^*}(\theta) = \sum_{i=1}^N p_i^0(\theta_0) \int_{\Omega_i} F_{\gamma_i}(x) p^i(x|\theta_i) dx \quad (A.3)$$

where $\theta = (\theta_0, \theta_1, \dots, \theta_N)$. Now, since P^0 is complete, we can in general invert (A.3), i.e. find unique functions $G_1^\gamma(\theta), \dots, G_N^\gamma(\theta)$ consistent with (A.3) such that

$$\int_{\Omega_i} F_{\gamma_i}(x) p^i(x|\theta_i) dx = G_i^\gamma(\theta) \quad i=1, \dots, N \quad (A.4)$$

for all $\theta \in \theta$. Now, if P^i is complete for $C^i, i=1, \dots, N$, then eqns. (A.4) determine F_{γ_i} uniquely, $i=1, \dots, N$, and so the original F_{γ^*} is determined uniquely, so that P is complete for C_p . Conversely, suppose that P is complete for C_p , but there exists i such that P^i is not complete for C^i .

Then there exists $F_Y(X), F_{Y'}(X) \in C^i$ such that

$$\int_{\Omega_i} F_Y(X) p^i(X|\theta_i) dX = \int_{\Omega_i} F_{Y'}(X) p^i(X|\theta_i) dX$$

for all $\theta_i \in \Theta^i$, with $F_Y(X) \neq F_{Y'}(X)$ for a subset of nonzero measure of Ω_i . But then if we consider $F_{Y^*}(X)$ and $F_{Y^{*'}}(X)$, $Y^* = (\gamma_1, \dots, \gamma_{i-1}, \gamma, \gamma_{i+1}, \dots, \gamma_N)$ and $Y^{*'} = (\gamma_1, \dots, \gamma_{i-1}, \gamma', \gamma_{i+1}, \dots, \gamma_N)$, we have $\phi_{Y^*}(\theta) = \phi_{Y^{*'}}(\theta)$, which is a contradiction of the completeness of P for C_P .

Q.E.D.

FOOTNOTES

1. Aggregate functions correspond to models estimated with averaged data, which include usual macroeconomic equations such as consumption and investment functions, as well as so-called "microeconomic" models used to study average production and demand variables. All points discussed apply also to models of "totaled" variables.
2. While the analysis is quite general, we do not cover externalities in aggregation (for instance, Liebenstein (1950)), varying domain problems (c.f. Houthakker (1956), Sato (1975), Hildenbrand (1981) among others, where current price levels determine the percentage of operating firms), nor do we address whether it is possible to specify different behavioral functions across agents which rationalize any general form of aggregate function (as in Sonnenschein (1972,1973) and Debreu (1974), among others).
3. A set $\Omega \subset \mathbb{R}^M$ is full dimensional if it is not contained in a proper affine subspace of \mathbb{R}^M , i.e. if there exists an M vector α such that $\alpha'(X_1 - X_2) = 0$ for all $X_1, X_2 \in \Omega$, then $\alpha = 0$.
4. Our overall approach is known as Pareto's Stratification Approach, traced to Pareto in 1895 by Wold and Jureen (1953) and Green (1964).
5. We treat Y as a single variable, however exactly the same definitions apply if Y is a q vector and $F_\gamma(X)$ a q vector function of X , where γ may be used to capture restrictions across components of $F_\gamma(X)$.
6. Heckman and Singer (1982) employ the completeness property in a different context to nonparametrically characterize the distribution of agent heterogeneity using duration data.
7. For concreteness, suppose that (2.1) models individual demand over a period of constant prices. y denotes quantity demanded, X denotes observed income and demographic variables, u denotes unobserved agent differences, and γ denotes preference parameters. $F_\gamma(X)$ is the Engel Curve, $p(X|\theta)$ the income-demographic distribution and $E(y) = \phi_\gamma(\theta)$ is the model between mean demand $E(y)$ and the distribution parameters γ, θ .
8. For constant prices, form (2.7a) is derived in Gorman (1953), Muellbauer (1975,1976) and Lau (1977,1981), where X denotes an M vector function of individual agent attributes and total expenditures. Demand systems obeying these restrictions include Deaton and Muellbauer's (1980) AIDS system, and Jorgenson, Lau and Stoker's (1982) translog system.
9. Form (2.7b) is utilized in Theil (1954). This framework is employed in applications of the Rotterdam system (see Theil (1971,1975) and Barnett (1981) for citations), and is generalized by Barnett (1979). To avoid confusion, we note that our behavioral functions (2.7b) and (2.5)) are referred to as "micro functions" in these references.
10. These issues go hand-in-hand with the specification of X and u . For example, Jorgenson, Lau and Stoker (1982) model agent demographic differences via observed variables X , whereas Theil (1975, Chapter 4) treats agent differences as random (via u in our framework).

11. This is consistent with the exact aggregation theorems of Gorman (1953), Muellbauer (1975,1976) and Lau (1977,1981). It should be noted that these papers employ a different framework (finite number of agents) and address the more general question of whether total demand depends on a small number of symmetric functions of individual variables, for arbitrary changes in the distribution of those variables (which may not follow a specific parameterization). Moreover, our Definition 4 concerns the particular specification of X of interest, and requires that $\mu = E(X)$ can be taken as a subvector of the parameters of density movement.
12. We should note that there exist some rather pathological complete aggregation structures that are distribution free but do not involve micro linearity, such as when the class C of micro functions contains only a single element, which then may be a nonlinear function of X .
13. It is easy to find versions of LPM for virtually any continuous nonlinear aggregate function. Suppose $E(y) = g(\theta)$ is an aggregate function of interest. In a way analogous to the proof of Theorem 2b) we can construct a density family \tilde{P} such that every aggregate function derived from \tilde{P} is linear in $g(\theta)$.

REFERENCES

- Amemiya, T. (1981), "Qualitative Response Models: A Survey," Journal of Economic Literature, 19, pp. 1483-1536.
- Barndorff-Neilsen, O. (1978), Information and Exponential Families in Statistical Theory, Wiley, New York.
- Barnett, W. A. (1979), "Theoretical Foundations of the Rotterdam Model," Review of Economic Studies, 46, pp. 109-130.
- Barnett, W. A. (1981), Consumer Demand and Labor Supply, North-Holland, Amsterdam.
- Chipman, J. (1974), "Homothetic Preferences and Aggregation," Journal of Economic Theory, 8, pp. 26-38.
- Deaton, A., and J. Muellbauer (1980), "An Almost Ideal Demand System," American Economic Review, 70, pp. 312-326.
- Debreu, G. (1974), "Excess Demand Functions," Journal of Mathematical Economics, 1, pp. 15-23.
- Efron, B. (1975), "Defining the Curvature of a Statistical Problem (with Application to Second Order Efficiency)," Annals of Statistics, 3, pp. 1189-1242.
- Efron, B. (1978), "The Geometry of Exponential Families," Annals of Statistics, 6, pp. 362-376.
- Ferguson, T. S. (1967), Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York.
- Gorman, W. M. (1953), "Commodity Preference Fields," Econometrica, 21, pp. 63-80.
- Green, H. A. J. (1964), Aggregation in Economic Analysis, Princeton University Princeton.
- Hadley, G. (1961), Linear Algebra, Addison-Wesley, Reading, Massachusetts.
- Heckman, J. and B. Suiger (1982), "The Identification Problem in Econometric Models for Duration Data," Advances in Econometrics, Hildenbrand, ed., Cambridge University Press, New York.
- Hildenbrand, W. (1981), "Short-Run Production Functions Based on Microdata," Econometrica, 49, 5, pp. 1095-1125.
- Houthakker, H. S. (1956), "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," Review of Economic Studies, 23, pp. 27-31.
- Jorgenson, D. W., L. J. Lau and T. M. Stoker (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in R. Basman and G. Rhodes, eds., Advances in Econometrics, JAI Press.
- Kendall, M. G., and A. Stuart (1973), The Advanced Theory of Statistics, 3rd ed., Vol. 2, Griffin and Company, London.
- Lau, L. J. (1977), "Existence Conditions for Aggregate Demand Functions," Technical Report No. 248, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California (revised February 1980).

- Lau, L. J. (1982), "A Note on the Fundamental Theorem of Exact Aggregation," Economics Letters, Vol. 9, 2, pp. 119-126.
- Lehmann, E. L. (1959), Testing Statistical Hypotheses, Wiley, New York.
- Lehmann, E. L., and H. Scheffe (1950), "Completeness, Similar Regions and Unbiased Estimation, Part I," Sankya, 10, pp. 305-340.
- Lehmann, E. L., and H. Scheffe (1955), "Completeness, Similar Regions and Unbiased Estimation, Part II," Sankya, 15, pp. 219-236.
- Leibenstein, H. (1950), "Band Wagon, Snob and Veblen Effects in the Theory of Consumer Demand," Quarterly Journal of Economics, May.
- Manski, C., and D. McFadden, eds. (1982), Structural Analysis of Discrete Data With Econometric Applications, MIT Press, Cambridge, Massachusetts.
- McFadden D., and F. Reid (1975), "Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models," Transportation Research Record, No. 534.
- Muellbauer, J. (1975), "Aggregation, Income Distribution and Consumer Demand," Review of Economic Studies, 42, pp. 525-543.
- Muellbauer, J. (1976), "Community Preferences and the Representative Consumer," Econometrica, 44, No. 5, pp. 979-999.
- Rao, C. R. (1973), Linear Statistical Inference and Its Applications, 2nd ed., Wiley, New York.
- Sato, K. (1975), Production Functions and Aggregation, North-Holland, Amsterdam.
- Shapiro, P., and S. Braithwaite (1979), "Empirical Tests for the Existence of Group Utility Functions," Review of Economic Studies, 46, No. 4, October.
- Sonnenschein, H. (1972), "Market Excess Demand Functions," Econometrica, 40, No. 3, pp. 549-563.
- Sonnenschein, H. (1973), "The Utility Hypothesis and Market Demand Theory," Western Economic Journal, 11, pp. 404-410.
- Stoker, T. (1981), "The Use of Average and Distribution Data to Characterize Micro Functions," draft, August.
- Stoker, T. (1982), "The Use of Cross Section Data to Characterize Macro Functions," Journal of the American Statistical Association, June, pp. 369-380.
- Stoker, T. (1983), "Aggregation, Efficiency and Cross Section Regression," MIT Sloan School of Management Working Paper No. 1453-83, July.
- Theil, H. (1954), Linear Aggregation of Economic Relations, North-Holland, Amsterdam.
- Theil, H. (1971), Principles of Econometrics, Wiley, New York.
- Theil, H. (1975), Theory and Measurement of Consumer Demand, 1, North-Holland, Amsterdam.
- Wold, H., and L. Jureen (1953), Demand Analysis, Wiley, New York.
- Wolff, P. de (1941), "Income Elasticity of Demand, A Micro-Economic and A Macro-Economic Interpretation," Economic Journal, 51, pp. 104-145.

FIGURE 1: POSSIBLE ENGEL CURVES

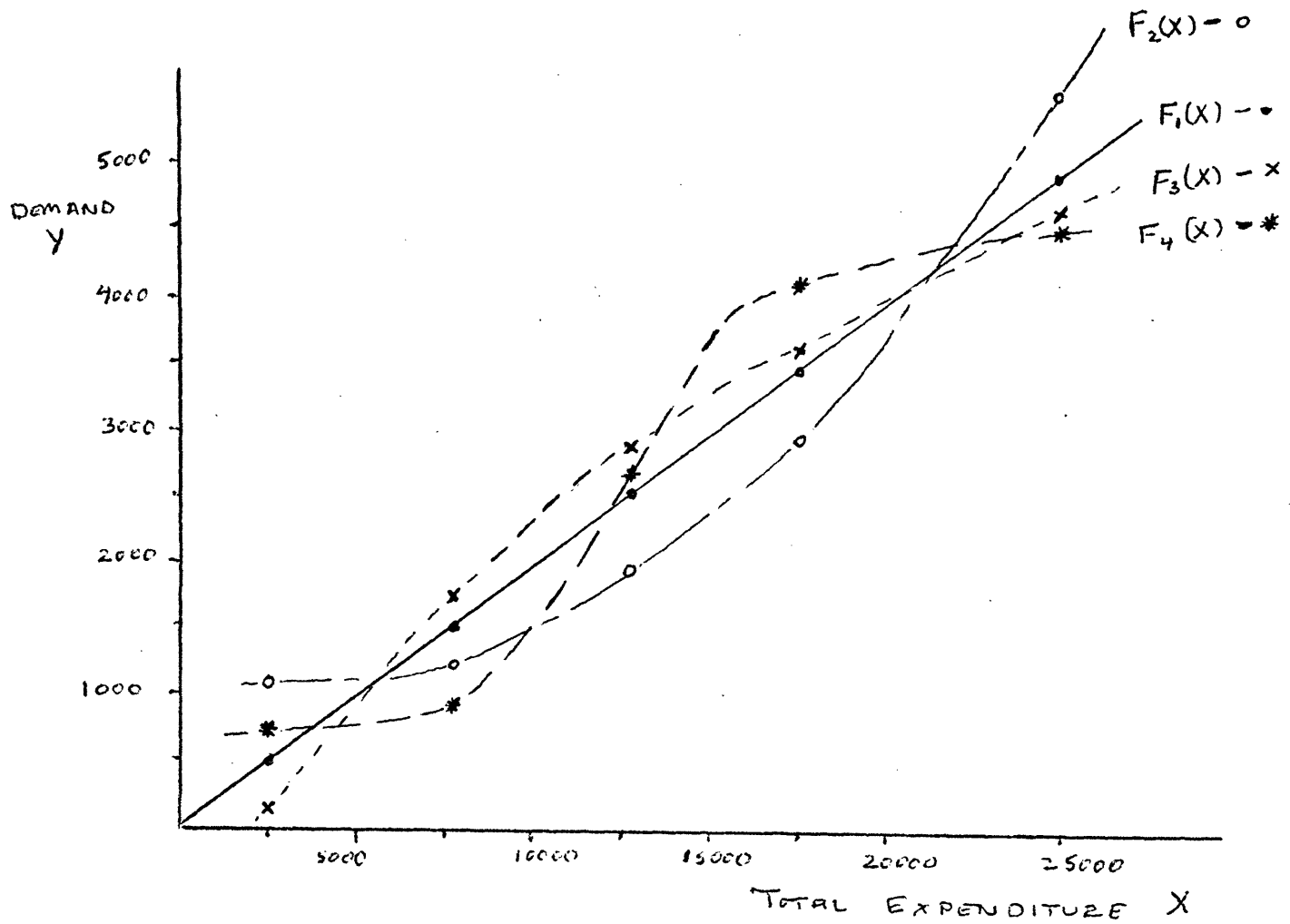


TABLE 1: DATA FOR ENGEL CURVE EXAMPLE

TOTAL EXPENDITURE CLASSES

within cell averages	TOTAL EXPENDITURE CLASSES				
	1 0-5000	2 5000-10000	3 10000-15000	4 15000-20000	5 20000+
X	2500	7500	12500	17500	25000
$y = F_1(X)$	500	1500	2500	3500	5000
$y = F_2(X)$	1219.96	1262.98	1849.04	2978.12	5689.91
$y = F_3(X)$	145.39	1700.83	2771.76	3642.24	4739.75
$y = F_4(X)$	700.00	845.00	2600.00	4255.00	4600.00

TABLE 2: COMPLETE DISTRIBUTION SCENARIO

t	P _{1t}	P _{2t}	P _{3t}	P _{4t}	P _{5t}	μ _t =E _t (X)	φ _{1t}	φ _{2t}	φ _{3t}	φ _{4t}
1	.3532	.1639	.1967	.2294	.0564	10,000	2000	2006	1978	2133
2	.3088	.1848	.1891	.1936	.1236	11,000	2200	2239	2174	2256
3	.2344	.2380	.1616	.2052	.1616	12,000	2400	2416	2400	2402
4	.2	.2	.2	.2	.2	13,000	2600	2600	2600	2600
5	.1255	.2331	.2124	.1916	.2370	14,000	2800	2760	2825	2744
6	.0811	.2540	.2049	.1558	.3041	15,000	3000	2994	3022	2867

TABLE 3: DISTRIBUTION SCENARIO WHERE COMPLETENESS FAILS

t	P _{1t}	P _{2t}	P _{3t}	P _{4t}	P _{5t}	μ _t =E _t (X)	φ _{1t}	φ _{2t}	φ _{3t}	φ _{4t}
1	.3032	.2540	.2049	.1557	.0819	10,000	2000	2000	2000	2000
2	.2688	.2360	.2032	.1705	.1213	11,000	2200	2200	2200	2200
3	.2344	.2180	.2016	.1852	.1607	12,000	2400	2400	2400	2400
4	.2	.2	.2	.2	.2	13,000	2600	2600	2600	2600
5	.1655	.1819	.1983	.2147	.2393	14,000	2800	2800	2800	2800
6	.1311	.1639	.1967	.2295	.2786	15,000	3000	3000	3000	3000

Key to Tables 2 and 3

P_{it} - Proportion of agents in cell i, i=1,...,5

φ_{jt} = E_t(F_j(X)) = Σ_i P_{it}F_{ji}, F_{ji} is the within cell average of F_j in cell i.

Table 3 distribution data constructed via

$$(P_{1t}, P_{2t}, P_{3t}, P_{4t}, P_{5t}) = (.2, .2, .2, .2, .2) + d (\mu_t - 13,000)$$

where d = $\tilde{X}/\tilde{X}'\tilde{X}$, $\tilde{X} = (-10,500, -5500, -500, 4500, 12,000)$.